

Natural Language Processing
Using Spreading Activation
and Lateral Inhibition

Jordan Pollack & David Waltz
Coordinated Science Laboratory
University of Illinois

Abstract

The knowledge needed to process natural language comes from many sources. While the knowledge itself may be broken up modularly, into knowledge of syntax, semantics, etc., the actual processing should be completely integrated. This form of processing is not easily amenable to the type of processing done by serial "von Neumann" computers. This work in progress is an investigation of the use of a spreading activation and lateral inhibition network as a mechanism for integrated natural language processing.

This work was supported in part by the Office of Naval Research under contract N00014-75-C-0612.

INTRODUCTION

It has long been thought that the modular decomposability of language knowledge into syntax, semantics and pragmatics implied that language processing could be similarly decomposed; that natural language could be processed by first parsing the syntax, then fleshing out the meaning of a syntactic derivation tree, and finally (if we could ever get to this point!) attempting to interpret the speaker's intentions. Nowadays, it has become apparent that this processing is integrated in humans [Marslen-Wilson, 1980], and that it should, thus, also be in computer models [Schank & Birnbaum, 1980; DeJong, 1980]. However, the natural inclination of von Neumann computers to run one-step at a time presents a severe roadblock to the kind of integration needed for NLP.

What is needed is an integration mechanism sensitive to interpretation pressures from several directions. A promising approach would seem to be the use of a quantitative spreading activation / lateral inhibition network. This kind of network, similar in conception to relaxation techniques for low-level vision, and to neural network models, works through the iterative adjustment of real-valued node weights.

PREVIOUS AND RELATED WORK

The term "spreading activation" is almost as overworked as the term "frame," but most systems which spread activation do it in one of two ways: As marker passing intersection search [Quillian, 1968; Collins & Quillian, 1972; Fahlman, 1980], in which a parallel intersection search is simulated by binary marking of adjacent nodes in a breadth-first manner, or as quantitative weight balancing, [Ortony, 1974; McClelland & Rumelhart, 1981], in which activation energies assigned to all nodes are iteratively adjusted, based on local activation energies and strength of connections. One of the well-known dangers of spreading activation is its potential for overkill; an intersection search, under certain circumstances, may generate too many

useless intersections, and quantitative adjustment may result in "heat death," where every node becomes activated. (A solution for this latter form of activation involves the use of decay, dampening factors, or the spread of negative energy - lateral inhibition.) Nonetheless, both forms of spreading activation display interesting behavior.

For example, the previously mentioned work by Collins and Quillian showed how spreading activation could account for aspects of human memory priming, while Fahlman's work demonstrated that many forms of problem solving could be simplified when an intersection search was computationally "free." Ortony, on the other hand, built a system for schema selection using damped activation, and McClelland and Rumelhart effected a close simulation of experimental results on human letter and word perception in context.

Other work in parallel approaches to natural language processing has been done by Small [1981] and Rieger [1977] in which the traditional practice of breaking down knowledge into syntax and semantics was turned on its head, and knowledge of all kinds was distributed to individual "word experts"; by Hendler & Phillips [1981] who are working on an ACTOR-based [Hewitt, 1976] NLP system; and by Gigley [1982] in which a neurologically-inspired NLP system capable of simulating aphasic behavior was built.

NATURAL LANGUAGE PROCESSING USING AN
ACTIVATION/INHIBITION NETWORK

The authors of this paper are presently building a NLP system in which the knowledge sources are modular, but the processing is fully integrated. The integration mechanism is an activation/inhibition network similar in nature to the one used by McClelland and Rumelhart and described below.

An activation/inhibition network is a weighted directed graph, where node weights represent activation levels, and link weights represent strength of activation if positive, or strength of inhibition if negative. The process of spreading activation / lateral inhibition involves iterative recomputation of activation levels. At each cycle, every node receives a contribution from each of its neighboring nodes equivalent to the neighbor's activation level multiplied by the weight of the intervening link. This contribution (scaled to range between -1 and 1) causes a proportional change in the activation level of the node; a contribution of 1 zaps the node up to its (predefined) maximum activation level while a contribution of -1 saps the node of all its strength. Eventually, a static condition is reached where some nodes reach their maximum or minimum strength, while the rest of them receive contributions of 0. (For a complete mathematical formulation see Pollack [1982b].)

NETWORK CONSTRUCTION

An activation/inhibition network such as this

can smoothly model the flow of quantitative constraints up and down a multilevel system. For natural language processing, the main problem becomes how to build such a multilevel network. We feel that a proper network can be built through the judicious instantiation of network fragments which are represented in standard knowledge representation structures, such as frames [Minsky, 1975].

The frames in our system contain the knowledge of syntax, of semantic features, and of case roles, organized to efficiently generate pieces of network on demand. These frames are richly interconnected with activation and inhibition links, and constitute the general knowledge base of the system. When sentences are input, a temporary network is constructed out of fragments stored within lexically accessed frames. These fragments are organized into a network by the same sort of breadth-first operation used in a chart parser [Kay, 1973]. The resulting network has activation links between phrase markers and their constituents, and inhibition links between pairs of phrases that have common constituents. (So far, we have done the network building by hand.)

In more detail, the required actions are as follows:

First, there is breadth-first instantiation of nodes representing phrase markers, case roles, and expectations for other nodes. These expectations are triggered when lexical items or grammatical constituents are encountered, and consist of simple feature patterns to match and connection procedures to be carried out if the match occurs. Secondly, there is pattern-based connection whereby if a newly instantiated node matches a pattern, specific linkages are made. As an example of these two processes, if a node of type NP is instantiated, it will then cause the instantiation of an expectation that a VP will occur; if a VP is found, an S is generated and connected to both the NP and VP. Of course, if more than one candidate for a pattern shows up, the two candidates are connected with an inhibition link, so one will eventually be eliminated.

The activation and inhibition processes reinforce nodes that are supported by activation links and inhibit those which are not, so, for example, expectations that are not quickly fulfilled will die. Furthermore, activation and inhibition are also happening in the background frame system by a purely word associative scheme, which helps prime good word senses (and aids in schema selection). Finally, nodes which become inhibited below a certain point are garbage collected thus keeping the active network as small as possible.

EXAMPLE OF OPERATION

Some preliminary results are presented here which demonstrate the feasibility of the activation/inhibition approach to NLP. As mentioned above, since the system is in its early stages, the networks presented were built by hand. We demonstrate how the system reacts to syntactic ambiguity, how a lexical preference can affect its behavior, and finally how semantic constraints can be integrated.

Consider, then, the following sentence, which, in the absence of any semantic knowledge, is syntactically ambiguous due to the lexical ambiguity of "up":

John ate up the street.

The hand-built network for this sentence is shown

in figure 1 with arrows denoting activation links, and circles denoting inhibition links (following McClelland & Rumelhart). Note that each node in this network is suffixed by two numbers which denote the "span" [Hobbs, 1974] or sequence of words, of that node.

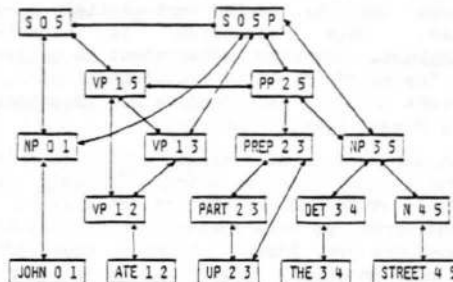


FIGURE 1 - SYNTAX ACTIVATION/INHIBITION NETWORK FOR "JOHN EATS UP THE STREET"

One would expect a robust NLP system to be confused by ambiguity but then to gracefully resolve it. This is indeed what happens. Figure 2 contains a graph of the activation levels over time for all the nodes in the network. Each node is depicted by a single letter, and each activation cycle by a horizontal row in the graph. When a letter traces a path to the left, it is being inhibited and when it moves to the right, it is being activated.

The most interesting node pairs to watch are B and C, the mutually inhibitory sentences, and G and F, the mutually inhibitory verb phrases:

(john01 is shown as @)	(part23 is shown as I)
(np01 is shown as A)	(prep23 is shown as J)
(s05 is shown as B)	(the34 is shown as K)
(s05p is shown as C)	(det34 is shown as L)
(ate12 is shown as D)	(street45 is shown as M)
(vp12 is shown as E)	(n45 is shown as N)
(vp13 is shown as F)	(np35 is shown as O)
(vp15 is shown as G)	(pp25 is shown as P)
(up23 is shown as H)	

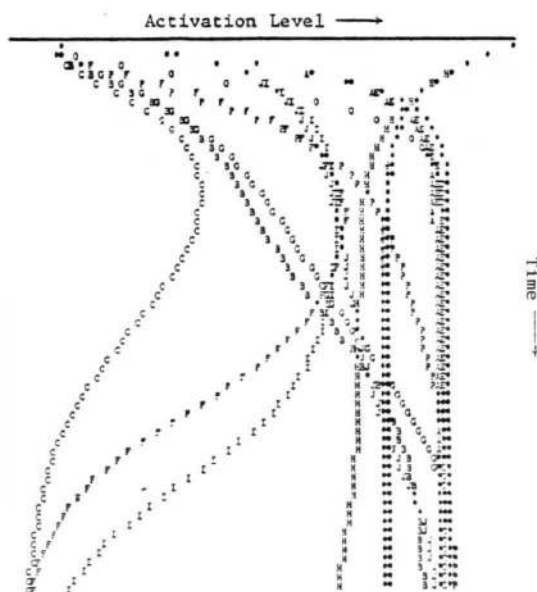


Figure 2 - "Confused"

B=(John) (ate (up the street))
 C=(John) (ate up) (the street)
 G=(ate (up the street))
 F=(ate up)

The system is confused at first: B is more heavily weighted than C, so the sentence with the preposition is selected, while F is more strongly activated than G, so the verb-particle phrase is selected. This selection is, obviously, inconsistent. But then, after about 30 cycles, the system "decides" ("Look Ma, no homunculus!") on a consistent reading of "up" as a preposition, and weights G more heavily than F.

In the absence of semantic preferences (e.g. a preference for interpreting "street" as a location), syntactic preferences can play a role. Certain words do have lexical tendencies, as, for instance, the word "does", which is most often a verb, but which is also a plural noun, meaning several female deer.

Figure 3 demonstrates the sensitivity of an activation/inhibition network to syntactic preferences. The link strength from "up" to "particle" has been increased, corresponding to a lexical preference. Notice that the phrases related to interpreting "up" as a preposition (B, G, J, and P) become inhibited much more quickly this time.

However, when humans process this sentence, they also take into account the knowledge that "street" is a good candidate for a location, but a bad candidate for the object of eating. The next example demonstrates the sensitivity of our NLP approach to this semantic knowledge. Four nodes have been added and connected into the network. The verb phrase "ate" is linked to "ate-loc" and "ate-obj," and the verb phrase "ate up" is linked to

{john01 is shown as @}	{part23 is shown as I}
{np01 is shown as A}	{prep23 is shown as J}
{s05 is shown as B}	{the34 is shown as K}
{s05p is shown as C}	{det34 is shown as L}
{ate12 is shown as D}	{street45 is shown as M}
{vp12 is shown as E}	{n45 is shown as N}
{vp13 is shown as F}	{np35 is shown as O}
{vp15 is shown as G}	{pp25 is shown as P}
{up23 is shown as H}	

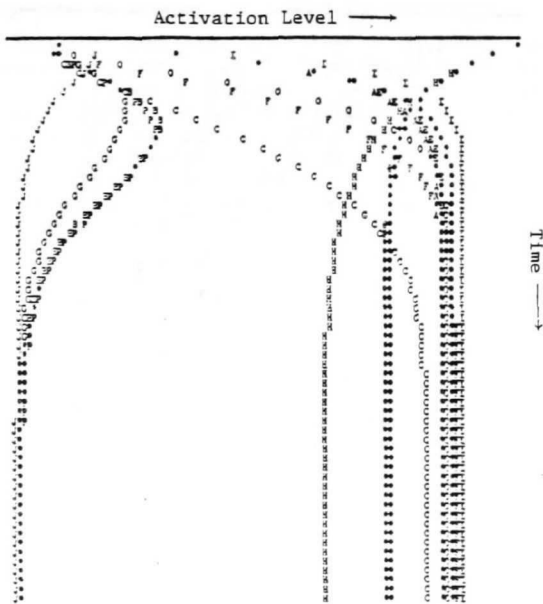


Figure 3 - Syntactic Preference for "Up" as Particle

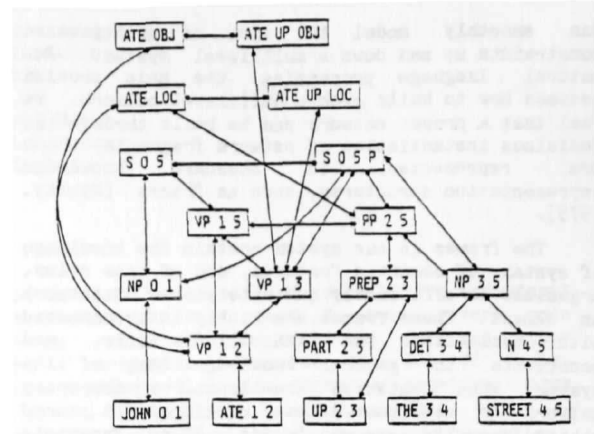


FIGURE 4 - SEMANTICALLY AUGMENTED NETWORK

"ate-up-loc" and "ate-up-obj." These nodes represent "cases" [Fillmore, 1968] of their respective nodes and are a subset of those that would be instantiated by our system. The pattern-matching connection component would connect the prepositional phrase "up the street" to "ate-loc" based on its span and on inherited features from "up" and "street".

The modified network is shown in figure 4, and figure 5 graphs the response of the activation/inhibition network to this new information. As one can see, after 15 cycles, all nodes related to interpreting "up" as a particle are being rapidly inhibited. (T, S, C, F, and I).

PROSPECTS

The results given above are interesting in that they demonstrate the sensitivity of activation/inhibition networks to slight

{john01 is shown as @}	{prep23 is shown as J}
{np01 is shown as A}	{the34 is shown as K}
{s05 is shown as B}	{det34 is shown as L}
{s05p is shown as C}	{street45 is shown as M}
{ate12 is shown as D}	{n45 is shown as N}
{vp12 is shown as E}	{np35 is shown as O}
{vp13 is shown as F}	{pp25 is shown as P}
{vp15 is shown as G}	{atoloc is shown as Q}
{up23 is shown as H}	{ateobj is shown as R}
{part23 is shown as I}	{ateuploc is shown as S}
	{ateupobj is shown as T}

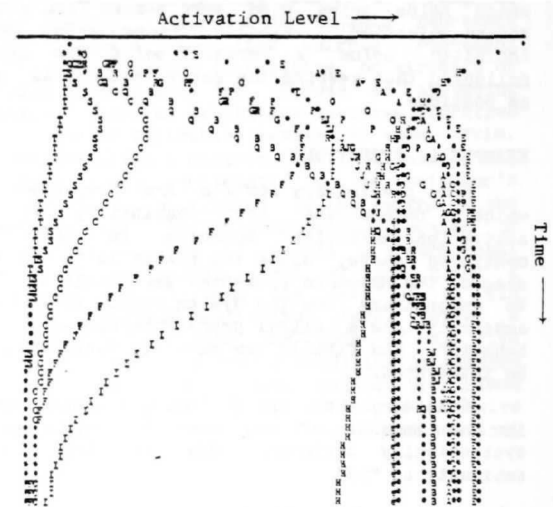


Figure 5 - Added Semantics

differences in knowledge. Currently we are working to complete the automatic instantiation and connection components of the system.

The use of a parallel and decentralized decision process can be brought to bear on many other interesting problems in NLP as well. For instance, there are indications that the timing and volume of spoken language both play useful roles in disambiguation [Wales and Toner, 1979]. A system based on activation and inhibition could be designed for sensitivity to these clues, since time is, after all, a crucial element in the activation/inhibition process.

Furthermore, the processing of garden path sentences, which are an interesting but not well-understood phenomenon in natural language, could quite possibly be handled by an activation/inhibition network. Marcus [1979] built a parser which attempted to account for garden-path sentences as a result of memory limitations. Unfortunately, there are garden path sentences his parser could (though shouldn't) handle [Milne, 1980], such as:

The prime number few.

Within the framework of activation/inhibition networks, garden path sentences would be accounted for by irreversible inhibition of expectations.

Also we have recently begun to consider ways of integrating a novel form of knowledge representation, "event shape diagrams" [Waltz 1982], to model certain kinds of metaphor understanding and adverbial modification. As an example, these methods should allow us to interpret sentences such as:

Robbie's metal legs ate up the space between himself and Susie.

as meaning a kind of PTRANS [Schank 1975].

Finally, a practical system based on activation/inhibition networks could be the starting point for new computing architectures. In this vein, [Pollack, 1982] has designed a VLSI cell for parallel simulation of activation/inhibition networks, thus showing that a programmable set of logical connections (i.e. links) can be run on a machine with fixed and regular physical connections (i.e. wires).

CONCLUSION

The processing of natural language requires the sensitive integration of multiple sources of knowledge. A mechanism very likely to achieve this integration is an activation/inhibition network.

REFERENCES

Collins, A. and M.R. Quillian, 'Experiments on semantic memory and language comprehension' in L.W. Gregg (Ed.) Cognition in Learning and Memory, Wiley, New York, 1972.

Church, K. and R. Patil, "Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table," presented at 56th Linguistic Society of America meeting, December, 1981.

DeJong, G., "Prediction and Substantiation: A new

approach for Natural Language Processing" Cognitive Science V.3 #.3 pp. 251-273, 1980.

Fahlman, S.E., NETL: A system for Representing and Using Real-World Knowledge, MIT Press, 1979.

Gigley, H., "Neurolinguistically Based Modeling of Natural Language Processing," Doctoral Thesis (in preparation), University of Mass, Amherst, 1982.

Hendler, J. and B. Phillips, "A Flexible Control Structure for the Conceptual Analysis of Natural Language Using Message-Passing," Technical Report 08-81-03, Texas Instruments, 1981.

Hewitt, C., "Viewing Control Structures as Patterns of Passing Messages," MIT AI Memo 410, 1976.

Hobbs, J.R., "A Metalanguage for Expressing Grammatical Restrictions in Nodal Spans Parsing of Natural Language.", Report NSO-2, Courant Institute, NYU, January 1974

Kay, M., "The MIND System", in Rustin (Ed.) Natural Language Processing, Algorithmics Press, New York, 1973. Marcus, M.P., A Theory of Syntactic Recognition for Natural Language, MIT Press, 1980.

Marslen-Wilson, W. and L. K. Tyler, "The Temporal Structure of Spoken Language Understanding," Cognition V.8 #.1 pp. 1-72, 1980.

Milne, R., "Using Determinism to Predict Garden Paths," DAI Research Paper 142, University of Edinburgh, 1980.

Minsky, M., "A framework for Representing Knowledge", in Winston (Ed.) The Psychology of Computer Vision McGraw Hill, New York, 1975.

McClelland, J.L. and D.E. Rumelhart, "An Interactive Activation Model of the Effect of Context in Perception", Technical reports 91 & 95, Center for Human Information Processing, UCSD, 1980.

Ortony, A., "SAPIENS: Spreading Activation Processing of Information Enclosed in Associative Network Structures", unpublished, 1976.

Pollack, J., "An Activation/Inhibition Network VLSI Cell", WP #31, Advanced Automation Group, Coordinated Science Laboratory, Urbana, January, 1982

Pollack, J., "An Activation/Inhibition Approach to Natural Language Processing", WP #35, Advanced Automation Group, Coordinated Science Laboratory, Urbana, April, 1982

Schank, R.C., "The Primitive ACTS of Conceptual Dependency", in Schank & Nash-Webber (Eds.) Theoretical Issues in NLP, ACL, Arlington, Va. 1975.

Schank, R.C. and L. Birnbaum, "Memory, Meaning, and Syntax," Research Report 189, Yale C.S. Department, November 1980.

Rieger, C., "Viewing Parsing as Word Sense Discrimination," in R. Dingwall (Ed.) A Survey of Linguistic Science Greylock, 1977.

Wales, R. and H. Toner, "Intonation and Ambiguity", in W.E. Cooper and C.T. Walker (Eds.) Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett, Erlbaum, New Jersey, 1979.

Waltz, D.L., "Event Shape Diagrams", To Appear in Proc. NCAL, Pittsburgh, August, 1982.