RC 22339 (98613) 15, January 2001 **Computer Science**

IBM Research Report

Extensible Shallow Parsing for Semantic Nets

Jonathan Connell
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

January 2001

Extensible Shallow Parsing for Semantic Nets

Jonathan H. Connell
IBM T.J. Watson Research Center
30 Saw Mill River Road
Hawthorne NY 10532 USA
jconnell@us.ibm.com

Abstract

This paper proposes a specific linguistic-based format for semantic networks in which nodes correspond to "open class" words. "Closed class" words and morphological elements form the basis for atomic link labels and node tags. A simple parser has been developed to transform written text into this representation. The properties of the resulting networks are discussed and psychologically inspired limited-horizon browsing techniques are examined.

1. Introduction

Semantic networks as a form of knowledge representation have a long history in AI and cognitive psychology. There is the spreading activation study of Quillian [68], the propositional and multi-code networks of Anderson [76; 83], Winston's arch learner [75], natural language parsing [Sowa 84], reasoning in KL-ONE [Brachman and Schmolze 85], vision-based descriptors [Connell and Brady 87], lexical affinities in WordNet [Miller *et al.* 93], and thesaurus-based information retrieval [Clark 00]. Semantic networks generally encode knowledge in smaller granularity chunks than alternatives such as frames or scripts. They are essentially equivalent to association lists or database tables – two other popular representation formalisms – but do not have as strong ties to particular processing techniques as these do. In some sense, two nodes and a link between them is the smallest particle of knowledge that can be expressed.

While semantic networks based on triples [Winston 82] of node-link-node are popular and flexible, they have trouble directly representing certain facts. Basically, links are often assumed to be predicates that take two arguments (sort of S-V-O). This means straight property ascription has to be rendered with a semantically empty link like "hq" (has-quality) to the associated concept. Di-transitive verbs need an auxiliary "to" adverbial link off the matrix verb to denote their third argument. Predicates like "between" require a reification of the interval between the specified two extents in order to fit the standard format.

In general, adverbial modification (links off links) and sentential embedding generates a network with more links than nodes. Consider Winston's example where:

```
((Macbeth murders Duncan) because
 (Macbeth desires
   (Macbeth to-be-a-kind-of king))).
```

This suggests that the links themselves may be the more important part of the representation. Other researchers (among them [Sowa 84]) have used this to argue that verbs themselves should be given nodes in more of a case-frame style representation. While this paper suggests a similar approach, it does not directly specify which role arguments are required and optional for each verb class. Instead, we aim to create a network where the link labels are simply not amenable to further modification, thus flattening the resulting graph structure.

2. Open and Closed Classes

Closed classes are those categories of words which do not readily admit new members. Such categories include prepositions, determiners, auxiliary verbs, pronouns, and conjunctions, among others. By contrast new open class words – nouns, verbs, adjectives, and adverbs – are added to languages all the time. Figure 1 shows an excerpt from a Curious George children's book. Closed class words have been rendered in lower case and open class words in upper case. Notice that most of the information content is carried by these capitalized words; the other words just serve to glue the concepts together.

```
this is GEORGE . he LIVES in the HOUSE
of the MAN with the YELLOW HAT . GEORGE
is a LITTLE MONKEY , and all MONKEYS
are CURIOUS . but no MONKEY is as
CURIOUS as GEORGE . this is why his
NAME is CURIOUS GEORGE .
```

Figure 1 – Only open class words are capitalized. These form the basis for the nodes in the network.

```
live -s
-> ***
   -- he
-in->
   house
   -- the
   -of->
      man
      -- the
      <- ***
         -with->
            hat
            -- the
            <- yellow
```

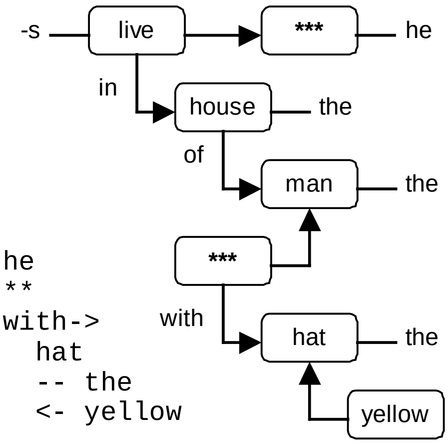

Figure 2 – The semantic network (right) and a tree representation (left) for: *He lives in the house of the man with the yellow hat*. Note that nodes correspond to open class words while closed class words and morphological elements become either link labels or node tags. The "***" nodes have no open class head term.

For this reason we propose that the nodes in a semantic network be based on these open class terms, and the links on the remaining "function" words. In such networks it does not make sense to attach an additional modifier to some link. Also, the actual link labels and tags can be dropped to generalize the meaning. Figure 2 shows the output of the proposed system for the second sentence in Figure 1's example text. The left side shows a tree-like representation of the parse highlighting the argument structure. The right side shows the equivalent semantic network.

The head of each phrase is used as the label for a node (e.g. "hat") and that modifiers of such nodes are also realized as discrete nodes (e.g. "yellow"). The bulk of the links are labeled by prepositions although subject links (any unmarked noun phrase occurring before the verb) are marked by plain single arrows. Although there are no examples here, unmarked noun phrases following the verb would marked with a double arrow (==>) to preserve some indication of surface order. Unary predicates, such as the determiner "the" and the suffix "-s", are connected to the relevant node with a plain line indicating they are "tags".

The handling of prepositions is somewhat non-standard. Our system considers all prepositions to be role markers for different noun-like constituents of a verb phrase. When a prepositional phrase appears to directly modify a noun (such as "with" in Figure 2) the system insert a dummy verb phrase node (here "***"). This allows the resulting net to directly match, node for node, the nets produced by conceptually similar sentences like "the man wearing a hat". The only exception to this handling technique is the preposition "of". Such phrases are allowed to directly attach to noun phrases without an intermediary (cf. "house" and "man" in Figure 2 for an example of a "naked genitive").

`this is george . he` **`live-s`** `in the house of the man with the YELLOW HAT . george is a LITTLE MONKEY , and all` **`monkey-s`** `are curious . but no monkey is as curious as george . this is why his name is CURIOUS GEORGE .`

Figure 3 – The usage of most words in a sentence can be determined from partial knowledge. The plain lowercase words are from closed classes. The category of the bold faced words can be obtained from their morphological endings. The category of the underlined words can be inferred because they are sandwiched between two closed class word (or a delimiter). The only remaining ambiguous cases are the capitalized phrases.

3. Extensible Parsing

Although other techniques such as cascaded finite state automata [Hobbs *et al*. 96; Silberztein 00] could be used, the parser that produces the network representations is based on Berwick's version [85] of the Marcus parser [80]. There is a phrase stack and a short look-ahead buffer containing categorized constituents (typically words).

A question for this parser, and indeed any parser, is where does this categorization (commonly called part of speech tagging) come from? In the proposed system part of the information comes from the built-in closed class lexicon. This contains on the order of 250 small words like prepositions and pronouns. In terms of language learning this is not an unreasonable number of items to memorize, especially since these tend to be very high frequency terms (in fact they are typically on the "stop list" in an information retrieval system [Salton 89]).

Another potent source of information are the inflectional endings. The parser looks specifically for words ending in "-ing", "-ed", "-s" (plural or active tense), and "-ly". Such endings tell not only the class of the complete word, but also the class of the stem (e.g. adjective + "-ly" = adverb). The parser attempts to restore the stem to its correct orthographic form and then stores it in a separate open class lexicon for later use. As a side note, while the spelling rules in such cases can be complex, the auditory rules are usually very simple and amount to simple appending the sound of the suffix to the sound stream for the stem. Again the chosen suffixes occur with high frequency and so should be easy to acquire. In fact, it has been shown [Brown 73] that indeed these are some of the first morphemes children learn.

Still another source of information consists of tight context markings. Iin Figure 3 there are a number of unknown words which are directly bracketed by known closed class words. For instance, in the phrase "the man with", the initial determiner unambiguously starts a new noun phrase, while the final preposition suggests that the noun phrase has terminated. Since all noun phrases require a head of the category “noun”, this must be the label of the word “man”. Again, the word and its label are added to the open class lexicon. This is what is meant by extensible parsing – being able to handle a small number of unknown words in any sentence.

Remaining ambiguities can often be resolved by even weak phrase structure rules. For instance the parser includes the rules shown below which encode the fact that modifiers typically precede the head of a phrase. In the rules the marker before the colon is what type of node is on the top of the stack (“*” means any). The bracketed items denote sequential items in the look-ahead buffer. These are labeled with their part of speech or “X” if unknown. The right hand side of each rule describes what to do with the contents of the numbered (left to right) buffer positions.

```
 *: [DET]       → new NP.tag  = 1
NP: [X]  [X]    →     NP.mod  = 1
NP: [X]  [NPi]  →     NP.head = 1
```

Now considered the phrase “the yellow hat <period>”. The first word triggers the top rule and causes a new noun phrase to be created with the word “the” attached as a tag. Next the word “yellow” is read in but no action occurs because its type is unknown. Then the further word “man” (also unknown) is read in. This triggers the middle rule which attaches the previous word (buffer element 1 = “yellow”) as a modifier. Finally, the <period> symbol is read (could be a lengthy pause in a speech system) and the bottom rule is triggered. A period is incompatible (label NPi) with a noun phrase hence any pending unknown word (“man” in this case) is attached as the head of the phrase in order to complete it. Thus a reasonable argument structure can be guessed despite lack of lexical knowledge. Whether the assigned roles are added to the lexicon or not is a design decision (e.g. sometimes nouns, not adjectives, modify other nouns).

4. Navigating the Net

Given that we have the ability to generate semantic nets, just how useful and perspicuous are they? One way to gauge their utility is to cast them as Web pages (an optional output from our parser). For instance focusing on the “live” node from Figure 2 we get something like Figure 4 where the network links are true HTML hyperlinks. If one clicks on a word like “house” a similar Web page pops up with the view of the network from that node. Notice that when a link to a page is generated all that is retained in the textual label is the head category of the other node on that page.

An interesting alternative view is shown in Figure 5 which contains the same information as Figure 4. Here we assume there are regions of space corresponding to nodes of various types. Conceptually a particular “grandmother” cell in each region “lights up” (shown as black diamond) for each particular instance of a node type. Alternatively, some

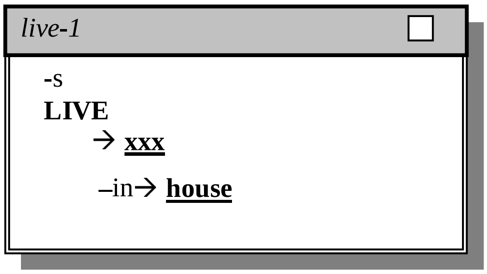

Figure 4 – The view from a node can be represented as a Web page. Here the head concept is shown capitalized with any tags (function words or endings) placed above it. Below are links to other nodes which are coded with the relation (if any) and the head type of the connected node. Clicking on the underlined words leads to similar Web pages for these nodes.

distributed representation [Kanerva 88] might equivalently be used to denote a particular instance. Regions can also have subregions corresponding to modification of the basic head term by “tags” or role markers (directional link labels). It is assumed that such regions could be grouped close to each other (i.e. syntactic proximity mirrors semantic proximity) using a method such as Kohonen maps [95].

In Figure 5 the activation point for the particular “live” node indicates that a “-s” morphological element was also present. The feature space also indicates that a “house” node is linked to this one as the object of the preposition “in” (i.e. “house” is the locative role filler). Similarly, a node of unknown type “***” is the subject of the “live” relation. The type is unknown since in Figure 2 the equivalent node is only tagged with “he” (which gives it number and gender but not category).

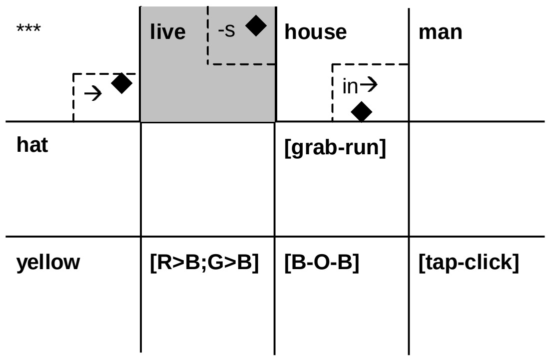

Figure 5 – The view from a node can also be depicted as a feature vector. The input space here is divided into discrete labeled feature regions, some with sub-specializations. The black dots represent activations in these regions. This example encodes approximately: “X live-s in-house”. Non-verbal predicates can also be included as shown in some of the bracketed labels.

This is essentially a feature set representation with outgoing links "curried" down to a unary predicate composed of the link label and head category (i.e. In-House(x) is true of this particular instance of "live"). Note that this is more than a connectionist encoding of a single "triple" [Hinton 90] because multiple predicates are represented simultaneously. Feature set representations are desirable because they are easily interfaced to learning programs and robot control systems.

Such representations also allow simple incorporation of non-lexical inputs as indicated by some of the other region labels in Figure 5. For instance, near the lexical term "yellow" might be the output of a visual classifier which lights up when the red and green components of a visual area are much brighter than the blue component (cf. [Horswill 96]). Or an area might respond to the probability associated with the terminal state of some HMM acoustic recognizer, like the phoneme sequence B-O-B. Similarly, an area might respond to the final state of some FSM behavioral controller. If the robot just grabbed something then ran away this could be interpreted as an act of "stealing". One could also imagine regions which correspond to saved state based on exploratory routines such as taping an object and listening for whether there is a click (an indication that the surface is "hard").

A browsing facility like shown in Figure 4 is certainly useful, and its feature space equivalent shown in Figure 5 is compelling from an engineering perspective. Yet it feels like not quite enough information is being presented in a single glance. For example, viewing the net from the "man" node in Figure 2 one only sees the direct connection to the "house" and "***" nodes. Since traditional slot-filler representations usually have two meaning carrying parts, it makes sense to extend the viewing horizon to two links away from the focus. This is shown in Figure 6.

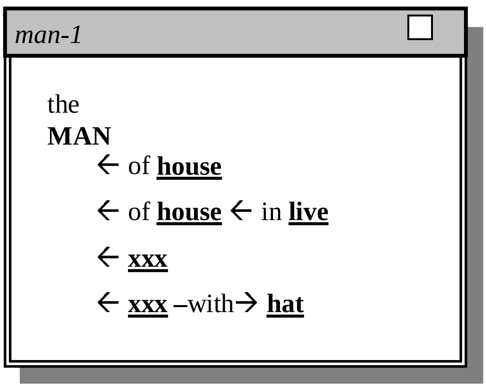

Figure 6 – Expanding the Web view to show up to two links away feels more natural. Here it can be seen directly that this is the man with the hat and the man who has a house that something lives in.

Unfortunately this does not translate easily into the feature set representation. One approach to having an expanded horizon yet retaining the "neurological" feel to the system is shown in Figure 7. Here the substrate feature space is time-shared [Shastri and Ajjanagadde 93] between a small number of nodes which the system is attending to. The first two temporal "phases" are devoted to nodes outside the current network. The third phase however corresponds to the "live" node represented in Figures 4 and 5 (identical pattern of activation). When this node is deliberately in focus (darkened phase marker) any remaining un-focused phases can be filled with nodes directly linked to this one. Thus at least the most important horizon-2 nodes can be automatically brought to attention.

Based on the surreptitiously loaded (or "expected") node for "house", the system might become interested in the linked "man" node and bring it into focus (as phase 5). Again, this would cause the most important nearby nodes (such as "***") to auto-load themselves. Based on this the system might intentionally focus on the linked "hat" node and load its representation as well. Thus with multiple phases available and a suitable focus policy, the effect of a two link horizon can be simulated while still retaining the benefits of the feature space view.

5. Limitations

The parser used in this study performs only a shallow analysis of each sentence. There are many syntactic phenomenon it does not handle such as subject-verb agreement and quantification, among others (cf. [Allen 95]). Nevertheless, it does a reasonable job with straightforward text. Thus it could be used as a user interface to generate semantic nets from somewhat constrained stylized input.

The system could also be used in text retrieval to generate a better notion of proximity. Consider these:

A stray **dog** was recently **bitten** by a deranged **man**.
When the **man bit** into his sandwich, his **dog** whined.

If we are looking for a case of "man bites dog" a standard retrieval system might have problems with the first sentence because none of these terms is directly adjacent. However in the semantic network there is the path dog-bite-man. On the other hand if the notion of proximity is loosened so that the probe words just have to occur in the same sentence, then the second example would generate a false hit. Yet when processed to yield a semantic net, the argument structure is made more explicit and there is no direct dog-bite-man node path present.

One of the biggest shortcomings of the current system is anaphora resolution. Although there are a wide variety of types [Hirst 81] the system does not even do simple inter-sentential pronoun resolution. A possible way to handle this would be to have an associative memory running in the background that constantly looked for matches of the

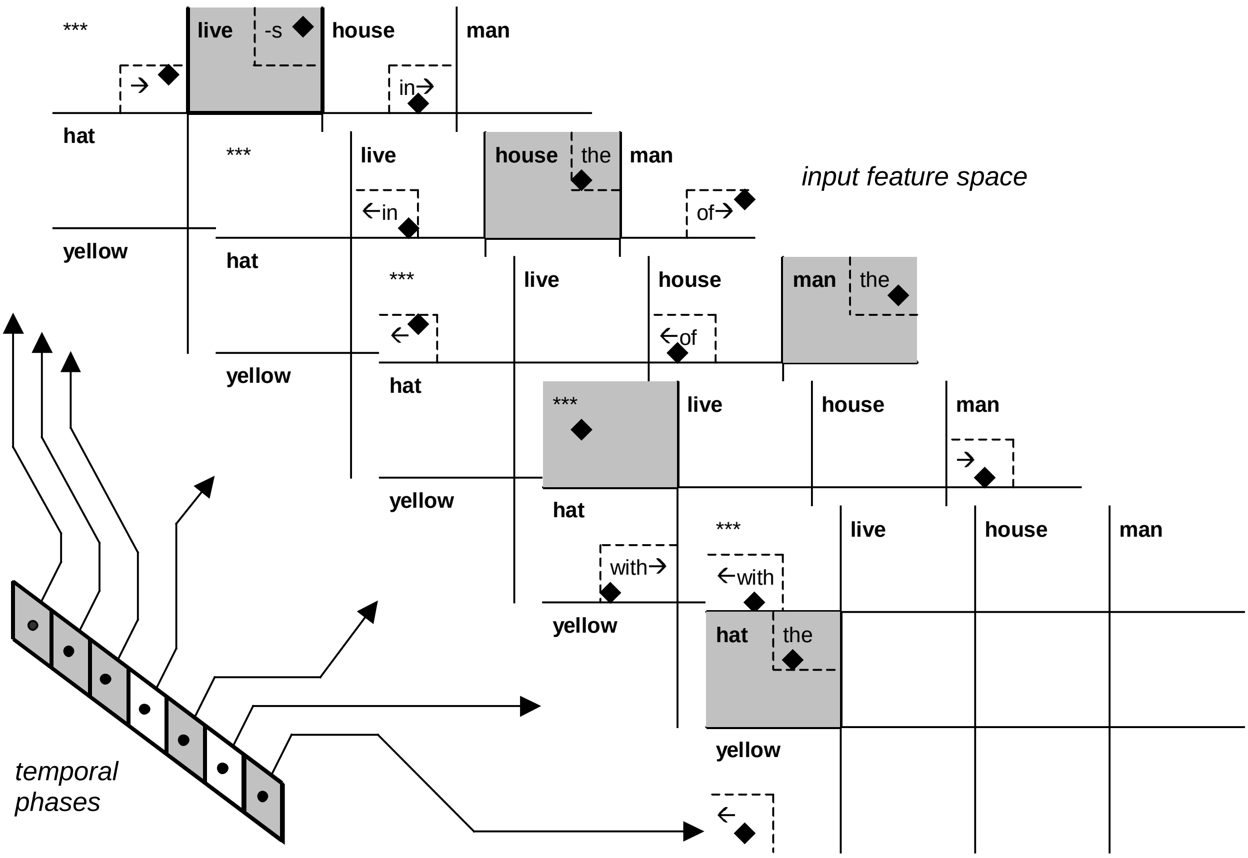

Figure 7 – A larger fragment of the network can be visualized by time-sharing the input feature space. The backmost copy of the space shows the same activation pattern as Figure 5 (corresponding to the Web page view of Figure 4). The closer copies of the space show the network expanded around different nodes. The input region for the head term of each node is shaded for illustration purposes. Phases are opportunistically filled (white phases) with the most important neighboring nodes relative the focal nodes (shaded phases).

current feature space activation pattern to previously stored nodes. If a good enough match was found (based on recency, specificity, etc.) the two representation could be merged. This mechanism might also be applicable to more difficult cases like definite noun phrase anaphora.

A related phenomenon is prepositional phrase attachment. Using a similar mechanism it may be possible to dynamically evaluate the likelihood of the phrase attaching to the current node based on a statistical "overlay" of all past nodes of the same type. However, it would seem necessary to group attached nodes not only by head type but also by the actual preposition used. This suggests that the subregion system for link labeling used in Figure 5 may need revision to make such information more readily apparent (work in ellipsis [Shaw 98] also hints at this).

Conjunction is also not handled by the current system, yet is prevalent in real text. One approach would be to "multiply out" all the separate paths. So, for instance, in the example below there would be 16 basic assertions of the form "Bill dried the spoons from lunch". Yet in some sense the grouping of Bill and Ted is not accidental – the following sentence might have the pronoun "they" which referred back to them <u>as a group</u>.

```
(Bill and Ted)
(washed and dried)
(the spoons and forks)
(from lunch and dinner).
```

Another phenomenon that seems important to handle is quotation. This occurs even in fairly formal narratives and is quite common in speech:

Then she goes, "Well, at least I never …"

At a systems level, the phases in the time-shared feature space already give a sense of "aboutness" while the subregions in the space provide “aspectuality”. The only part missing from Dennett’s [78] prescription seems to be linked with this quotation facility.

6. Conclusion

This paper has presented a parser for converting English text into semantic nets. The parser is a standard shift-reduce type and has a number of limitations – only a shallow parse of the sentence is accomplished. Nevertheless, the resulting semantic networks do a reasonable job of capturing the anchoring concepts and the rough argument structure between them.

The nets themselves can be visualized in a Web format, or as a feature space where links are condensed into predicates denoting the head types of the connected nodes. The feature space view from a single node is also attractive because it provides a simple means for interfacing the system to other sensory modalities. However, the single link horizon seems too miserly compared to standard slot-value representations. For this reason, the feature space idea was extended to a superimposed collection of such representations, similar to a short term memory.

Obviously this work is just a start and much remains to be done. Interesting avenues of exploration involve anaphora resolution, handling of conjunction, production rule matching, and extensions for computer vision.

References

[Allen 95] James Allen, *Natural Language Understanding,* Benjamin/Cummings, 1995.

[Anderson 76] John R. Anderson, “Chapter 5: ACT’s Propositional Network”, *Language, Memory, and Thought,* Lawrence Erlbaum, 1976.

[Anderson 83] John R. Anderson, “Chapter 2: Knowledge Representation”, *The Architecture of Cognition*, Lawrence Erlbaum, 1983.

[Berwick 85] Robert C. Berwick, *The Acquisition of Syntactic Knowledge,* MIT Press, 1985.

[Brachman and Schmolze 85] R. Brachman and J. Schmolze, "An Overview of the KL-ONE Knowledge Representation System", *Cognitive Science,* 9(2), 1985

[Brown 73] Roger Brown, *A First Language: The Early Stages,* Harvard Univ. Press, 1973.

[Clark *et al.* 00] Peter Clark, John Thompson, Heather Holmback, and Lisbeth Duncan, “Exploiting a Thesaurus-based Semantic Net for Knowledge-based Search”, *Proc. of IAAI-2000,* pp. 988-995, 2000.

[Connell and Brady 87] Jonathan H. Connell and J. Michael Brady, “Generating and Generalizing Models of Visual Objects”, *Artificial Intelligence,* 31(2), pp. 159-183, 1987.

[Dennett 78] Daniel Dennett, *Brainstorms,* Harvester, 1978.

[Hinton 90] Geoffrey Hinton (ed.), *Connectionist Symbol Processing,* MIT Press, 1990.

[Hirst 81] Graeme Hirst, *Anaphora in Natural Language Understanding: A Survey,* Lecture Notes in Computer Science 119, Springer-Verlag, 1981.

[Hobbs *et al.* 96] Jerry R. Hobbs, Douglas Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel, and Mabry Tyson, “FASTUS: Extracting Information from Natural-Language Texts”, in *Finite State Devices for Natural Language Processing,* E. Roche and Y. Schabes (eds.), MIT Press, 1996.

[Horswill 96] Ian Horswill, “Variable Binding and Predicate Representation in a Behavior-based Architecture”, *Animals to Animats 4* (SAB-96 conf.), pp. 163-172, 1996.

[Kanerva 88] Penti Kanerva, *Sparse Distributed Memory,* MIT Press, 1988.

[Kohonen 95] Teuvo Kohonen, *Self-Organizing Maps,* Springer-Verlag, 1995.

[Marcus 80] Mitchell P. Marcus, *A Theory of Syntactic Recognition for Natural Language,* MIT Press, 1980.

[Miller *et al.* 93] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, “Five Papers on WordNet”, Princeton Univ., *http://www.cogsci.princeton.edu/~wn*, 1993

[Quillian 68] M. Quillian, “Semantic Memory”, in *Semantic Information Processing*, M. Minsky (ed.), MIT Press, 1968.

[Salton 89] Gerard Salton, *Automatic Text Processing,* Addison-Wesley, 1989.

[Shastri and Ajjanagadde 93] Lokendra Shastri and Venkat Ajjangadde, “From Simple Associations to Systematic Reasoning”, *Behavioral and Brain Sciences*, 16(3), pp. 417-494, 1993.

[Shaw 98] James Shaw, “Segregatory Coordination and Ellipsis in Text Generation”, *Proc. of COLING-ACL*, pp. 1220-1226 , 1998.

[Silberztein 00] Max Silberztein, “Overview of INTEX”, http://www.ladl.jussieu.fr/Intex/new_intex.html, 2000.

[Sowa 84] John Sowa, *Conceptual Structures: Information Processing in Minds and Machines,* Addison-Wesley, 1984.

[Winston 75] Patrick Winston, “Learning Structural Descriptions from Examples”, in *The Psychology of Computer Vision,* P. Winston (ed.), McGraw Hill, 1975.

[Winston 82] Patrick Winston, “Learning New Principles from Precedents and Exercises”, *Artificial Intelligence,* 19(3) pp. 321-350, 1982.