A Probabilistic Approach to Marker Propagation

Dekai Wu*
Computer Science Division
University of California at Berkeley
Berkeley, CA 94720

Abstract

Potentially, the advantages of marker-passing over local connectionist techniques for associative inference are (1) the ability to differentiate variable bindings, and (2) reduction in the search space and/or number of processing elements. However, the latter advantage has mostly been realized at the expense of accuracy and predictability. In this paper we consider a class of associative inference to which marker passing is often applied, variously called abductive inference, schema selection, or pattern completion. Analysis of marker semantics in a standard semantic net representation leads to a proposal for more strictly regulated marker propagation. An implementation strategy employing an augmented relaxation network is outlined.

1 Introduction

Both marker-passing and local connectionist[1] models have been applied to a class of inference we call *recognition.* The essence of the task is to find a construal for some set of input evidence, by retrieving enough additional structure from the knowledge base to "fill in the gaps" and thereby infer an explanation for the input. The construal should be the most plausible, coherent explanation that can be found within a relatively instantaneous unit of time. A natural language processing example is interpreting *gooseneck lamp* as a lamp whose shaft is a gooseneck, and *pool table lamp* as a lamp hanging over a pool table, in each case by finding the schemata that best relate the individual concepts. The task of interpreting partially obscured visual patterns has also been cast as a problem of selecting schemata that best account for the input evidence [McClelland and Rumelhart 1981]. Various authors view the same or similar tasks as pattern completion, abductive inference to explain the occurrence of particular combinations of concepts, schema selection and composition, concretion [Wilensky et al. 1988], or classification-style realization [Mark 1982, Schmolze and Lipkis 1983].

With marker-passing models, input is given by simultaneously placing *origin markers* on several concepts in a semantic network. The system then propagates markers outward from the concepts of origin, effectively doing a parallel intersection search for relational paths between the concepts. With local connectionist models, input is given by raising the *activation* of several nodes, causing activation to spread to neighboring nodes and eventually highlighting the relational paths. The primary difference is that connectionist models propagate only numeric activation, rather than symbolic information.[2]

No clear winner has emerged between marker-passing and connectionist models. Although local connectionist models can technically be regarded as numeric marker passing, researchers have exploited the connectionist restriction by analyzing activation propagation more carefully than marker propagation, giving connectionist systems more precise search characteristics. On the other hand, symbolic markers retain the advantage of handling variable bindings, that is, the binding of general concepts and roles to particular occurrences. Binding is a serious problem in connectionist systems, since current approaches require too large a network to be practical.

In this paper, we examine another potential advantage of marker passing, which is reduction in the search space and/or number of processing elements, by restricting the search to localized areas of the network. Marker-passing models have not yet realized this advantage; those which do restrict marker propagation rely on arbitrary assumptions that lead to inaccurate search. This paper analyzes how marker propagation should be regulated so as to reduce search in a motivated fashion. The proposal, like several other recent proposals, is a hybrid model that attempts to synthesize advantageous aspects of marker-

*This research was sponsored in part by the Defense Advanced Research Projects Agency (DoD), monitored by the Space and Naval Warfare Systems Command under Contract N00039-88-C-0292, the Office of Naval Research under grant N00014-80-C-0732, and the Sloan Foundation under grant 86-10-3.

[1] In local connectionist models, each node represents a concept, whereas in distributed models, concepts are represented by a pattern of activation over multiple nodes [Hinton et al. 1986]. We do not discuss here the advantages and disadvantages that marker-passing and local connectionist models share by virtue of employing local representations.

[2] One aspect of marker passing that has no parallel in connectionist models is that it is often used as a heuristic for generating possible inferences, with an evaluation stage that selects the best ones. Our analysis excludes evaluation stages.

passing and connectionist techniques.

2 Issues

2.1 Marker propagation and resource limits

The main issue this paper addresses is how to regulate marker propagation to make computation feasible. For sequential implementations, marker passing is a theory of search, and the goal is to reduce the search space. For parallel implementations, the goal is to reduce the number of processing elements. (In this paper we use "propagation" to mean creating a new marker on a neighboring concept, as opposed to moving the same marker from concept to concept. Markers created by propagation are non-origin markers.)

Existing models do not limit propagation in a satisfactory way. The reason is that most marker-passing approaches depend on an "over-propagation" strategy to ensure finding the desired concepts or paths; i.e., markers are propagated to relatively large distances in all directions, so that the chance of missing a desired path is small. (Because extensive propagation tends to find too many remote connections, special filters are used to eliminate spurious paths, based on heuristic evaluation.) The problem is that it is not clear how much over-propagation is sufficient, and so either the search is susceptible to arbitrary failure, or it makes excessive resource demands. For example, some systems restrict propagation by presetting the number of links a marker can be propagated [e.g., Hirst 1987], but because number of links does not reflect semantic distance in a well-defined way, no preset limit guarantees the desired path can always be found within that number of links. More sophisticated techniques employ numeric activations on markers. The simplest such method is to assign lower activation for markers that are farther from the point of origin [e.g., Norvig 1987], using different decrements for each link type, and inhibiting propagation below a threshold. Another method, used by Hendler [1986], ensures that a more constant number of nodes is searched by making the propagated activation inversely proportional to branching factor. However, the search regions are still determined using number of links and branching factor as a rough approximation of semantic distance.

Several ways of managing over-propagation schemes have been proposed. Charniak [1986] allows activation to decay exponentially over time, to remove old markers. Anti-promiscuity rules [Charniak 1986b] do not permit propagation in the case of concepts with branching factor above some threshold, with the rationale that such concepts are too general to provide much evidence anyway. Other systems assume that resource limitations can be overcome with massive parallelism, and perform exhaustive marker propagation assuming one processor for every concept [e.g., Fahlman 1979], like in local connectionist models. This should be a last resort, however, as the number of concepts is quite large for real domains. Finally, Alshawi [1987] suggests an indexing scheme that keeps track of "clusters" of markers to improve search times for large numbers of markers; again, however, no effort is made to reduce the number of markers.

2.2 Sensitivity to notational variants

Semantic nets are subject to notational variation, because the same propositional information can be represented in many ways (e.g., intermediate abstraction levels can be introduced arbitrarily). The effect of notational variants is to alter the network's indexing. However, search should not be affected arbitrarily by changes in indexing. In a sense, connectionist and propositional models represent extremes in sensitivity to network structure, and semantic nets should fall somewhere in between. Connectionist models rely entirely upon network structure to direct search. Marker-passing search is also guided by network structure, but the structure should be regarded as providing heuristic rather than conclusive indexing. Like propositional models, this frees the investigator from the connectionist requirement that concepts be defined by a comprehensive set of indexes to related concepts, by providing a language for abstract concept definitions, with a uniform semantics for notational variants. The extra degree of freedom must be accompanied by a search strategy that minimizes sensitivity to the notational variants. As far as we know, little if any work has addressed this issue, except by performing exhaustive search.

2.3 Usage of activation

Some models employ numeric activation levels on markers. There are two kinds of usage: controlling propagation (as in §2.1 above), and measuring belief. The latter usage equates activation with certainty. Alshawi [1987] uses a set of "context factors" that influence activation in different ways, in effect giving functions for combining evidence sources (for language interpretation). Charniak [1986] has a weak form of belief, by computing a "path strength" from the activations of markers on a path, such that the inference mechanism only considers paths with strengths of the highest order of magnitude.

While both usages seem appropriate at the intuitive level, it is not feasible to empirically verify whether particular methods work for large conceptual domains, without a complete theory of acquisition. Neither do the methods of computing activation appear to be based on probabilistic models that would provide better motivation. The distinction between the two usages is sometimes blurred. Moreover, belief measures tend not to be well-defined, and several authors have pointed out the need to formalize such measures [Cheeseman 1985, Pearl 1988]. Charniak [1986] also suggests that prior probabilities (not activation) should be used to define belief measures.

3 Formal Definitions

To reduce arbitrary propagation, what marker propagation represents must be defined more precisely than in existing models. Our analysis proceeds as follows: First, what a marker represents is defined, followed by what propagation and intersection represent. The next section examines desirable propagation characteristics. Finally, an implementation strategy is discussed.

The approach differs from path-based formalizations, which characterize the semantics of paths rather than markers [e.g., CharniaJk 1986, Norvig 1987]. Path semantics assume a path has already been found, and do not help constrain propagation.

To maximize the scope of our analysis, we assume a fairly standard semantic net: a simplified KL-ONE style net with conceptual IS-A and subpart hierarchies used as a term-defining language.[3] The representation is compatible with or translatable into most existing models. A partial semantics is given in Appendix A.

3.1 Caveat

To analyze propagation, we formalize concepts as predicates that apply to *occurrences,* rather than predicates that apply to individuals. Concepts *occur* when they help construe input evidence. For example, *X* in *light-bulb(X)* represents an occurrence of the generic light bulb concept, and *Y* in *light-bulb-1{Y)* represents an occurrence of an individual light bulb concept.[4] The reason is that we are interested in determining when concepts are worth searching, i.e., likely to help construe the input. Thus, we prefer a uniform notation for describing occurrences of concepts, regardless of whether they are generic or individual since either type may participate in the construal. If instead, we allowed concepts to predicate individuals, as in *light-bulb(LIGHT-BULB-l),* we would need the explicit (meta-) predicates

Generic-Occurrtnce{X,light-bulb), and *indiv-occurrence(Y,LIGHT-BULB-J).*

3.2 Markers

Markers represent hypotheses being considered by the interpreter or inference mechanism. This is true in any marker-passing model, regardless of whether it is explicitly noted. A hypothesis has at least two components (which may be implicit or explicit): a proposition, and an estimate of the chance of intersection. Many models also include a belief measure.

The first component is a proposition about the occurrence of a concept. For example, a marker on the *gooseneck* concept in Figure 1 hypothesizes that $\exists i_0$ *gooseneck(io).* For concepts with conceptual subparts ("schemata"), an occurrence of the concept implies occurrence of its subparts; thus, a marker on *lamp* hypothesizes

$$h_1\text{: } (\exists i_1, i_2, i_3)\ lamp(i_1) \wedge table\text{-}base(i_2) \wedge light\text{-}bulb(i_3) \wedge base\text{-}part(i_1, i_2) \wedge lb\text{-}part(i_1, i_3).$$

The second part of a marker's hypothesis is an estimate of the chance that the marker lies on a propagation path that will intersect another path. (Thus, part of the hypothesis is implicitly represented by the location of the marker.) In many marker-passing implementations, the estimate is implicitly binary: the interpreter believes that intersection is likely or unlikely, and accordingly places or does not place a marker [e.g., Fahlman

*I.e., IS-A and subpart relations carry no assertional force.

[4]Thus, we have at least a tri-part ontological distinction between generics, individuals, and occurrences. This is the sort of distinction where connectionist models are hampered by the aforementioned variable binding problem.

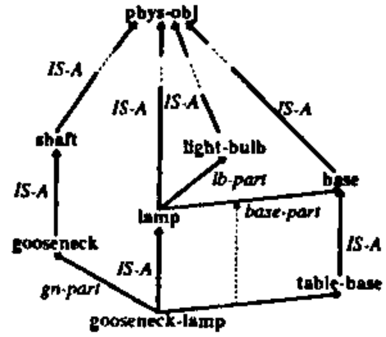

Figure 1: Example hierarchy.

1979, Hirst 1987, Norvig 1987]. To facilitate analyzing propagation, we shall allow estimates to be real numbers between zero and one, to denote degree of belief that intersection will occur. Deriving estimates is discussed in §4.

3.3 Propagation

Propagating a marker represents generating a hypothesis. The new hypothesis depends upon the source marker's hypothesis, and the semantics of the link propagated over. Transformation functions are given by example, for the classes of links in our model:

Downward IS-A: Propagating down an IS-A link specializes the hypothesis. For example, propagating a marker from *lamp* to *gooseneck-lamp* transforms h_1 into

$$h_2\text{: } gooseneck\text{-}lamp(i_1) \wedge table\text{-}base(i_2) \wedge light\text{-}bulb(i_3) \wedge gooseneck(i_4) \wedge base\text{-}part(i_1, i_2) \wedge lb\text{-}part(i_1, i_3) \wedge gn\text{-}part(i_1, i_4).^5$$

We write $h_1 \blacktriangleright\!\blacktriangleright h_2$ to indicate the propagation path, which is important because the variables in the two hypotheses are not independent. The dependency arises from predicating the same concept occurrences; i.e., h_2 hypothesizes that the occurrence in h_1 of the lamp concept is specifically an occurrence of the gooseneck lamp concept, so h_2 entails h_1. Because propagation is against the direction of implication, Charniak [1986] calls this an abductive assumption.

Upward IS-A: Propagating up an IS-A link abstracts the hypothesis. The transformation is the reverse of downward transformation, and the same dependency holds.

Downward subpart: Propagating across a conceptual subpart link from part to whole also specializes the hypothesis. For example, propagating a marker from *light-bulb* to *lamp* transforms

$$h_0\text{: } light\text{-}bulb(i_3)$$

into h_1 above. Again the dependent variables mean that h_1 implies h_o. Thus, h_0 is also an abductive assumption; this is why we call propagation from part to whole "downward".

Upward subpart: Abstracts the hypothesis, with the same dependency.

For semantic nets with associative relations, i.e., relations between concepts that are not conceptual part-wholes, propagation involves similar transformation of

[6]Henceforth, the existential will be left out.

hypotheses. However, the dependency between the hypotheses cannot be expressed clearly. For this reason, we treat an associative relation between two concepts as an abbreviation for a third concept that acts like a "frame" with the two concepts as subparts.[6] The propagation rules above then hold.

3.4 Intersection

An intersection represents a meta-hypothesis about performing unification. Intersection occurs when two markers propagated from different origins are placed on the same concept. For example, suppose the path h_o »-> h_1-«► h_2 intersects at *gooseneck-lamp* with another path h_3 -> h_4:

$$\begin{aligned}
&h_3\text{: } gooseneck(i_8) \\
&h_4\text{: } gooseneck\text{-}lamp(i_5) \wedge table\text{-}base(i_6) \wedge light\text{-}bulb(i_7) \\
&\quad \wedge gooseneck(i_8) \wedge base\text{-}part(i_5, i_6) \wedge lb\text{-}part(i_5, i_7) \\
&\quad \wedge gn\text{-}part(i_5, i_8)
\end{aligned}$$

The intersection suggests unifying $i_1 = i_5$, $i_2 = i_6$, etc. Unification, if performed, represents collapsing two occurrences of a concept.

4 Desirable Propagation Characteristics

Having derived the above definitions from the semantics of the representation, we would like to optimize propagation. We make the assumption that the interpreter has no information about the connectivity of the semantic net prior to searching because (1) it should be as insensitive as possible to notational variants, and (2) any acquisition strategy for concept formation dynamically changes a semantic net's connectivity. Given this constraint, the only information that can be utilized to guide propagation is the information derivable from knowledge stored at the concept nodes where markers already are. Consequently, propagation for each origin marker must be optimized independently of other origin markers (until intersection).

4.1 Deductive propagation

Deductive (upward) propagation cannot be restricted. When an origin or non-origin marker is created, upward propagation to all its ancestors in the IS-A and subpart hierarchies must follow immediately. The reason is that the occurence of a concept also implies the occurence of all concepts standing for more primitive feature/role combinations (Appendix A). Intersections through any of these ancestral levels must be equally detectable; in the absence of connectivity information, all ancestors must be simultaneously marked. Immediate complete upward propagation is also employed by Martin and Riesbeck [1986].

Deductive propagation is not as expensive as it may appear, since it applies only to term-defining IS-A and subpart relations. Searching assertional IS-A and subpart relations can be treated as a special case of abductive propagation.[7]

4.2 Utility maximization for abductive propagation

To regulate abductive (downward) propagation, our approach is to ensure that propagation occurs in the order that, given the information possessed by the interpreter at any point in time, the next propagation maximizes the chance of intersection. Under the unknown-connectivity assumption, the best estimate that can be made for the likelihood of intersection involving a given marker is proportional to the posterior probability of the marker's proposition. The reason is as follows: For any single origin marker, the interpreter does not know which links connect the concept to more nodes, by the connectivity assumption. Using the maximum entropy assumption, we assume an equal distribution of nodes for all links. Given this, intersection chances are maximized by choosing the propagation that creates the marker with the most probable proposition, given the input evidence.

This does *not* guarantee maximizing the objective probability of finding an intersection, i.e., the probability that is defined as the frequency of finding intersections relative to the frequency of propagations, anywhere in the net, and involving any origin marker. Maximizing objective probability is not possible when there is missing information, as with the connectivity assumption. The approach does however maximize subjective probability under the connectivity assumption, i.e., the probability that is computed by assigning uniform distributions where unknown, following maximum entropy.[8]

4.3 Marker probabilities

The "probability" of a marker's proposition really refers to the posterior probability given all the input evidence. When an origin marker with proposition h_o is created in response to an input, its posterior probability $p(h_o)$ reflects the degree of confirmation provided by the input evidence C_o. (With reliable evidence indicators, $p(h_0)$ will essentially be 1.)

For the abductive case (where a marker with proposition h_1 derives from origin h_o such that h_1 entails h_o), the posterior probability is

$$p(h_1) = P(h_1|h_0)p(h_0).$$

Posterior probabilities for the deductive case require explicit normalization factors.[9]

4.4 Necessity of frequencies

To compute the conditional probability term, the frequency of occurrence for each concept must be known. Assuming that $f(cn_i)$ is the frequency of occurrence of the concept that the marker h_i is on,

$$p(h_1) = [f(cn_1)/f(cn_0)]p(h_0).$$

Requiring frequency information is equivalent to requiring weighted IS-A or subpart links, because it is the rel-

[6]An example is the assertional IS-A in Appendix A.

[7]Minimizing the deductive IS-A and subpart closures is an efficiency argument for distinguishing terminological and assertional uses. See the semantics of assertional IS-A in Appendix A.

[8]Discussions of objective vs. subjective probabilities can be found in Walpole and Myers [1978] and Cheeseman [1985].

[9]Pearl [1988] discusses normalization in simple taxonomic belief hierarchies.

ative frequency ratios that must be derivable. The same information could thus be stored as ratios on links between concepts.

Frequencies are the only assertional knowledge in our representation, i.e., they assert how often concepts occur in dealing with the external world (everything else up to this point is used for defining terms). Theoretically, frequencies should be acquired by counting whenever a concept occurs in a construal.

4.5 Insensitivity to notational variants

Deductive propagation is insensitive to intermediate levels of abstraction since complete upward propagation occurs at once. For abductive propagation, sensitivity to notational variants is minimized by tying the order of propagation to the frequency of concept occurrence (rather than number of links). This ensures the same order of propagation for notational variants, except that extra intermediate levels may be inserted.

4.6 Effect of intersection

If an intersection occurs and results in unification, then the two colliding markers are merged, and the posterior probabilities of all hypotheses involving the unified concept occurrences must be revised, by taking the combined input evidence from both origins as support for the hypotheses. Prior to intersection, hypotheses derived from the two origins are treated as being independent; however, when the variables in the hypotheses are unified, the hypotheses become dependent. That is, given input evidence e_0 and e_1 for origins fto and ft], $p(h_1) = P(h_i \backslash e_0, e_1)$ should be computed for each affected hypothesis h_i, because it is no longer assumed that $P\{h_i \backslash e_o, e_1) = P(hi \backslash eo)$. Propagation proceeds from the merged marker as before.

4.7 Conflating utility and belief

If a marker's posterior probability is also used as a belief measure—a reasonable first approximation—then the two usages of activation can be conflated in one numeric measure. Roughly speaking, the inferences that should be made are the subset of hypotheses after some period of propagation that end up with the highest posterior probabilities, say, above a threshold of 0.5. No evaluation stage is required, as a result of constraining propagation to discover the most probable concepts first.

Things become more complicated upon closer analysis. Some of the issues are: (1) When should the search cease? (2) How can posterior probabilities be efficiently normalized as evidence accumulates? (3) How can search be biased to "commit", i.e., to "over-estimate" the more likely prior probabilities, and "under-estimate" less likely ones, as a default effect? While we have no conclusive answer to these questions, a strategy we are pursuing is outlined below.

5 Implementation Strategy

If the last three points are not considered, then a sequential implementation of most-probable-first search is straightforward. However, because of those points, and in order to exploit parallelism, an extension to the model is proposed. There are two obvious ways to parallelize marker passing: processor-per-concept and processor-per-marker. Having discussed how to localize the regions searched, we adopt the processor-per-marker approach, to reduce processor requirements. (This is equivalent to cost-per-marker under sequential simulation.) Furthermore, we assume that markers propagate autonomously at variable speeds, and thereby implement most-probable-first search.

The first step is to ensure that the basic formulation is compatible with the analysis above. To do this, we use an imaginary model, where for each origin marker h_o, each potential abductive marker h,, and each time point t,, the value of c(h_o, *h, t)* is a real number between zero and one. For this model to fit the analysis, the function *c* should have these characteristics:

1. **For any h_0 and h, the function begins at $c(h_0, h, 0) = 0$ and rises to an asymptote such that $\lim_{t \to \infty} c(h_0, h, t) = P(h|h_0)p(h_0)$.**
2. **For any threshold $0 < thr < 1$, the c functions should cross thr in most-probable-first order. The simplest such function is (see Figure 2a): $c(h_0, h, t) = P(h|h_0)p(h_0)(1 - e^{-t})$.**

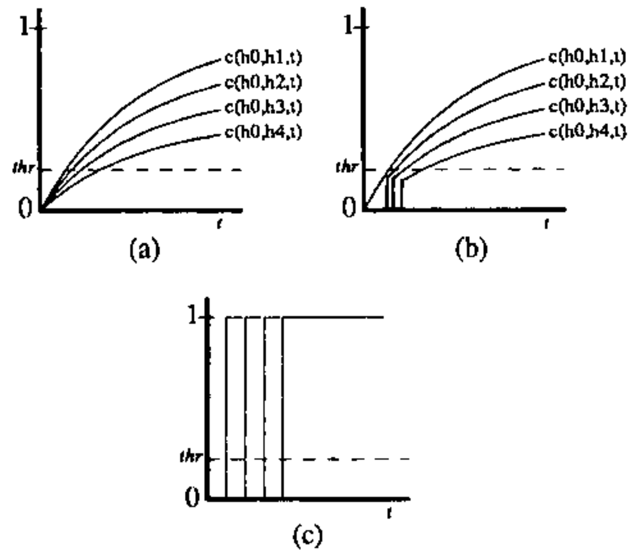

Figure 2: Functions assuming $h_4 \to h_3 \to h_2 \to h_1$.

By restricting propagation such that a hypothesis *h* does not propagate unless c(h_o, *h, t) > thr,* we get the approximation shown in Figure 2b. The model does not find intersections below threshold but is otherwise accurately most-probable-first. For comparison, the way that discrete binary marker-passing models propagate is shown in Figure 2c.

The advantage of formulating the model this way is that it can be implemented using a technique that fits our desired propagation characteristics, and also appears to be suited to handling the three issues brought up at the end of the previous section. In this technique, the dependencies between hypotheses are used to link the markers, forming a new net, which instantiates parts of the semantic net's structure. The marker net is an augmented relaxation network where the activation of each marker represents $c(h_0, h, t)$. The following incremental algorithm is used:

1. Update the activation of each marker as a function of its previous activation and the activations of its immediate neighbors.
2. Discard the markers with very low activation.
3. Augment the marker net by propagating markers above threshold.
4. Repeat.

Figure 3: Augmented relaxation network algorithm.

Non-linear activation functions can be designed to give the curves in Figure 2b. (In the case where an intersection results in unification, the activation function should give similar curves, modified to reflect evidence combination.) Moreover, the model suggests possible approaches to the more difficult issues: (1) Cease searching when relaxation settles, i.e., when the activation no longer changes significantly with each iteration. Hypotheses below threshold are not worth searching. (2) Attach inhibitory normalization links to markers with no ancestors, so as to normalize the activations of all descendants; introduce inhibitory disjointness links between sibling markers with mutually exclusive hypotheses, such that the sum of their activations is not permitted to exceed that of their common parent. (3) Bias the activation functions to distort the belief/utility settling points so as to "over-commit" above a threshold probability and "under-commit" below it.

6 Discussion

We have suggested how to restrict marker propagation more accurately than in existing models, based upon a probabilistic analysis, and we have explained the usage of activation more precisely. It remains to be shown whether the assumptions made under our analysis justify the suggested extensions.

The augmented relaxation network is similar in spirit to dynamic connectionist nets [e.g., Berg 1987]. Chun and Mimo [1987] have also suggested combining marker passing and relaxation. However, in our model more emphasis is placed upon maintaining the symbolic properties of semantic networks than in others, since we feel that ease of manipulating the knowledge representation is one of the advantages of symbolic AI. Hendler [1987] suggests combining marker passing with microfeatures, using a "defining characteristic" link, but does not use a term-defining hierarchy to represent microfeature combinations at varying levels of abstraction.

A comparison with "pure" recognition models illustrates the hybrid advantage. Hobbs et al. [1988] cast a number of natural language interpretation problems as abductive inference in a theorem-prover, but noted that some way to control the potentially explosive search is necessary. As symbolic models make increasing use of parallel techniques for efficient computation, connectionist work has moved toward addressing structured representations, traditionally a symbolic strength [Feldman 1986]. Shastri [1988] presents a connectionist model that handles the single-schema subcase of the recognition problem using a hierarchical representation, but still lacks the means to handle variable binding.

An open question is how to decide whether to perform unification when an intersection occurs. Moreover, multiple intersections may suggest mutually exclusive unifications, so discrimination criteria are needed.

Another issue involves relaxing the assumption that the interpreter has no *a priori* information about the semantic net's connectivity. In this case, a global optimization strategy might be used to restructure the network or provide additional indexing weights to improve the worst case search cost.

An implementation called FRESCO is under development [Wu 1987]. It is being applied to analyzing noun compounds such as *gooseneck lamp,* with parallel parsing and semantic interpretation.

Acknowledgements

Thanks to Robert Wilensky and Michael Braverman for valuable discussions, as well as Jerry Feldman, Peter Norvig, Jim Hendler, Joachim Diederich, the members of BAIR (Berkeley AI Research), and the AI/Cognition group at TU Munich.

A Representation Semantics

Since knowledge representation terminology is often confusing, the semantics of the assumed representation is sketched here:

1. A concept *cn* is a one-place predicate (with an associated frequency of occurrence $0 < f(cn) < 1$).
2. An occurrence / of the concept is a constant such that *cn(l)* holds.
3. A (terminological) IS-A relation between two concepts cn_1 and cn_2 means (Vx)[cni(z)—>cn_2(t)].
4. A primitive conceptual subpart relation between two concepts cn_ and cn_2 means (Vi)[cni(i) —► $(\exists j)[cn_2(j) \wedge rl(i,j)]]$, where *rl* is a two-place predicate.
5. A defined conceptual subpart relation between two concepts cn_1 and cn_2 must correspond to another subpart relation *s* between two concepts cn_3 and cn_4, such that cn_1 IS-A cn_3, and cn_2 IS-A cn_4 it means $(\forall i)[cn_1(i) \rightarrow (\exists j)[cn_2(j) \wedge rl_{12}(i,j) \wedge rl_{34}(i,j)]]$, where r/12 is a two-place predicate, and r/34 is the two-place predicate in *s.*

Additional constraints for transitivity and disjointness have been left out, but are assumed; also, a more general form of quantification is lacking. Assertional IS-A is a non-primitive associative relation:

- An assertional IS-A between two concepts cn_1 and cn_2 is itself a concept cn_{12} with defined subpart relations to cn_1 and cn_2. The ratio $f'(cn_{12})/f(cn_1)$ indicates the salience of cn_2 to cn_1, and f $(cn_{12})/f'(c_{n2})$ the salience of cn_1 to cn_2.

References

Alshawi, Hiyan. [1987]. Memory and Context for Language Interpretation. Cambridge: Cambridge University Press.

Berg, George. [1987]. "A parallel natural language processing architecture with distributed control", Cog. ScL-87.

Charniak, Eugene. [1986]. "A neat theory of marker passing", Proc. AAAI-86, 584-588.

Charniak, Eugene. [1986b]. A Single-Semantic-Process Theory of Parsing, Tech. Rep., Dept. of Comp. Sci., Brown University.

Cheeseman, Peter. [1985]. "In defense of probability", Proc. IJCAI-85, 1002-1009.

Chun, Hon Wai, and Alejandro Mimo. [1987]. "A model of schema selection using marker passing and connectionist spreading activation", Cog. Sci.-87, 887-896.

Fahlman, Scott E. [1979]. NETL: A System for Representing and Using Real-World Knowledge. Cambridge, Massachusetts: MIT Press.

Feldman, Jerome A. [1986]. Neural Representation of Conceptual Knowledge. TR189, Dept. of Comp. Sci., Univ. of Rochester, New York.

Hendler, James A. [1986]. Integrating marker-passing and problem-solving: A spreading activation approach to improved choice in planning. Ph.D. thesis, Brown University.

Hendler, James A. [1987]. "Marker-passing and microfeatures", Proc. IJCAI-S7, 151-154.

Hinton, G. E., J. L. McClelland, and D. E. Rumelhart. [1986]. "Distributed representations", in D. E. Rumelhart and J. L. McClelland (eds.), Parallel Distributed Processing (Vol. 1). Cambridge, Massachusetts: MIT Press. 77-109.

Hirst, Graeme. [1987], Semantic Interpretation and the Resolution of Ambiguity. Cambridge: Cambridge University Press.

Hobbs, Jerry R., Mark Stickel, Paul Martin, and Douglas Edwards. [1988]. "Interpretation as abduction", Proc. 26th ACL, 95-103.

Mark, William. [1982]. "Realization", in J. G. Schmolze and R. J. Brachman (eds.), Proc. 1981 KL-ONE Workshop, 76-87.

Martin, Charles E., and Christopher K. Riesbeck. [1986]. "Uniform parsing and inferencing for learning", Proc. AAAI-86, 257-261.

McClelland, James L., and David E. Rumelhart. [1981]. "An interactive activation model of context effects in letter perception: Part 1", Psychological Review 88: 375-401.

Norvig, Peter. [1987]. A Unified Theory of Inference for Text Understanding (Ph.D. thesis). Report No. UCB/CSD 87/339, C. S. Div., University of California, Berkeley.

Pearl, Judea. [1988]. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, California: Morgan Kaufman.

Shastri, Lokendra. [1988]. "A Connectionist Approach to Knowledge Representation and Limited Inference", Cognitive Science 12: 331-392.

Schmolze, J. G., and T. A. Lipkis. [1983]. "Classification in the KL-ONE knowledge representation system", Proc. IJCAI-83, 330-332.

Walpole, R. E., and R. H. Myers. [1978]. Probability and Statistics for Engineers and Scientists (2nd ed.) New York: Macmillan.

Wilensky, Robert, David Chin, Marc Luria, James Martin, James Mayfield, and Dekai Wu. [1988]. "The Berkeley UNIX Consultant project", Computational Linguistics 14(4): 35-84.

Wu, Dekai. [1987]. "Concretion inferences in natural language understanding", Proc. GWAI-87, 11th German Workshop on Artificial Intelligence. Berlin: Springer-Verlag. 74-83.