Toward Connectionist Parsing

Steven Small, Gary Cottrell, and Lokendra Shastri
Department of Computer Science
The University of Rochester
Rochester, New York 14627

Abstract

The parsing of natural language is the product of dense interactions among various comprehension processes. We believe that traditional models have greatly underestimated the richness of these interactions. We propose a model for low-level parsing which is massively parallel, highly distributed, and highly connected. The model suggests a solution to the problem of word sense disambiguation which is psychologically plausible and computationally feasible. The paper outlines the general connectionist paradigm followed by a brief description of a three-level network to do parsing. Finally we trace through an example to illustrate the functioning of the model.

Introduction

The parsing of natural language must be studied in the full context of all other comprehension processes that depend or are depended upon by the parsing process. While the study of these different processes has led to interesting and valuable results, we believe that the interactions among them are not as low in bandwidth as traditional models would imply. Our work involves trying to model the high connectedness of these aspects. The interpretation of metaphor and indirect speech acts, for example, and the processes of word sense disambiguation and anaphoric reference seem to require a significant application of general memory skills (i.e., "non-linguistic knowledge") in order to carry out. Further, it seems that any strictly serial decoupling of associated subtasks, i.e., one major subtask and then the next and so on, does not permit enough interdependence of function to explain adequately reasoning tasks that people find quite easy and natural to perform.

In this paper we sketch a model of natural language understanding which deals with the problem of word sense disambiguation in what we believe to be a psychologically plausible manner. Word sense ambiguity is an interesting problem because people do it easily, and the task requires knowledge from many sources. Small [1], for example, lists 57 different senses for the word "take," which depend on the local context of the following words. This is not an isolated phenomenon. In an informal study, Gentner [2] found that the 20 most frequent nouns have an average of 7.3 senses each; the 20 most frequent verbs have an average of 12.4 senses each. We believe that a model which tries to emulate how people understand language must have at its core a clean and efficient disambiguation mechanism. Previous work in this area by Wilks [3], Riesbeck and Schank [4], and Small and Rieger [5] have been adequate as high-level models, but we emphasize a processing structure which is closer to the neuronal hardware, as we believe these models will exhibit consequences not found in sequential, symbol-passing models. (For more discussion of this, see [6, 7, 8, 9, and 10].) Our model will use a uniform processing framework, be able to maintain multiple hypotheses in parallel, and switch interpretations easily based on subsequent context.

We base our model in part on recent studies in lexical access by Swinney [11] and Seidenberg et al. [12], which demonstrate that when people hear an ambiguous word, at least two senses of the word are initially active, but by 200 msec later, only the one appropriate to the context is still active. A cornerstone of our model is parallel access of all meanings followed by an interaction of these meanings with the context of the rest of the sentence. This will result in only one meaning remaining highly active.

Connectionist Models

We have in mind a particular new approach to the study of natural language comprehension. These are the massively parallel models of the sort currently under development by Feldman and Ballard [7], which they call *connectionist models.*

The basic element of the connectionist model is the *computing unit.* A unit is characterized by a continuous valued *potential* (or confidence) between -1 and 1, a vector of *inputs*, a small set of *states*, and a single *output*, which is an arbitrary (preferably simple) function of the inputs. Thus the firing of a unit may be based on logical combinations of input values. The connection pattern is prewired; only the *weights* on connections may change. Connections may be inhibitory or excitatory, giving the model much flexibility. A coalition of connected units which are mutually reinforcing corresponds to a percept; a *stable coalition* is one in which the overall excitation exceeds the overall inhibition. The fundamental premise of the connectionist framework is that individual units do not transmit large amounts of symbolic information, but compute by being appropriately connected to large numbers of similar units.

The technique used to encode information in the network is called the unit/value principle. That is, a unit represents a value of a parameter. A unit receives input from other units which provide positive or negative "evidence" for the value the unit represents. If it gets enough positive input, it will "fire." The output is a confidence measure which is transmitted to all connected units.

Connectionist Parsing

We propose a three-level network to represent the parsing system, shown in Figure 1.

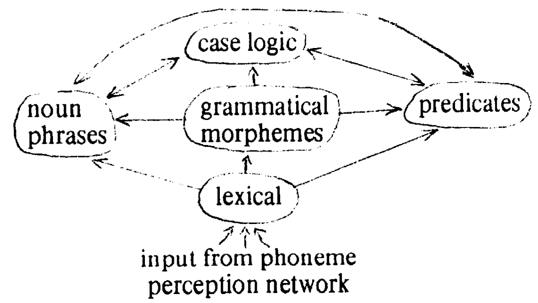


Figure 1. Proposed Parsing Network.

We assume input to the lexical level comes from a word perception network such as that described in McClelland and Rumelhart [8]. Note that we do not have an explicit syntax level. Syntax information is embedded in every level of the network in various ways. An overview of the function of each level follows. More detail is given in [6].

The Lexical Level. This is the input level of our network. We do not try to account for phoneme effects at this stage of the research. The units at this level roughly represent morphemes, although we do not strictly follow the linguistic definition. Thus we represent hearing or reading a sentence such as "John loves Mary" by activating the "John," "loves," and "Mary" units sequentially, with a model-dependent delay between them.

The Word Sense Level. Units on the lexical level are connected to units representing their different senses at this level. For example, "cart" at the lexical level is connected to WAGON (the noun sense) and TRANSPORT (the verb sense) at this level. We represent idioms by conjunctive connections; that is, "throw" and "up" would conjunctively connect to VOMIT--both would have to be active for the VOMIT node to become active. We conceive of this level as divided into three subnetworks, the noun phrase or "object reference" network, the predicate or "relational" network, and the function word network. We separate the first two for reasons given in [2], and because they play a separate role in connecting to the case logic level. The third subnetwork arises from the role of function words as linguistic operators, and the (controversial) evidence that particular lesions may impede their use.

The Case Logic Level. This level expresses the possible relationships (bindings) between predicates and objects. We posit an *exploded case* representation; that is, we use a large number of case roles that are more specific than AGENT, OBJECT, etc., but fall into those classes. These nodes thus represent some typing information on case roles for predicates. A predicate at the word sense level excites its case nodes at this level, with a sequencing mechanism increasing the activity of unfilled case nodes as others are filled (thus setting up expectations).

One central goal of our research is to construct an experimentally testable theory of the organization of these levels. In general, the design principles of these spreading activation networks must evolve through the data of psychological experimentation and computational simulation. We acknowledge that the networks illustrated in the next section are at an early stage in this evolution.

A Simple Example Analysis

Let us now work through the analysis of an example sentence, showing how the final interpretation comes to be decided upon despite a sequence of (wrong) intermediate hypotheses. Consider the following sentences:

(a) "A man threw up a ball."

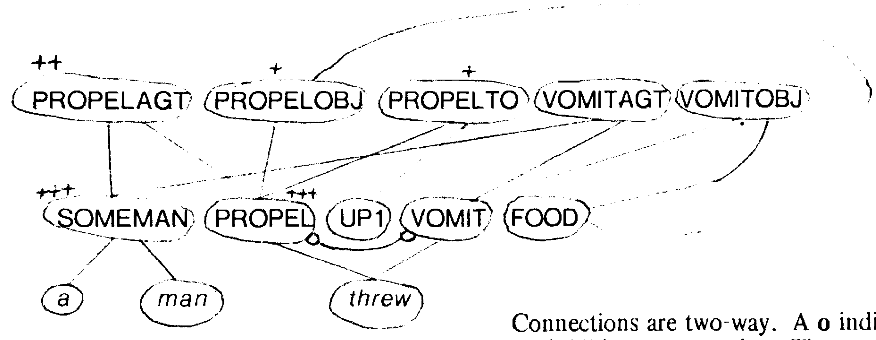
(b) "A man threw up dinner."

We shall illustrate the analysis of sentence (a), while explicitly taking into consideration the possibility of interpretation (b) along the way. One thing that we shall not do in this section is to explain the role of context (other than intra-sentential) in the process; while the existing state of the connection network has activity corresponding to the analyses of previous perceptual inputs (including language), that will not come into play in this example.

As the phrase "a man" is heard, the appropriate processing units become activated at the lexical level, causing the activation of unit SOMEMAN at the word sense level. Note that activation does not only flow from lower levels up, but that feedback activation flows down as well, forming a stable coalition among the units for "a", "man", and SOMEMAN. In addition, there is inhibition between SOMEMAN and SOMEWOMAN, representing the notion that a person cannot be both.

Next, driven from below, the lexical unit for "threw" becomes active, exciting the units on the word sense level that represent its possible meanings. This includes the unit PROPEL in the predicate subnetwork, as shown in Figure 2. The VOMIT node is not yet active, since both "threw" and "up" must be active for VOMIT to exceed threshold.

Activation then spreads from the PROPEL unit to units that represent the conceptual cases of PROPEL, which we call PROPELAGT, PROPELOBJ, PROPELFROM, and PROPELTO, and in particular, a stable coalition forms including PROPEL, PROPELAGT, SOMEMAN, "man," and "threw." We ignore in this example the kinds of connections and auxiliary units required to make this work. Assume for the moment that we have managed to build the connection network in such a way that it



Connections are two-way. A **o** indicates an inhibitory connection. The + signs indicate relative levels of activation.

Figure 2. Network Activation after "A man threw".

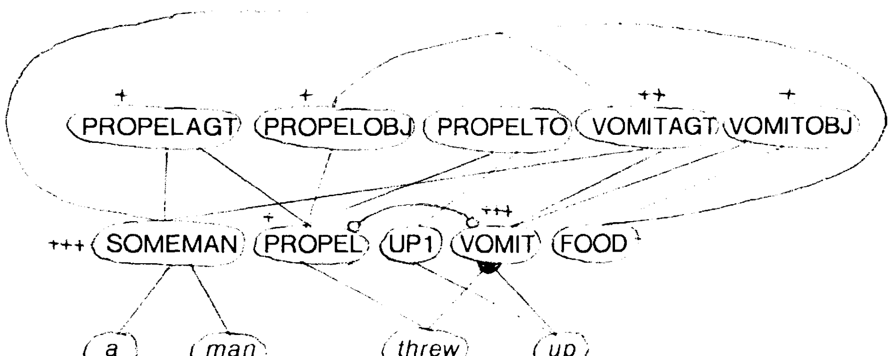
manifests the desired stable coalition of activity.

The next piece of input forces the connection network to adjust to activity that has an inhibitory effect on the existing stable coalition. The activity of the node "up" causes the VOMIT and UP1 units to become active. The units of the model must readjust their potentials in the face of this new data--the continuous output values of the affected units cause a decrease in the confidence of PROPEL and an increase in that for VOMIT. Since the phrase "threw up" most usually denotes vomiting, the VOMIT unit has a higher activation at this stage than does the one for PROPEL (see Figure 3). As before, associated case units VOMITAGT and VOMITOBJ also become active, and a new stable coalition dominates the local (i.e., language processing) network behavior.

Finally, the listener perceives the phrase "a ball", which reinforces the object case node PROPELOBJ and inhibits the analogous VOMITOBJ, resulting in a correct interpretation of the example sentence. This activity takes place through the SOMEBALL node, which inhibits VOMITOBJ and excites PROPELOBJ. PROPEL thus gets more top-down reinforcement from its filled case nodes than VOMIT does. If the word "dinner" had been heard instead, this same behavior, mediated this time by the FOOD word sense node, would have excited the VOMITOBJ case node and led to the other interpretation.

Conclusion

We have constructed a simulator for connection networks and early results have guided the design presented here. We are building a network containing 20-30 polysemic words; our current goal involves making it respond appropriately to different combinations of them. We expect this to lead to the development of rules for generating connection patterns upon the introduction of new words into the network. Future research will



The nodes for *threw* and *up* are connected to the VOMIT node by a conjunctive connection.

Figure 3. Network Activation after "A man threw up".

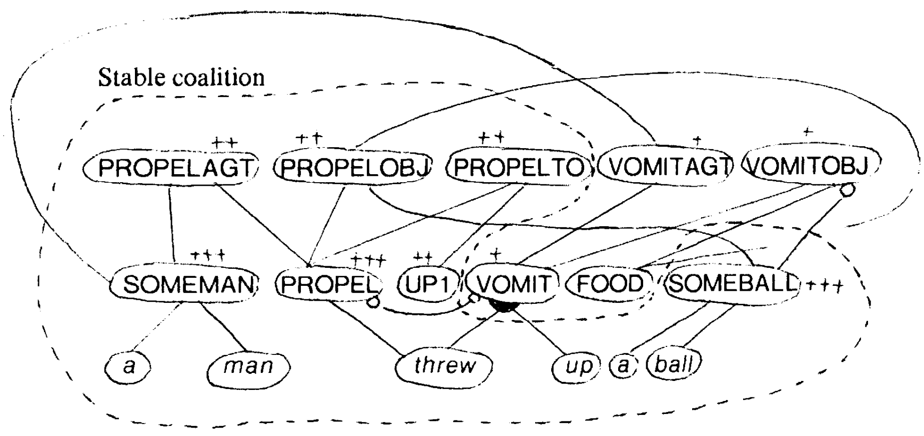


Figure 4. Network Activation after "A man threw up a ball".

investigate the problems of reference, focus of attention, word sense induction, and the structure of higher-level networks involving long-term storage.

Acknowledgments

This work was supported in part by the National Aeronautics and Space Administration under Grant NSG 7253 and in part by the Office of Naval Research under Grant N00014-82-K-0193. Their support of this basic research is gratefully acknowledged.

References

[1] Small, Steven L. (1978), "Conceptual Language Analysis for Story Comprehension", TR 663, U. Maryland.

[2] Gentner, D. (1982), "Some Interesting Differences Between Nouns and Verbs", *Cognition and Brain Theory 4*, 2, 161-177.

[3] Wilks, Y., (1976), "Parsing English II", in *Computational Semantics*, Charniak and Wilks (Eds). North-Holland, 155-184.

[4] Riesbeck, Christopher K., and Roger C. Schank (1976), "Comprehension by Computer: Expectation-Based Analysis of Sentences in Context," Research Report #78, Department of Computer Science, Yale University.

[5] Small, Steven L. and Chuck Rieger (1982), "Parsing and Comprehending with Word Experts (A Theory and its Realization)", in *Strategies for Natural Language Processing*, Ringle and Lehnert (Eds.). Lawrence Erlbaum Associates (to appear).

[6] Cottrell, Gary (1982), "Connectionist Parsing", Technical Report, Department of Computer Science, U. Rochester (to appear).

[7] Feldman, Jerome A. and Dana H. Ballard (1982), "Connectionist Models and their Properties," *Cognitive Science*, vol. 6 (to appear).

[8] McClelland, James L. and David E. Rumelhart (1980), "An Interactive Activation Model of the Effect of Context in Perception: Part I", Report #8002, Center for Human Information Processing, U. California at San Diego.

[9] Hinton, Geoffrey E. (1981), "Shape Representation in Parallel Systems," *Proc., Seventh IJCAI*, 1088-1096, Vancouver, B.C.

[10] Gigley, Helen M. (1982), "A Computational Neurolinguistic Approach to Processing Models of Sentence Comprehension", COINS TR 82-9, University of Massachusetts.

[11] Swinney, David A. (1979), "Lexical Access during Sentence Comprehension: (Re)Consideration of Context Effects", *Journal of Verbal Learning and Verbal Behavior*, vol. 18, 645-660.

[12] Seidenberg, M.S., M.K. Tanenhaus, J.M. Leiman, and M. Bienkowski, "Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing," *Cognitive Psychology*, in press.