WORD EXPERT PARSING

Steven L. Small
Department of Computer Science
University of Maryland
College Park, Maryland 20742

This paper describes an approach to conceptual analysis and understanding of natural language in which linguistic knowledge centers on individual words, and the analysis mechanisms consist of interactions among distributed procedural experts representing that knowledge. Each word expert models the process of diagnosing the intended usage of a particular word in context. The Word Expert Parser performs conceptual analysis through the interactions of the individual experts, which ask questions and evaluate responses in a single mutually acceptable sense meaning. The Word Expert theory is advanced as a better cognitive model of natural language understanding than the traditional rule-based approaches. The Word Expert Parser models parts of the theory, and the important issues of control and representation that arise in developing such a model form the basis of this technical discussion. An example from the prototype LISP implementation helps explain the theoretical results presented.

1. Introduction

Computational understanding of natural language requires complex interactions among a variety of distinct yet redundant mechanisms. The construction of a computer program to perform such a task begins with the development of an organizational framework which inherently incorporates certain assumptions about the nature of these processes and the environment in which they take place. Such cognitive premises affect professional scope and substance of computational analysis for comprehension as found in the program.

This paper describes a theory of conceptual parsing which considers knowledge about language to be distributed across a collection of procedural experts centered on individual words. Natural language parsing with word experts entails several new hypotheses about the organization and representation of linguistic knowledge for computational language comprehension. The Word Expert Parser [1] explores how the word expert view, coupled with certain other choices based on previous work, affect structure and process in a cognitive model of parsing.

The Word Expert Parser is a cognitive model of conceptual analysis in which the unit of linguistic knowledge is the word and the form of research is the set of processes underlying comprehension. The model is aimed directly at problems of word sense ambiguity and idiomatic expressions, and in greatly generalizing the notion of word sense, promotes those issues to a central place in the study of language parsing. Parsing models typically cope unsatisfactorily with the wide heterogeneity of usages of particular words. If a sentence contains a standard form of a word, it can usually be parsed; if it involves a less prevalent form which has a different part of speech, perhaps a more complex one, it may not be. Distinguishing among the many senses of a common verb, adjective, or pronoun, for example, or correctly translating idioms are rarely possible.

At the source of this difficulty is the reliance on rule-based formalisms, whether syntactic or semantic (e.g., cases), which attempt to capture the linguistic contributions inherent in constituent chunks of sentences that consist of more than single words. A crucial assumption underlying work on the Word Expert Parser is that the fundamental unit of linguistic knowledge is the word, and that understanding its sense or role in a particular context is the central parsing process. In the parser to be described, the word expert constitutes the kernel of linguistic knowledge and its representation in the expert's knowledge base. The representation is procedural in nature and executes directly as a process, cooperating with the other experts for a given sentence to arrive at a mutually acceptable sentence meaning.

Certain principles behind the parser do not follow directly from the view or even from other recent theories of parsing. The cognitive processes involved in language comprehension comprise the focus of linguistic study of the word expert approach. Parsing is viewed as an inferential process where linguistic knowledge of syntax and semantics and general pragmatic knowledge are applied in a uniform manner during interpretation. This methodological position closely follows that of Brown (see [2] and [3]) and Schank [4], the central concern with word usage and word sense ambiguity follows similar motivations of Wilks [5]. The central structure of the Word Expert Parser results from agreement with the hypothesis of Marcus that parsing can not be performed deterministically, and in ways in which information gained through interpretation is permanent [6]. Rieger's view of inference as intelligent selection among a number of competing plausible alternatives [7] of course forms the cornerstone of the new theory. His ideas on word sense selection for language analysis [8] and [9] and strategy selection for general problem solving [10] constitute a consistent cognitive perspective.

Any natural language understanding system must incorporate mechanisms to perform word sense disambiguation in the context of open-ended world knowledge. The important assumptions for word usage diagnosis derive from the ubiquity of local ambiguities, and brought about the notion that they be made the central processes of computational analysis and understanding. Consideration of almost any English content word leads to a realization of the scope of the problem — with a little time and perhaps help from the dictionary, many distinct usages can be identified. As a simple illustration, several usages for the words "heavy" and "lie" appear in Figure 1. Each of these seemingly benign words exhibits a rich depth of contextual use. An earlier paper contains a list of almost sixty verbal usages for the word "take" [11].

The representation of all contextual word usages in an active way that insures their utility for linguistic diagnosis led to the notion of word expert. Each expert is a procedural entity composed of all possible contextual interpretations of the word it represents. When placed in a context formed by experts for other words in a sentence, each expert should be capable of sufficient context-parsing, and self-examination to determine successfully its functional or semantic role, and further, to realize the nature of that function or the precise meaning of the word. The representation and control issues involved in basing a parser on word experts are discussed below, following presentation of an example execution of the existing Word Expert Parser.

2. Model Overview

The Word Expert Parser successfully parses the sentence

"The deep philosopher throws the peach pit into the deep pit."

through cooperation among the appropriate word experts. Initialization of the parser consists of retrieving the experts for "the", "deep", "philosopher", "throw", and "pit."

An important assumption of the word expert viewpoint is that the set of word usages is not only finite, but fairly small as well.

The perspective of viewing language through lexical contributions to structure and meaning has naturally led to the development of word experts for common morphemes that are not words (and even, experimentally, for punctuation). Each word expert for -ing, which aids significantly in helping to
Some word senses of "heavy"

1. An overweight person is politely called "heavy".
   "He has become quite heavy.

2. Emotional music is referred to as "heavy".
   "Mahler writes heavy music.

3. An intensity of precipitation is "heavy".
   "A heavy snow is expected today.

Some word senses of "ice"

1. The solid state of water is called "ice".
   "Ice melts at 0°C."

2. "Ice" participates in an idiomatic nominal describing a favorite delight.
   "Homemade ice cream is delicious."

3. "Dry ice" is the solid state of carbon dioxide.
   "Dry ice will keep that cool all day.

4. "Ice" or "iced" describes things that have been cooled (sometimes with ice).
   "One iced tea to go please."

5. "Ice" also describes things made of ice.
   "The ice sculptures are beautiful!"

6, 7. "Ice hockey" is the name of a popular sport which has a rule penalizing an action called "icing":
   "He iced the puck causing a face-off."

8. The term "ice box" refers to both a box containing ice used for cooling foods and a refrigerator:
   "This ice box isn't plugged in!"

Figure 1: Example contextual word usages

"over", and so forth, from a disk file, and organizing them along with data repositories called word bins in a left to right order in the sentence level workspace. Note that three copies of each expert for "deep" and "pit" appear in the workspace. Since each expert executes as a process, each process instantiation in the workspace must be put into an executable state. At this point, the parse is ready to begin.

The word expert for "the" runs first, and is able to terminate immediately, creating a new concept designator (called a concept bin and participating in the concept level workspace) which will eventually hold the data about the intellectual philosopher described in the input. Next the "deep" expert runs, and since "deep" has a number of word senses, is unable to terminate (i.e., complete its discrimination task). Instead, it suspends its execution, stating the conditions upon which it should be resumed. These conditions take the form of associative trigger patterns, and are referred to as disambiguating expressions involving gerunds or participles such as "the man eating tigers". A full discussion of this will appear in [12].

Although I call them "processes" word experts are actually coronavirus-like, resembling CONNIVER's generators [13], and even more so, the stack groups of the MIT LISP Machine [14].

It should be clear that the notion of "word sense" as used here encompasses what might more traditionally be described as "contextual word usage". Aspects of a word token's linguistic environment constitute its broadened "sense".
3. Structure of the Model

The basic structure of the Word Expert Parser (and thus the model) can be described as a network of interconnected experts, each of which specializes in a specific aspect of word sense disambiguation. The parser contains a central tableau of control by the word expert processes, one of which is associated with the resumption and termination of the "deep" expert processing or the states of processing of other model components. The parser also contains a central tableau of control by the word expert processes, one of which is associated with the resumption and termination of the "deep" expert processing or the states of processing of other model components.

Word Experts

The principal knowledge structure of the model is the word sense discrimination expert. A word expert represents the knowledge required to disambiguate the meaning of a word in any context. Although represented computationally as coroutines, these experts differ considerably from ad hoc LISP programs and have approximately the same relation to LISP augmented transition network (ATN) grammar. As the architecture is represented in Figure 2, the basic structure demonstrates the network paradigm of the ATN program representation of an expert as a point in a network represented by a word sense discrimination network structure with a word sense discrimination network structure. The word sense discrimination network structure contains a number of expert networks called word sense discrimination networks, each of which is a network that represents the process of converging on a single conceptual usage of a word. These networks are distinguished by the words they represent and the terminal nodes of other experts. Each word sense discrimination network contains a number of experts that specialize in the sense of the word modeled by the network. A sense net for the word "heavy" appears in part (a) of Figure 2. Examination of this network reveals that four senses are represented -- the three adjecive usages shown in Figure 1 plus the nominal sense of "thug" as in "Joe's heavy told me to beat it."

Expert Representation

The network representation of a word expert leaves out certain computational necessities of actually using it in parsing. A word expert has two fundamental activities: (1) It analyzes lexicographical and conceptual data being absorbed by its context, the control states of various model components, and more general issues requiring common sense knowledge of the physical world. (2) It makes decisions that affect the lexical and conceptual contents of the workspaces, the control states of itself, concept bins, and the parser as a whole, and the model's overall performance. The current procedural representation of the word expert for "heavy" appears as part (b) of Figure 2.

Each word expert process includes three components: a declarative process, a question process, and a body. The header provides a description of the expert's behavior, for purposes of inter-expert constraints forward propagation. If sense discrimination by a word expert results in the knowledge that a word to which a concept bin is attached is executed in a specific sense or conceptual category, then it should constrain it to be associated with that concept bin. If it fails, then all such expert processes, containing such a concept bin, should be constrained to the pointer to the address at which it expects to continue execution. Through its descriptive header, an expert provides information about its activity and is fairly simple. If the question process is asked at node n2 of the "heavy" expert is typical:

"Is the object to my right better described as an artistic object, a form of precipitation, or an intellectual philosopher?"

Action nodes in the "heavy" expert perform such tasks as determining the concept bin to which it contributes, and posting expectations for the word to its right. In terms of the model's operation, an expert is more complex, and whereas some queries can be made in more than one way, the several question types solicit different kinds of information. Some questions require fairly direct inference to be answered adequately, and others demand no more than simple registries of facts or lookups. This variety corresponds well, in my opinion, with human processing involved in conceptual analysis. Certain contextual clues to meaning are strongly evident, and they often indicate when more general issues requiring common sense knowledge of the physical world. In addition, a network of instructions affecting the lexical and conceptual contents of the workspaces, the control states of itself, concept bins, and the parser as a whole, and the model's overall performance. The current procedural representation of the word expert for "heavy" appears as part (b) of Figure 2.

Each word expert process includes three components: a declarative process, a question process, and a body. The header provides a description of the expert's behavior, for purposes of inter-expert constraints forward propagation. If sense discrimination by a word expert results in the knowledge that a word to which a concept bin is attached is executed in a specific sense or conceptual category, then it should constrain it to be associated with that concept bin. If it fails, then all such expert processes, containing such a concept bin, should be constrained to the pointer to the address at which it expects to continue execution. Through its descriptive header, an expert provides information about its activity and is fairly simple.

Each node in the body of the expert has a type designated by a letter following the node name, either Q (question), A (action), S (suspend), or T (terminal). By tracking through the question nodes (treating the others as vacuous except for their goto pointers), sense network for each word expert process can be derived. The graphical framework of a word expert (and thus the question process) is essentially a phrase structure representation task of word sense disambiguation. Each question node has a type shown following the Q in the node name. MC (multiple choice), C (conditional), YN (yes/no), and PI (possible/impossible).

Multiple choice questions typically derive into the basic relation among objects and actions in the world. For example, the question asked at node n2 of the "heavy" expert is typical:

"Is the object to my right better described as an artistic object, a form of precipitation, or an intellectual philosopher?"

The basic structure of the Word Expert Parser depends principally on the role of individual word experts in affecting (1) each other's actions and (2) the dependence of the result of one's action on the actions of others. Each expert affects each other by posting expectations on the central bulletin board, constraining each other, changing control states of other model components. Expert processes for augmenting data structures in the workspaces are constrained to the conceptual level of the parsing process by posting object names, descriptions, schemata, and other useful data to the concept level workspace. To determine what action to take at any instance, the expert uses a set of various kinds about the processes of the model and the world at large.

Four types of questions may be asked by an expert, and whereas some queries can be made in more than one way, the several question types solicit different kinds of information. Some questions require fairly direct inference to be answered adequately, and others demand no more than simple registries of facts or lookups. This variety corresponds well, in my opinion, with human processing involved in conceptual analysis. Certain contextual clues to meaning are strongly evident, and they often indicate when more general issues requiring common sense knowledge of the physical world. In addition, a network of instructions affecting the lexical and conceptual contents of the workspaces, the control states of itself, concept bins, and the parser as a whole, and the model's overall performance.

The blackboard of the Hearsay speech understanding system [6] is analogous to the entire workspace of the parser, including the word bins, concept bins, and bulletin board.
The Importance of Multiple Choice

Multiple choice questions make possible but answer them tremendously complex processing. A substantial knowledge representation formalism based on semantic networks, such as RAIN (19), with multiple parameters, with intelligent description matching, must be used to represent in a uniform way both general and specific knowledge acquired through textual interpretation. In RAIN terms, a multiple choice question such as "Is the sentence RAIN more like ARTISTIC-OBJECT, PHYSICAL-OBJECT, or PRECIPITATION?" must be answered by appeal to the units representing the options involved. Clearly, RAIN can be viewed as a PHYSICAL-OBJECT; much less so as an ARTISTIC-OBJECT. However, in almost all contexts, RAIN is closest conceptually to PRECIPITATION. Thus, this should be the answer.

Expert Side Effects

Word experts take two kinds of actions — actions explicitly intended to affect sense discrimination by other experts, and actions to augment the conceptual information that constitutes the result of a computation. Each path through a sense network represents a distinct usage of the modeled word, and at each step of the way, the word expert must update the model to reflect the state of its processing and the extent of its knowledge. The heavy expert of Figure 2(b) exhibits several of these actions. Nodes n2 and n3 of this word expert process represent "heavy"'s decision about the concept bin (i.e., conceptual notion) it is to be associated with.

The initial execution of a word expert must accommodate certain special cases. If the word participates in a noun-noun pair, this must be determined; in either case, the expert must determine the concept bin it contributes to, based on its descriptive data throughout the parse. This concept is illustrated in the next example.

An exception arises when an expert creates a default concept bin to represent a conceptual notion referenced in the text, but to which no words in the text contribute. The automobile in "Joanie parked." is an example.
could either be one that already exists in the workspace or a new one created by the expert at the time of its decision. After deciding on a concept, the principal role of the expert is to discriminate the possibly many remaining senses of the word. Note that a good deal of this disambiguation may take place during the initial phase of concept determination. After asking enough questions to discover some piece of conceptual data, this data augments what already exists in the word's concept bin, including declarative structures put there both by itself and by the other lexical participants in that concept. The parse completes when each word expert in the workspace has terminated. At this point the concept-level workspace contains a complete conceptual interpretation of the input text.

Conceptual Case Resolution

Adequate conceptual parsing of input text requires a stage involving from this discussion and constituting the current phase of research --- the attachment of each picture and setting concept (bin) to the appropriate conceptual case of an event concept. Such a mechanism can be viewed in an entirely analogous fashion to the mechanism for a contextual language experter to perform local disambiguation of word senses. Rather than word experts, however, the experts at this level are conceptual in nature. The concept level thus becomes the main level of activity and a new level, call it the schema level of disambiguation.

If it is an event concept, its function is to fill its conceptual cases with settings and pictures; if it is a setting or picture, it must determine its schematic role.

The implementation of the mechanisms to carry out this next step presents both by itself and in the example parse) would not be possible. In the new usage differ from those present for a known sense.

4. Summary and Conclusions

The Word Expert Parser is a theory of organization and control of a word expert system. The control environment is characterized by a collection of word experts, any of which cooperatively arrive at a conceptual interpretation of an incoming sentence. The non-linguistic knowledge available to these experts in performing their task, including control state knowledge and knowledge of the world, and eliminate all but the most persistent forms of ambiguity, is the partner model human processing.

This new model of parsing claims a number of theoretical advantages: (1) Its representational ability of linguistic knowledge reflect the enormous redundancy in natural languages -- without this redundancy in the model, the inter-expert handshaking (seen in many forms in the example parse) would not be possible. (2) The model suggests some interesting approaches to language acquisition. Since much of a word expert's knowledge is encoded in a branching discrimination structure, adding new information about a word involves the addition of a new branch, this branch would be placed in the expert at the point where the contextual clues for discriminating the new usage differ from those present for a known usage. (3) Isomorphic uses of language are easily encoded, since the word expert provides a clear way to do so. (4) The model is a relatively simple model of sequential concurrent-like processing in human language understanding. The organization of linguistic knowledge is new, rather than the more instable his conjectures about the flow of control in a human language understander.

ACKNOWLEDGMENTS

I would like to thank Chuck Rieger for his insights, encouragement, and generosity of his ideas presented here. Chuck has graciously allowed me to steal. In addition, I thank the following people for helping me with this work through their comments and suggestions: Phil Agre, Hilt Grinberg, Phil London, Jim Reggia, Hanay Samet, Randy Trigg, Rich Wood, and Pamela Zave.

REFERENCES

[1] Rieger, C., and S. Small, Word Expert Parsing, Proceedings of the 8th International Joint Conference on Artificial Intelligence, 1979.
[2] Riesbeck, C., Computational Understanding: Analysis of Sentences and Context, AI-Memo 238, Stanford University, 1974.
[3] Riesbeck, C., and R. Schank, Comprehension by Computer: Expectation-based Analysis of Sentences in Context, Research Report 78, Yale University, 1976.
[4] Schank, R., Conceptual Dependency: A Theory of Natural Language Understanding, Cognitive Psychology, vol. 3, no. 4, 1972.
[5] Wilks, Y., Making Preferences More Active, Artificial Intelligence, vol. 11, no. 7, 1978.
[6] Marcus, M., Capturing Linguistic Generalizations in a Parser for English, Proceedings of the 2nd National Conference of the Canadian Society for Computational Studies of Intelligence, 1978.
[7] Rieger, C., The Importance of Multiple Choice, Proceedings of the 2nd Conference on Theoritical Issues in Natural Language Processing, 1978.
[8] Rieger, C., Viewing Parsing as Word Sense Discrimination, A Survey of Linguistic Science, Dingwall (ed.), Greylock Pub., 1977.
[9] Rieger, C., Five Aspects of a Full Scale Story Comprehension Model, Associative Networks -- The Representation and Use of Knowledge in Computers, Findlay and Gold, Academic Press, 1979.
[10] Rieger, C., An Organization of Knowledge for Problem Solving and Language Comprehension, Artificial Intelligence, vol. 7, no. 2, 1976.
[11] Small, S., Conceptual Language Analysis for Story Comprehension, Technical Report No. 563, University of Maryland, 1978.
[12] Small, S., Word Experts for Conceptual Language Analysis, Ph.D. Thesis (forthcoming), University of Maryland, 1980.
[13] McDermott, D., and C. Susman, The Combiner Reference Manual AI-Memo 259a, Massachusetts Institute of Technology, 1974.
[14] Lisp Machine Group, LISP Machine Progress Report, AI-Memo 444, Massachusetts Institute of Technology, 1977.
[15] Woods, W., Transition Network Grammars for Natural Language Analysis, Communications of the ACM, vol. 13, no. 10, 1970.
[16] Ermann, L. and V. Lesser, A Multi-Level Organization for Problem Solving Using Many Diverse Cooperating Sources of Knowledge, Proceedings of the 6th International Joint Conference on Artificial Intelligence, 1975.
[17] Reggia, J., Representing and Using Medical Knowledge by Computer, Proceedings of the 8th National Conference on Artificial Intelligence, 1977.
[18] Susman, G., T. Winograd, and E. Charniak, Micro-Planner Reference Manual AI-Memo 204a, Massachusetts Institute of Technology, 1971.
[19] Bobrow, D., and T. Winograd, An Overview of KRL, A Knowledge Representation Language, Cognitive Science, vol. 1, no. 1, 1977.
[20] London, P., Dependency Networks as a Representation for Modeling in General Problem Solvers, Technical Report 698, University of Maryland, 1978.
