

Word Sense and Case Slot Disambiguation

Graeme Hirst and Eugene Charniak

Department of Computer Science
Brown University
Providence, RI 02912

Abstract

The tasks of disambiguating words and determining case are similar and can usefully be combined. We present two cooperating mechanisms that each work on both tasks: *MARKER PASSING* finds connections between concepts in a system of frames, and *POLAROID WORDS* provide a protocol for negotiation between ambiguous words and cases. Examples of each in action are given. The cooperating mechanisms allow linguistic and world knowledge to be unified, frequently eliminate the need to use inference in disambiguation, and provide a usefully constrained model of disambiguation.

1. Introduction

The problem of determining the correct sense of a lexically ambiguous word in context is often seen primarily as one of context recognition, a word being disambiguated to the unique meaning appropriate to the frame or script representing the known or newly-established context. For example:

- (1) Nadia's new car is a lemon.

The *new-car* script/frame/context selects the *poorly-made-car* meaning for *lemon*. In practice, of course, there are many complications. This approach would fail with (2):

- (2) Nadia's new car is the color of a lemon.

In general, word sense can depend not only upon global context, but also (or only) upon the meaning of nearby words. In (3):

- (3) The bank was robbed.

the *river-edge* meaning of *bank* can be rejected in favor of the meaning that has an association with robbery, without any consideration of global context. The nearby disambiguating words may themselves be ambiguous; a well-known example (Small 1980) is *deep*

pit, where *deep* can mean *profound* or *extending far down*, and *pit* can be *fruit stone* or *hole in the ground*. However, only one meaning of each fits with the other, so they are mutually disambiguating.

A closely related problem is that of case slot disambiguation. Compare:

- (4) Ross drove to town with reckless abandon.
(5) Ross drove to town with Nadia.

Determining that *reckless abandon* is an instance of the *manner* case (and hence the *manner* slot inherited by the *driving* frame—see Charniak (1981a) for justification), while *Nadia* is an instance of the *accompanier* case, is very like disambiguating the word *with*. Indeed, we can think of determining the case of subject and object NPs as disambiguating the “prepositions” *SUBJ* and *OBJ* that are present in the deep structure, but that in English are deleted before the surface realization.

Necessary for word sense disambiguation, then, are mechanisms to find associations between concepts, to handle low-level disambiguation cues, and to handle reconciliation negotiations between ambiguous words. Two mechanisms that cooperate on these tasks are described below. They are *MARKER PASSING* and the *POLAROID*¹ *WORD* system. We will show how we use these mechanisms for word sense disambiguation, and how they are also used for case slot disambiguation.

2. Marker passing

Marker passing (*MP* for short) is a mechanism for determining connections in a semantic net of frames. It can be thought of as passing markers around the network, from node to node, from slot to filler, under the constraint rules to be discussed below. Marker passing was proposed by Fahlman (1979) for his *NETL* system,²

¹*POLAROID* is a trademark of the Polaroid Corporation.

though the idea goes back to Quillian (1968).

Our marker passer operates independently of, and in parallel with, a parser based on that of Marcus (1980). That is, following only basic morphological analysis, the input sentence goes to both the Marcus parser and the marker passer, both of which separately grind away on each word as it comes in. For example, suppose the input is (6):

(6) Nadia's plane taxied to the terminal.

The words *plane*, *taxi* and *terminal* are all ambiguous. The MP places a marker on the frame representing each known meaning of each substantive word in the sentence (including unambiguous ones). Then it follows frame and slot connections from each of these starting points, marking each it passes through.

To prevent as many "uninteresting" paths as possible, certain steps are prohibited. For example, a connection could be found between **airplane** and **bowling-ball** simply by running up the ISA chain from **airplane** through **vehicle** and the like to **human-made-object**, and then down another ISA chain from there to **bowling-ball**. This is prevented by the anti-promiscuity rule, which does not allow paths to propagate from nodes with more than n connections. This is because nodes with many connections tend to be uninteresting ones near the top of the ISA hierarchy—**human-made-object**, for example. The search is bounded by a depth limit rather than an absolute limit on the number of markers passed; that is, a path may not continue more than d nodes below its starting point.³ In the example above, paths would be found between **airplane** and **airport-building**, and **airplane** and **aircraft-ground-travel**, indicating that the corresponding meanings of *plane*, *terminal*, and *taxi* should be chosen. Many other paths will also be found; we discuss the problem of FALSE POSITIVES in Charniak (1982), and posit a PATH CHECKER that will filter out many paths that are uninteresting or silly.

3. Polaroid Words

The Polaroid Word mechanism (PW to its friends) is responsible for making the final disambiguation decision for each word. Often, as in the case of (6), this requires no more than looking at the paths found by the

marker passer. At other times, MP will return nothing overwhelmingly conclusive; it is then necessary for PWs to use other information and negotiation between possible meanings.

PWs operate in parallel with the parser. As each word comes in to the parser and its part of speech is assigned, a PW process is created for it. (If the word is unambiguous, the process merely announces the meaning and uninterestingly hibernates; PWs always knock off as soon as they have narrowed their possible meanings to just one.) Meanwhile the marker passer has been looking for connections between the new word and those already seen. The new PW looks at these MP connections, and, if it can, reduces the number of meanings that this instance of the word could represent. It then starts negotiations with preceding words to see if it can further disambiguation, either its own or that of another PW. If still unresolved, it will then sleep until a new word, possibly the bearer of helpful information, comes along.⁴

A PW, upon creation, knows all the possible senses for its word, and something about the conditions that would support each. For example, the adjective *green* knows that in its color sense it can only qualify an entity that ISA **physical-object**. Once awake, a PW proceeds as follows: First, it looks at the decisions that have been announced by the preceding PWs that it can negotiate with. A PW may negotiate with other PWs as follows: a verb with the prepositions and nouns it dominates; a preposition with the noun of its prepositional phrase, and with other prepositions; and a noun or verb modifier with the word it modifies. (All these negotiation paths are bi-directional.) An announcement may take the form "I mean X and so my friends should be Y", or "I could mean X if my friends are Y, or I could mean Z and set no conditions for that", and so on. Next, the PW looks at the MP chains that have been found between any of its possible meanings and those of the preceding words of the sentence. A strong chain (where strength is inversely proportional to length, though the types of the links traversed are also considered) is *prima facie* evidence that its starting point is the right meaning for the word. From this information, the PW eliminates any meanings that don't

²Marker passing is, of course, expensive when the net is interestingly large. Fahlman, who intended using MP for deduction, proposed super-parallel hardware to back it up. Although our scheme is much simpler than Fahlman's, we too assume that hardware of the future will, like people of the present, be able to derive connections between concepts in parallel.

³We are still experimenting with constraint rules, and values for n and d . The rules mentioned are examples of those we have considered; we make no claims for their status as Absolute Truth.

⁴The name *Polaroid Word* derives from the fact that the meaning of each word develops right in front of the parser's eyes, and the print can be handed around by the parser, its state of development a matter of indifference as long as the meaning of the word isn't needed to guide the parse. And if the meaning is needed (the parser having occasionally to resort to semantic advice), an examination of a half-developed image may suffice.

suit its friends.⁵ Of the meanings that remain, MP may have suggested one strongly enough for the PW to confidently announce it as the word's sense; otherwise it will announce all its remaining possibilities, and their conditions. It then rests a while.

The announcement wakes up other PWs that have not yet made their final decision, and each sees whether the new information—both the new word's announcement, and any MP chain between it and the new word—helps it make up its mind. If so, it too makes an announcement, again awakening its friends. This continues until none can do any more. Then the next word in the sentence comes in, its PW is created, and the sequence is repeated.

Let's consider this example, concentrating on the subordinate clause:

- (7) Ross found that *SUBJ* the slug would operate *OBJ* the vending machine.

Note the insertion of the "prepositions" *SUBJ* and *OBJ*. We want to work out that *the slug* is a metal stamping, not a gastropod, a bullet or a shot of whiskey; that the frame *operate* refers to is **cause-to-function**, not **perform-surgery**; and that *SUBJ* and *OBJ* indicate the slots **instrument** and **patient** respectively.

The first words are *SUBJ* and *slug*; their PWs have not yet enough information to do anything interesting, nor has marker passing produced anything. Then *operate* comes along, and tells the others that, *inter alia*, it might have a meaning that ISA **transitive-action**, or then again it might have a meaning that ISA **intransitive-action**. From this, and from the slot-filler requirements for **agent** and **instrument** in the **transitive-action** and **intransitive-action** frames, the *SUBJ* PW now realizes it has three possibilities:⁶

- *operate* is transitive and *SUBJ* indicates the agent, which should be animate and sentient.
- *operate* is transitive and *SUBJ* indicates the instrument, which should be inanimate.
- *operate* is intransitive and *SUBJ* indicates the agent, which should be animate and sentient.

It announces this to the world. The PW for *slug* can now reconcile this with its own options. It finds that none of its possible meanings fall in the ISA hierarchy under **sentient-entity**, and announces this. Now *SUBJ* can eliminate all but the second of its potential meanings, and when it arrives *OBJ* can immediately decide that it indicates the **patient** slot.

⁵If it finds that this removes a sense strongly suggested by MP, then it is in trouble; a likely reason is that the word is being used metaphorically, a situation we do not attempt to handle.

⁶At present, we ignore the possibility of passive sentences.

The noun phrase *vending machine* now arrives,⁷ bringing with it an MP chain that might be **vending-machine—coin—metal-disc—slug**. This is enough for the *slug* PW to favor **metal-disc** as its meaning, and all words are now disambiguated.

Now consider this example, in which marker passing is not used at all:

- (8) *SUBJ* the crook operated *OBJ* a pizza parlor.⁸

This proceeds as before, until the *SUBJ* PW has laid out its three options. Since *crook* can be either sentient (a criminal) or not (a shepherd's staff), it is unable to make the move that in the previous example disambiguated *operate*. However, when *OBJ* comes along, the *operate* PW can immediately eliminate the intransitive possibility. However, after the last word is processed, the PWs' negotiations reach a standstill with *crook* still undisambiguated.

If it happens that at the end of the sentence one or more words are not fully disambiguated, then knowledge of a **PREFERRED** or **DEPRECATED MEANING** is invoked for them. Preferred and deprecated meanings are indicated as part of the dictionary entry for each word; a word can have zero or more of each, and a meaning need not be either. The possibilities that remain are ranked accordingly, and the top one or ones are chosen. If there are two or more, the word is ambiguous; inference (see next section) may be deployed if desired.

It will be apparent that our use of slot-filler constraints is not dissimilar to conventional Katz and Fodor (1963) selectional restrictions. The difference is that our constraints are not just symbols in a dictionary entry, but rather are part of the knowledge base. That is, we don't mark the word *slug* as [$\pm$ animate]; rather it is part of our world knowledge that the frame **gastropod-without-shell** represents something that inherits properties from the **animate-entity** frame that stands above it in the ISA hierarchy.

There are two advantages in this. First, we avoid arbitrary and syntactically unmotivated selectional restrictions; it seems more natural to require the knowledge base to decide whether the object has the necessary properties, using whatever proof techniques may be necessary. Second, we now have a much more unified representation of knowledge, both world and linguistic. We do not, for example, store in one place the fact that the doer of an action must be an sentient

⁷We assume that it is recognized as an unambiguous canned phrase representing a single concept.

⁸We claim that this is exactly the same meaning of *operate* as in the previous example: **cause-to-function**. In a context like this, the action described is generic, a matter we ignore.

entity, and in another that the agent of the verb representing that action must be sentient. Rather, the frame system contains the information but once, and it can be used wherever it is needed. This is in accord with the goal of the project of which the present work forms a part: a uniform knowledge representation suitable for both natural language understanding and problem solving (see section 5).

4. What we can't do

It should be clear that a combination of marker passing and Polaroid Words is not a replacement for inference and pragmatics in word sense and case disambiguation; rather, it serves to substantially reduce the number of times that these must be employed. In particular, of course, it eliminates most cases where inference would be employed to find contextual connections.

5. Existential considerations

Both MP and PW exist at present only in prototype implementations. We are writing more sophisticated versions of each as part of our group's development of an integrated language-understanding and problem-solving system (Charniak 1981b, Wong 1981a,b), and will try out different MP constraint rules and PW negotiation protocols to find which are best.

Our use of MP will not be limited to disambiguation; we also expect higher semantic levels of the system to use it for determining context and finding causal links between events—see Charniak (1982) for discussion.

Polaroid Words are implemented as processes that interpret Lisp data structures that each word has in its dictionary entry, containing purely lexical information. We anticipate developing a simple, constrained declarative language, similar in spirit to Marcus's PIDGIN (1980), for writing Polaroid Words. This is in contrast to approaches such as Small's (1980), where the meaning of a word is represented as a large, barely-constrained procedure. We avoid this by letting the parser and the marker passer do much of the work that Small requires his "word experts" to perform.

6. Conclusion

We have presented a pair of cooperating mechanisms that both disambiguate word senses and determine case slots by finding connections between concepts in a network of frames and by negotiating with one another to find a set of mutually satisfactory meanings. This approach permits linguistic and world knowledge to

be combined in a single representation, and, in turn, is to be part of a system in which knowledge for language understanding and problem solving are similarly unified.

References

- CHARNIAK, Eugene (1981a). "The case-slot identity theory." *Cognitive Science*, 5(3), July–September 1981, 285–292.
- CHARNIAK, Eugene (1981b). "A common representation for problem-solving and language-comprehension information." *Artificial Intelligence*, 16(3), July 1981, 225–255.
- CHARNIAK, Eugene (1982). "Passing markers: A theory of contextual inference in language comprehension." Technical report CS-80, Department of Computer Science, Brown University, Providence, RI 02912.
- FAHLMAN, Scott (1979). *NETL: A system for representing and using real-world knowledge*. Cambridge, MA: The MIT Press, 1979.
- KATZ, Jerrold Jacob and FODOR, Jerry A (1963). "The structure of a semantic theory." *Language*, 39(2), 1963, 170–210.
- MARCUS, Mitchell P (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: The MIT Press, 1980.
- QUILLIAN, M. Ross (1968). "Semantic memory." in: MINSKY, Marvin Lee (editor). *Semantic Information Processing*. Cambridge, MA: The MIT Press, 1968. 227–270.
- SMALL, Steven (1980). "Word expert parsing: A theory of distributed word-based natural language understanding." Doctoral dissertation [available as technical report TR-954], Department of Computer Science, University of Maryland.
- WONG, Douglas (1981a). "Language comprehension in a problem solver." *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, 24–28 August 1981. 7–12.
- WONG, Douglas (1981b). *On the unification of language comprehension with problem solving*. Doctoral dissertation [available as technical report CS-78], Department of Computer Science, Brown University, 1981.