FRAME SELECTION IN PARSING*

Steven L. Lytinen

Dept. of Computer Science, Yale University
New Haven CT 06520

ABSTRACT

The problem of frame selection in parsing is discussed, with the focus on the selection of a frame for texts which contain highly ambiguous or vague words. An approach to frame selection is presented which involves the use of a small number of general inference rules, in conjunction with a hierarchically-organized conceptual memory. This is in contrast to various other methods, which rely either on disambiguation rules stored in the dictionary definitions of ambiguous words, or on previously-activated frames to guide the selection of new frames. The selection of frames for vague or ambiguous words using these previous methods is shown to be problematic. The method presented here does not suffer from these same problems because of the hierarchical organization of memory and the general inference rules used.

I Introduction

One problem which has been encountered in the use of frames [6] or other frame-like structures such as scripts [8] in natural language processing has been the selection of the appropriate frame for a text. How does a system with a large number of frames choose the correct one? Sometimes, particular words in a text point directly to a particular frame, thus trivializing this problem. For example, the word "arrest" refers directly to the $ARREST script. However, more often it is the case that no one word in a text points definitively to a unique frame. Instead, many of the words in the text are ambiguous or vague, and it is only by considering them in combination that a frame can be selected. An arrest, for instance, can be described without using the word "arrest", as in "Police took a suspect into custody", or even "They got their man". In cases like this, frame selection is much more difficult.

In this paper, I will present a method for selecting frames for texts containing ambiguous or vague words. This method uses a small set of general inference rules in conjunction with frame-like structures organized in a hierarchical fashion. This method of frame selection has been developed within a multi-lingual machine translation project, called the MOPTRANS system [5], which translates short (1-2 sentences) newspaper articles about terrorism.

II Previous Approaches to Frame Selection

A. Lexically-based Disambiguation Rules

One approach to the frame selection problem in parsing has been to use disambiguation rules which are stored in the dictionary definitions of ambiguous or vague words. I will call this the *lexically-based approach*. This approach was first used in Riesbeck's parser [Riesbeck75], and since then in many other parsers, such as the Word Expert Parser [Small80]. In this approach, disambiguation rules take the form of test-action pairs, called *requests* or *demons*. For each possible meaning of an ambiguous or vague word, one or more requests/demons are responsible for determining if that meaning of the word is the one being used in the given context. The requests in an ambiguous word's dictionary entry are activated when the parser encounters the word. Then, one of these requests *fires*, or is executed, when its conditions are met by the state of active memory, thus choosing its word sense as the meaning of the word in that context.

The lexically-based approach to frame selection was used in the Word Expert Parser to disambiguate the word "throw" to its many possible meanings. In WEP, the dictionary definition of a vague word consisted in part of a discrimination net of possible concepts to which the word could refer, as well as a group of demons which functioned as the discrimination rules for the d-net of frames. Thus, the frames for "throw", which included frames such as PERSON-THROW, THROW-OBJECT-TO-LOCATION, and THROW-OUT-GARBAGE, were arranged in a discrimination net, with demons for choosing the correct frame such as "If the agent of "throw" is a person, then refine "throw" to PERSON-THROW", and "If the object of PERSON-THROW is garbage, then refine PERSON-THROW to THROW-OUT-GARBAGE."

The problem with the lexically-based approach to frame selection is that in a system with a large number of frames, very vague words require a great many disambiguation rules. This is because the number of disambiguation rules needed for a given word is proportional to the number of frames to which the word could possibly refer, since each possible meaning of a word requires one or more rules to determine whether or not that meaning is being used in a given context. In WEP, the number of demons needed to disambiguate "throw" to all of its possible meanings was quite large, since "throw" could refer to a large number of frames. The word "throw" is relatively specific compared to other English verbs like "encounter", or "do", which could refer to almost any action. Extremely vague words such as these would require an unmanageably large number of disambiguation rules to handle all the possible frames to which they could refer.

*This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under contract No. N00014-82K-0149

B. Expectations from Other Frames

Another approach to the frame selection problem involves using frames to predict the occurrence of other frames. I will call this approach the ***frame-based approach.*** This approach was used in the GUS system [Bobrow77], a system which conversed about airplane trips; and in the Integrated Partial Parser [Lebowitz80], which parsed short newspaper articles about terrorism. In this approach, frames already selected are responsible for predicting other frames that are likely to appear in a text. These predictions can help to disambiguate words which could refer to many different frames. For example, in IPP the word "seized" could refer to many different scripts: $HIJACK, $TAKE-OVER (a building), and $KIDNAP. Expectations from already active structures often determined which of these scripts "seized" referred to. Thus, if the structure EXTORT, another frame in IPP, was already active, then "seized" was assumed to mean $HIJACK, since hijackings are often part of extortions.

The frame-based approach solves some of the problems with frame selection. This approach does not suffer from the rule explosion encountered in the lexically-based approach, since the number of rules does not depend on the number of meanings of a word. However, an obvious problem with this approach is the selection of an initial frame. If no frames are active at the beginning of a story, then no predictions can be made as to what frames will occur in the story. To avoid this problem, the GUS system only dealt with texts having to do with airplane trips. Thus, the trip specification frame was always active at the beginning of the story. This frame could then be used to predict other frames that might appear in the text. The IPP parser also relied in part on a restricted domain to deal with the problem of selecting an initial frame. Many words in English which are vague in general are unambiguous in the domain of terrorism, and thus were unambiguous in IPP. For instance, the word "divert" in IPP referred to only one frame, namely $HIJACK.

C. Frame Selection by Discrimination

Another approach to frame selection was used in the FRUMP system [DeJong79]. FRUMP produced summaries of newspaper articles from many domains. Thus, the frame selection problem was very real in FRUMP. To handle this problem, DeJong used discrimination nets called sketchy script initiator discrimination trees (SSIDTs). One SSIDT existed for each conceptual dependency primitive. An SSIDT, when given a conceptual dependency representation, selected a frame, or "sketchy script", on the basis of the roles and role fillers contained in the Conceptual Dependency representation. Thus, a text was first decomposed into its CD representation, then parsing rules would fill in various roles in the representation, and finally an SSIDT selected a sketchy script on the basis of what roles were filled in, and how they were filled.

FRUMP used the SSIDT in Figure 1 to select the sketchy script $EARTHQUAKE for the sentence "The ground trembled." First, the word "trembled" was represented by PTRANS, the CD primitive for physical motion. "Trembled" also provided the information that the motion was cyclical in manner. Then, parsing rules assigned "ground" to be the OBJECT of this PTRANS. This role-filling information guided the script selection process through the SSIDT to the sketchy script $EARTHQUAKE.

FRUMP's appoach to frame selection does not suffer from the same rule explosion as the lexically-based approach. Also, it can select an initial frame for a story, unlike the frame-based approach. However, it has the disadvantage of

```
                      PTRANS
                        |
                  node 2 (OBJECT)
                  /     |     \
          GROUND  VEHICLE  HUMAN
             |
       node 3 (ACTOR)
          /   \
  EXPLOSIVE   GEOLOGICAL FORCE
                    |
             node 4 (MANNER)
                    |
                CYCLICAL
                    |
               $EARTHQUAKE
```

Figure 1: SSIDT for PTRANS

requiring large, ad hoc discrimination trees, whose only purpose is to disambiguate words. Also, it depends on the ability to initially represent a text in terms of conceptual dependency primitives. While this works well for words such as "trembled", which refer very clearly to physical actions, it is not as easy to represent the meanings of all words in terms of conceptual dependency primitives.

III A Different Approach to Frame Selection

I will now discuss a different approach to the problem of frame selection, which uses a small number of general inference rules in conjunction with a hierarchically arranged set of frames. The frames range from very general to very specific, depending on their position in the hierarchy.

To explain this approach, I will discuss the frame selection process for a very general Spanish phrase, "hacer diligencias", which was encountered by the MOPTRANS program. Literally, this phrase means "to do diligent actions". Often, it is equivalent to the English "to run errands", as in the following example:

Spanish: Maria no puede ir a la reunion porque tiene que HACER MUCHAS DILIGENCIAS.

English: Mary cannot go the gathering because she HAS TO RUN A LOT OF ERRANDS.

However, it can mean many other things, depending on the context in which it appears. This is because often the context provides enough information to allow the reader to infer quite specifically what action the phrase refers to. Here are some examples*:

Spanish: Juanita salio a HACER UNAS DILIGENCIAS AL MERCADO.

English: Juanita went TO SHOP FOR GROCERIES.

Spanish: Va a pintar su apartamento? -- Si, pero antes tengo que HACER UNAS DILIGENCIAS PARA VER si consigo la pintura que quiero.

English: Are you going to paint your apartment? -- Yes, but first I have TO GO SEE if I can find the paint that I want.

Spanish: La policia REALIZA INTENSAS DILIGENCIAS PARA CAPTURAR a un reo que dio muerte a una mujer.

English: The police ARE UNDERTAKING AN INTENSE INVESTIGATION in order to capture a criminal who killed a

*Sometimes, the verb "realizar" (to realize or achieve) is used in place of "hacer".

woman.

From these examples, we see that a large number of different frames can be used to represent "hacer diligencias" in different contexts.

How can we devise rules which disambiguate a vague phrase like "hacer diligencias"? At first glance, one might think that we could find straightforward features of the surrounding context which would discriminate between at least some of the different frames to which the phrase can refer. For instance, in the police investigation story above, the fact that POLICE is the ACTOR of "realizar diligencias" might be enough to discriminate this sense of the phrase. However, this is not the case, as the following example illustrates:

Spanish: La reina Isabela va a visitar a la ciudad de Nueva York el lunes. La policia REALIZA DILIGENCIAS para insurar su seguridad durante la visita.

English: Queen Elizabeth will visit New York city on Monday. The police ARE TAKING PRECAUTIONS to insure her safety during her visit.

Let us continue to look at the police investigation example above. What parts of the context in this example are relevant to determining that "realizar diligencias" means POLICE-INVESTIGATION? To answer this, consider the line of reasoning that a human reader might follow in order to infer this. First, since the prepositional phrase "para capturar" (in order to capture) follows "realizar diligencias", a human reader knows that the action expressed by "realizar diligencias" somehow will lead to a capture, or that the capture is the goal of the "diligencias". Capturing something involves getting control of it, and we know that before we can get control of an object, we have to know where it is and we have to find it. This indicates that perhaps "realizar diligencias" refers to some sort of finding. But when police are trying to find something in order to get control of it, they usually do a formal type of search, or an investigation. Therefore, we know that in this case, the phrase refers to a police investigation.

What we would like, then, is to devise frame selection rules which parallel this line of reasoning. In essence, this line of reasoning is a refinement process. Each inference limits further the type of action to which "realizar diligencias" could refer. At first, we know nothing more than simply that the phrase is referring to some action. Then, it is limited to be a type of FIND. Finally, we can infer that it is an INVESTIGATION.

The frame selection method used in the MOPTRANS system parses this example using a similar refinement process. The frames in MOPTRANS are arranged hierarchically, from most vague to most specific. The dictionary definitions of words consist of pointers into this hierarchy. The level of specificity at which the definition of a word points into the hierarchy depends on how vague or ambiguous the word is.

Since "realizar diligencias" is a very general phrase, it points into the hierarchy at a very general level, to the frame ACTION. ACTION falls in the hierarchy as is shown in Figure 2. The dotted lines connecting the concepts ACTION, FIND, etc., represent IS-A links. In addition to this hierarchical information, GET and POLICE-CAPTURE, which are script-like structures, provide information about stereotypical sequences of events.

All of the concepts in this IS-A hierarchy have case frames, specifying the prototypical fillers for various slots, such as ACTOR, OBJECT, etc. For example, the case frame for FIND indicates that its ACTOR should be a PERSON, its OBJECT should be a PHYSICAL OBJECT, and its RESULT should be a GET-CONTROL. The case frame for POLICE-INVESTIGATION indicates that its ACTOR should be an AUTHORITY, its OBJECT should be a CRIMINAL, and its RESULT is an ARREST.

```
              ---- ACTION --------
              |                  |
             FIND           GET-CONTROL
              |                  |
     POLICE-INVESTIGATION     ARREST

GET            = FIND + GET-CONTROL
POLICE-CAPTURE = POLICE-INVESTIGATION + ARREST
```

Figure 2: Memory structures for police investigation example

With this hierarchical memory organization and scriptal knowledge, very general rules can be used to select the correct frame for this example, along the same lines as I suggested that a human reader would follow. The dictionary definition of the word "capturar" points to the concept GET-CONTROL in the hierarchy above. From the event sequence GET, we know that GET-CONTROL is often preceded by the event FIND. Since the story says that an ACTION, "diligencias", precedes the GET-CONTROL, we can infer that the action is probably a FIND. This suggests the following general inference rule: If a scene of a script is mentioned in a story, then other scenes of the same script can be expected to be mentioned. Thus, if an abstraction of another scene of the script is mentioned, we can infer that the abstraction actually is the other scene. In more concrete terms, in this example GET-CONTROL is a scene of the script GET. Another scene of GET is the scene FIND. "Realizar diligencias" refers to an abstraction of the concept FIND, namely ACTION. Since GET-CONTROL was mentioned, indicating that other scenes of the script GET are likely to be encountered, we can infer that the ACTION is actually a FIND, since ACTION is an abstraction of FIND.

Put more precisely, this line of inferencing can be expressed in the following rules:

SCRIPT ACTIVATION RULE: If an action which is part of a stereotypical event sequence is activated, then activate the stereotypical event sequence, and expect to find the other actions in that sequence.

EXPECTED EVENT SPECIALIZATION RULE: If a word refers to an action which is an abstraction of an expected action, and the slot-fillers of the action meet the prototypes of the slot-fillers of the more specific action, then change the representation of the word to the more specific expected action.

Next, consider how we can infer that the FIND is a POLICE-INVESTIGATION. First, in the story the ACTOR of the FIND is the POLICE. One piece of knowledge that we have about POLICE is that often they are the ACTORs of POLICE-INVESTIGATIONs, since that is part of their job. Then, since the IS-A hierarchy tells us that POLICE-INVESTIGATION is a refinement of the concept FIND, we can infer that in this story, the FIND is most likely a POLICE-INVESTIGATION.

This suggests the following inference rule:

SLOT-FILLER SPECIALIZATION RULE: If a slot of concept A is filled by concept B, and B is the prototypical filler for that slot of concept C, and concept C IS-A concept A, then change the representation of concept A to concept C.

In this case, concept A is FIND, and concept B is the POLICE. The POLICE are the prototypical ACTORs of concept C, a POLICE-INVESTIGATION. Since FIND is above POLICE-INVESTIGATION in the IS-A hierarchy, then we can conclude that FIND in this case refers to POLICE-INVESTIGATION.

Thus, I have suggested three general inference rules, the script activation rule, the expected event specialization rule, and the slot-filler specialization rule; which can perform the disambiguation of "realizar diligencias" in the example above. These rules require the organization of knowledge structures in a hierarchical fashion, so that they can use this hierarchy to guide the refinement of concepts. They also require the existence of event sequences (scripts) in memory, to provide expectations as to what actions are likely to occur together in stories.

Given these rules, frame selection for the police investigation example proceeds as follows: first a general representation is built for "realizar diligencias"; simply, the concept ACTION. Then, the ACTOR of ACTION is filled in with the concept AUTHORITY, since "policia" is the subject of the verb "realiza". Next, the concept GET-CONTROL is built to represent the word "capturar". This also causes the event sequence GET to be activated, because of the Script Activation Rule above. This, in turn, causes the concept ACTION to be changed to FIND, due to the Expected Event Specialization Rule. Finally, since the ACTOR slot of FIND is filled by AUTHORITY, and since the prototype of the ACTOR slot of POLICE-SEARCH is AUTHORITY, and since POLICE-SEARCH is the only concept under FIND in the IS-A hierarchy for which this is true, the concept FIND is further refined to POLICE-INVESTIGATION, due to the Slot-filler Specialization Rule. filled by AUTHORITY, and since the prototype of the ACTOR slot of POLICE-SEARCH is AUTHORITY, the concept FIND would be changed to be POLICE-INVESTIGATION because of the slot-filler specialization rule above.

IV Conclusion

I have presented three general inference rules, the Script Activation Rule, the Expected Event Specialization Rule, and the Slot-filler Specialization Rule, which can be used to select frames for very vague words such as "diligencias". These rules draw on information from a hierarchically organized conceptual memory, which provides knowledge about abstractions of events and sequences of events.

This frame selection method is in contrast to the lexically-based and frame-based methods discussed earlier. In the lexically-based method, since at least one disambiguation rule is needed for each sense of an ambiguous word, very vague or general rules require a very large number of disambiguation rules. The frame-based method is limited in that selection of an initial frame must be done by some other method. The disambiguation method which I have presented here does not suffer from these limitations.

MOPTRANS's frame selection method is most similar to the method used in FRUMP [DeJong79], involving the use of SSIDTs. However, the MOPTRANS method has several advantages over DeJong's method. First, text does not need to be represented in terms of Conceptual Dependency primitives at the beginning of the frame selection process. In DeJong's system, "hacer diligencias" would initially need to be represented in CD. Second, although the organization of frames in a hierarchy serves much the same function as the discrimination nets used by DeJong, the traversal of the hierarchy in the approach I have presented is less ad hoc than in DeJong's system. A small number of inference rules perform the traversal of the hierarchy, in conjunction with the definitions of the frames in the system. In FRUMP, arbitrary tests were used to determine what path in the discrimination net should be followed.

The frame selection process which I have described here is also similar in some respects to the Incremental Description Refinement process used in RUS [Bobrow&Weber80]. In this system, a taxonomic lattice [Woods78] is used to refine the semantic interpretation of a sentence as it is being parsed. The refinement process is similar to the frame selection method I have described here, in that it relies on the structure of the hierarchy to provide it with the information needed to discriminate to more specific concepts in the hierarchy. For example, the sentence "John ran the drill press" was parsed in this system using a taxonomic lattice containing nodes RUN-CLAUSE, PERSON-RUN-CLAUSE, RUN-MACHINE-CLAUSE. The parser refined its semantic interpretation of the sentence from RUN-CLAUSE to the more specific PERSON-RUN-CLAUSE and finally RUN-MACHINE-CLAUSE as more information was provided by the parse of the sentence.

Although the refinement processes in RUS and MOPTRANS are similar, the content of the nodes in the hierarchies used in the two systems is completely different. The nodes in the taxonomic lattice in RUS are in no way independent of lexical items, since nodes represent both semantic and syntactic functions of lexical items. Thus, the system would presumably have a separate node, RUNNING-OF-MACHINE-NOUN-PHRASE, to represent a noun phrase like "The running of the drill press", even though the noun phrase has virtually the same meaning as RUN-MACHINE-CLAUSE. This is in contrast to the nodes in the hierarchy used in MOPTRANS, which are meant to be elements in a conceptual representational system, and therefore independent of the specific lexical items which built the representation.

REFERENCES

[1] Bobrow, D.G., Kaplan, R.M., Kay, M., Norman, D.A., Thompson,H., and Winograd, T,. "GUS, a Frame Driven Dialog System." *Artificial Intelligence* 8:1 (1977).

[2] Bobrow, R.J., and Weber, B.L. Knowledge Representation for Syntactic/Semantic Processing. *Proc. AAAI-80,* Stanford University, August, 1980, pp. 316-323.

[3] DeJong, G. *Skimming Stories in Real Time: An Experiment in Integrated Understanding.* Ph.D. Thesis, Yale University, May 1979.

[4] Lebowitz, M. *Generalization and Memory in an Integrated Understanding System.* Ph.D. Thesis, Yale University, October 1980.

[5] Lytinen, S.L., and Schank, R.C. "Representation and Translation." *Text* 2:1/3 (1982), pp. 83-112.

[6] Minsky, M. "A framework for representing knowledge." In *The Psychology of Computer Vision,* P. Winston, Ed.,McGraw-Hill, New York, 1975, ch. 6, pp. 211-277.

[7] Riesbeck, C. "Conceptual Analysis." In *Conceptual Information Processing,* North-Holland, Amsterdam, 1975.

[8] Schank, R.C. and Abelson, R.. *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.

[9] Small, Steven. *Word Expert Parsing: A Theory of Distributed Word-based Natural Language Understanding.* Ph.D. Thesis, Department of Computer Science, University of Maryland, 1980.

[10] Woods, W.A. "Taxonomic Lattice Structures for Situation Recognition." In *TINLAP-2,* July, 1978.