STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO AIM-211
STAN-CS-73-383

NATURAL LANGUAGE INFERENCE

BY

Yorick Wilks

SUPPORTED BY

ADVANCED RESEARCH PROJECTS AGENCY
ARPA ORDER NO. 457

August 1973

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY
MEMO NO. AIM-211

COMPUTER SCIENCE DEPARTMENT
REPORT NO. CS-383

AUGUST 1973

NATURAL LANGUAGE INFERENCE

by

Yorick Wilks

ABSTRACT: The paper describes the way in which a Preference Semantics system for natural language analysis and generation tackles a difficult class of anaphoric inference problems (finding the correct referent for an English pronoun in context): those requiring either analytic(conceptual) knowledge of a complex sort, or requiring weak inductive knowledge of the course of events in the real world. The method employed converts all available knowledge to a canonical template form and endeavors to create chains of non-deductive inferences from the unknowns to the possible referents. Its method of selecting among possible chains of inferences is consistent with the overall principle of "semantic preference" used to set up the original meaning representation, of which these anaphoric inference procedures are a manipulation.

This paper owes much to discussions with Annette Herskovits, Bill Simpson, Ken Colby, Bruce Anderson and Horace Enea. The mistakes are all mine of course.

This research was supported by the Advanced Research Projects Agency, Department of Defense(SD 183), USA.

INTRODUCTION

The paper describes inferential manipulations of a representation of the meaning structure of natural language. It differs from previous descriptions (1, 2, 3, 4) of this semantics-based system, which have concentrated on the representation itself, and above all on the procedures by which the representation is produced from input sentences and paragraphs in English. In this paper I assume that structure of representation, except for a brief recapitulation, and concentrate on operations upon it for the solution of a class of difficult anaphora problems.

The system described is part of a running system for understanding and translating natural language on the PDP6/10 at the Stanford A. I. Laboratory, programmed in MLISP and LISP1. 6. I shall not in anyway stress the translation-into-French aspect of the work, but its presence provides a continual empirical check of the adequacy of the inferences and "understanding" described here.

The earlier emphasis on the construction of the linguistic base is, I believe, fully justified. The present system is, to my knowledge , the most comprehensive producer of meaning structrures for generalized natural language available at present in terms of implementation , vocabulary, disambiguation of many-sensed words and referents, etc. Moreover, as I have argued elsewhere (2), it is not the implementation of a conventional theory from linguistics, but is one with somewhat different principles of content.

In what I call its basic mode, the system already resolves anaphoras depending on superficial conceptual content of the text words. This is does in the course of setting up the initial representation. I shall call these type A anaphoras. For example, in "Give the bananas to the monkeys, although they are not ripe. They are very hungry", the system in its basic mode would decide that the first "they" refers to the bananas and the second to the monkeys. It can do that simply from what it knows about monkeys getting hungry because they are animate, and bananas having phases like ripeness because they are plants. All this information is, one might say colloquially, part of the superficial meaning of "banana "and "monkey".

This paper describes an "extended inference mode" of the system that tackles two other kinds of anaphora example that I shall call types B and C. Consider the correct attachment of "it" in "John drank the whisky from the glass, and it felt warm in his stomach". It is clear that the pronoun should be tied to "whisky" rather than "glass", but how it is to be done is not immediately obvious. Analysis of the example (see below) suggests that the solution requires , among other things, an inference equivalent to the sentence "whatever is in a part of X is in the X".

Anaphoras like the last I shall call type B, in that the inferences required to resolve them are analytic but not superficial. By analytic I simply mean that the sentence above, about parts and wholes, is logically true and not in any clear sense a fact about the real world (but rather about the meanings of words). What is meant by "superficial" in the distinction between types A and B will become clear after a discussion of the meaning formalism employed here.

I shall also discuss type C anaphoras, which require inferences that are not analytic at all, but weak generalisations (often falsified in experience) about the course of events in the world. Yet their employment here is not in any sense a probabilistic one. In "The dogs chased the cats, and I heard one of them squeal with pain", we shall, in order to resolve the referent of "one" (which I take to be "cat" not "dog"), need a weak generalisation equivalent to "animate beings pursued by other animate beings may be unpleasantly affected". Such expressions are indeed suspiciously vague, and a reader who is worried at this point should ask himself how he would explain (say, to someone who did not know English well) the way he knew the referent of "one" in that sentence. It can hardly be in virtue of a particular fact about cats and dogs because the same general inference would be made whatever was chasing and being chased. I shall be surprised if he does not come up with something very like the inference suggested, and it may be the nature of natural language itself that is worrying him.

The inferences for type C, then, are general expressions of partial information(in McCarthy's phrase) and are considered to apply only if they are adequately confirmed by the context. What I mean by that will become clear in the course of what follows, but in no case do these expressions yield deductive consequences about the future course of the world. Indeed, they would be foolish if they did because the world's course cannot be captured in that way. In the whisky example above, it might have been his earlier dinner that made him feel good.

BRIEF RECAP OF THE SYSTEM'S BASIC MODE OF ANALYSIS

In its basic mode, the system fragments texts (into phrase/clause like items) and attaches a template to each. A template is a canonical form of connectivity of semantic formulas as follows(where a formula is a complex item, to be described, corresponding to a sense of an input word):

```
      F1↔↔↔↔↔F2↔↔↔↔↔↔F3
    /  |  \     \       |  \
F11 F12 F13   F21   F31   F32
```

F1, F2, F3 are the principal formulas of the template and are always agent, action and object (in that order), though any of them may be a dummy in any particular example. (F11, F12, F13) is a list of formulas dependent on main formula F1 etc. It should be said, in view of other current uses of "template", that it is not a surface item at all, but a formal underlying meaning representation. Moreover, it does not function within a crude pattern matching technique, such that if some text fragment has no templates matching it it is thrown away, as it were. Special routines are called in such situations to construct an appropriate template item. All this basic material is set out in earlier papers.

The structure of formulas is explained below in some detail. In brief a formula is a nested list(a binary tree in fact) of semantic primitives called elements(expressed here as LISP atoms). Each such formula expresses a sense of at least one an English word.

Let me give an example of a template structure at this point by using the following simplifying notation: any English words in square brackets [] stand for the meaning representation of those words in the Preference Semantics system. This device is important in the exposition of the material in that the content of the coded forms can be seen immediately, whereas the complex coded forms themselves would be as hard to read as ,say , a sentence read a word at a time. But it is important to restate that the rules and formalisms expressed within [] are really structured primitives, and that their tasks could not be carried out, as some still seem to believe, by massaging the language words themselves to stand for their own meaning representation.

So then, the template for "The black horse passed the winning post easily" could be written (ignoring any ambiguity problems for the moment):

```
[horse]↔↔↔↔[passed]↔↔↔↔↔↔[post]
   ↑            ↑           ↑
[the black]  [easily]  [the winning]
```

If any or all of the agent, action or object formulas are missing, the template node(s) is filled if with a dummy element DTHIS. Thus a template for "The old house collapsed" could be written:

```
[house]↔↔↔↔[collapsed]↔↔↔↔DTHIS
   ↑
[the old]
```

In the case of structures like prepositional phrases, we consider the preposition represented as a pseudo action, and the whole template as having a dummy agent. Thus for "at the Derby", we have:

```
DTHIS←←←←[at]←←←←←[Derby]
                      ↑
                    [the]
```

The representation of a text(composed of fragments) is then a network of these template networks. The templates are interconnected by case ties. The notion of case is discussed in more detail below, but for the moment a case can be thought of as tying one template to some particular node in another template by a link of a certain type : namely the case type, which specifies the sort of dependence the former template has on the latter. In the sentence "He lost his wallet / in the subway" (fragmented at the stroke) we might say that the second fragment of the sentence depends on "lost" in the first, and that the dependence is the locative case. Thus in the representation, the template for the second fragment would be tied to the central, action, node of the first, by a link labelled LOCA. The node on the first template to which the case tie ties is called the mark of the second template.

Type A anaphora is dealt with adequately within this framework , roughly(since the purpose of this paper is not to describe the basic mode of operation of the system) because we get a denser network of links by considering the formula for the appropriate referent substituted for the problem pronoun than with its rivals. A link is considered inserted, or strengthened , in the network, when a preference is satisfied. So, if we think of the formula for "ripe" as expressing a preference for application to plants, we see why a denser network arises in the above example for correct solution, rather than for one equivalent to "ripe monkeys". The way in which formulas express preferences of various sorts is described below.

Once solved, these type A anaphoras also constitute links between templates, from the pronoun variable to its correct referent. Thus the compressed list form of the whole representation obtained from the basic mode is:

(CASE MARK ANAPHORA F1 F2 F3 (F1 dependents)(F2 dependents)(F3 dependents))

This initial form of representation of a text paragraph is called its IREP(Interlingual REPresentation), or "semantic block". No emphasis has been placed here on syntax analysis of the input, and a reader who consults (1, 2, 3 or 4) will see that all of conventional grammar analysis has been done in the course of setting up this form of representation. An example of such an IREP, for the monkey-banana example, is given below as computer output.

Having sketched in the basic mode and its representation, we can now get to our muttons and sketch in the extended inference mode that is the heart of this paper.

QUICK SKETCH OF THE "EXTENDED INFERENCE MODE"

The extended inference procedure is called whenever the basic mode cannot resolve an anaphora between two or more candidates by semantic link density. In the example about John and his stomach, density techniques have no way to decide whether the glass or the whisky is in his stomach. On a basis of preferred agents and objects of actions (what I have referred to as superficial conceptual information) both are equally good candidates. The extended inference procedure is called , and if it succeeds it returns a solution to the basic mode which then continues with its analysis.If it too should fail to reduce the number of candidates to one, then the top level of the system tries to solve the problem by default, or what a linguist would call focus. Roughly, that means : assume that whatever was being talked about is still being talked about. So, in "He put the bicycle in the shed and when he came back next week it was gone", neither density criteria, nor the extended inferences to be described here, will help at all. So the system may as well assume ,in this limited context , that the bicycle is still the focus of attention, and hence the reference of "it".

Consider again the following sentence after all the basic mode's routines have been applied:

[1: John drank the+ whisky / 2 DIRE : DTHIS from a+glass / 3 : and it felt warm /IN 4: DTHIS in his+stomach]

Because of the square brackets , this item is a template representation. The case names DIRE(direction) and IN(containment) indicate the dependencies of templates 2 on 1, and 4 on 3 , respectively. The DTHISs are dummies added to fill out the canonical triplet for in cases of missing agents,objects etc. Further assume that the "his" has been tied to "John" by the basic mode , and presents no problem of analysis. And assume too that the basic mode provided a list of "candidates" for the reference of "it"("whisky" and "glass"). because if there had not been such a list of more than one candidate the routine under description would not have been called into play.

Extractions are then made from each template in turn,and if and only if a template contains a representation of either an answer word or the variable pronoun itself. An extraction is the unpacking of every possible case tie : both those in the action (second)formula of the template and those labelling ae link to other templates. In this example we obtain the following extractions: which are template like forms as follows (where the first digit refers to the fragment #, the second to the extraction, and "+" links words with a single formula):

11: [whisky (IN in) John +part]

12: [whisky (DIRE to) John+part]

21: [whisky (DIRE from) a+glass]

41: [?it (IN in) his+stomach]

So, in this informal representation we have acquired new template-like objects that express, in canonical form, new analytic information extracted from the existing templates, and from which new inferences can be made. It is postulated that the generation of this inexplicit information from the deeper levels of the formulas is essential to the process of understanding. These new forms differ from standard templates only in that their second node, or pseudo-action, has had a case name CONSd onto whatever the node was before. Note here that the form (IN in) is not redundant since the IN locates the case precisely as containment, while the English preposition can indicate many cases other than containment, as in "in five minutes".

We can describe how these particular extractions were made, even in the absence of any detailed knowledge of the structure of formulas:

11 has been derived from the template for "John drank the whisky" because from the structure of the formula for "drink" it follows that the liquid drunk is subsequently inside the drinker. This is because, when making up the formula for the action"drink", we express in it that the action consists in causing a liquid to be inside the agent of the action.

Again, 12, is inferred because the same formula specifies that the liquid enter the drinker through a specific part of the drinker (his mouth, of course).

21 is inferred from the direction case of the second template, whence we know that it was the whisky that was moved from the glass.

Lastly, 41 is inferred from the direction case of the fourth template, because whatever the referent of it is, it is also in John's stomach.

Let us see where we are: we have obtained new template items that yield assertive information, but did not appear in the original text. (As we shall see in the detailed treatment below, some of the above are obtained from extraction, more strictly defined, and some from what I shall call "repacking the semantic block".)

In the pool for inference procedures we now have the original templates that mention either the variable pronoun or the possible "answer" referents, plus the extractions. We also have access to an inventory of Common Sense Inference Rules (CSIRs) which are of the form [T1 → T2], where T1 and T2 are T-forms, that is , templates or extractions.

We now try two strategies in turn: first we try a zero-point strategy, which is to try to identify an answer template(or extraction) and a variable template(or extraction) without the use of CSI rules.

The general assumption here, and it is a strong psychological assumption is that in order to resolve these painful ambiguities the understanding system is going to use the shortest possible chain of inferences it can. And a zero-point strategy will, as it were, have no length at all (in terms of a chain of CSI inferences) and so if it works, it will always provide the shortest chain.

This strategy is adequate for the example under discussion, because we can (under a suitable definition of matching) identify extractions I1 and 41, and identify ?it and the whisky, and we are home. This was the solution of a B type anaphora, requiring only analytic, necessarily true conceptual information. It should be noticed that some C type anaphoras (defined earlier by the need of weak inductive information for their solution) can also be solved by the zero point strategy, because some extractions, and in particular those from the goal case (see below) , are inductive and non-analytic.

If the zero-point strategy fails, we bring down all the CSI rules that contaiin an action subformula occurring in an answer or problem T-form in the pool, and attempt to find the shortest chain that leads from some answer to some variable.

Thus, in the sentence "The soldiers/ fired+at the women/and I saw several fall", we extract a form equivalent to [soldiers strike women] , since we can tell from the formula for "fired+at" that the action is intended to strike the object of the action. We are seeking for partial confirmation of the assertion [?several fall DTHIS], and such a chain is completed by the rule [X strike Y] → [Y fall DTHIS] , though not by a rule equivalent to, say, [X strike Y] → [Y die DTHIS], since there is nothing in the sentence as given to partially confirm that rule in a chain, and cause it to fit here. Since we are in fact dealing with subformulas in the statement of the rules, rather than natural language words , "fitting" means an "adequate match of subformulas".

It is conceivable that there would be an, implausible, chain of rules and extractions giving the other result, namely that the soldiers fall: [soldiers fire DTHIS] ∧ .[X fire DTHIS] → [Y fire+at X] → [Y strike X] → [X fall DTHIS] etc.,based on the assumption that things that fire guns get fired at ("...he who lives by the sword shall perish by.....") . But such a chain would be longer than the one already constructed and would not be preferred.

MORE ON THE BASIC MODE

Formulas

Formulas are structures corresponding to senses of words, expressing their meaning. Much of the body of this paper is concerned with the manipulation of such structures, and the extraction of information from them, so it is important to have some general idea of their construction and interpretation.

Formulas are binary trees , expressed as lists, of semantic elements, punctuated by right and left brackets. The elements are either case elements, or actions such as CAUSE, STRIK, CHANGE, or items such as THING, MAN, EVNT. I am using, as examples here, element names that are self-explanatory Anglo-Saxon monosyllables, but there are about 75 in use, and some need informal explanation, such as GRAIN, used to mean "structure". There are also elements like KIND indicating qualifiers, and elements (indicated by and initial *) that stand for classes of other elements: such as *ANI(animate) to cover MAN, BEAST and FOLK(human groups). In addition , most elements have a negated form NOTX, where X is the element name. I assume here that the use of linguistic primitives of this general sort, that are not logical predicates, needs no special defense at this point.

The most important element in a formula is its rightmost, called its head. This indicates what general sort of item, or action , or type the word sense expressed corresponds to: for example, any word sense corresponding to a human being will have MAN as its head.

Since formulas are binary trees of unlimited depth, they can be continually subdivided into pairs of elements and subformulas, down to the level of the semantic primitive elements. This process is equivalent to either building up the formula, or decomposing it while interpreting it. At each stage there is a dependence of the left half of any pair on the corresponding right : this dependence is either of

(a)item (agent or object) on action

(b) qualifier on item or action

(c) of case specification on action or item.

In any particular example the interpretation is unambiguous, once we know the range of functions of the elements in play. So, the subformula (MAN WRAP) always means "a human envelops something"because WRAP is always an action when in the right hand position (and always a qualifier when in the left,dependent, position. It is never an item) and MAN is a agent not an object in this example (it would be an object if in the representation of "a human being is enveloped")because agents of actions may be unmarked , though objects are never unmarked. Conversely (WRAP THING) is a container, since WRAP is always a qualifier when in the left hand, dependent, position in a pair

An important notion is tat of the semantic preferences that formulas can express. Consider the formula for "grasp"of objects:

"grasp"(action1) → ((*ANI SUBJ)((*PHYSOB OBJE)(((THIS (MAN PART))INST)(TOUCH SENSE)))))

The case elements SUBJ and OBJE occur at the top levels on the left of the formula, and at that level in an action formula they express the preferred agent and object of the action concerned.

Thus, grasping, in this sense is an action preferably done by animate beings (*ANI) to physical objects (*PHYSOB), and consists in a act of sensing, by touch, and done with an instrument (INST is the case element) which is a part of the body. When I say "prefers" here, I mean that , if the preferred agent or object cannot be found, a template is constructed with whatever is available. Thus, "The robot grasped the block" would never be rejected; it would only be less preferred than any possible competing interpretation that had an animate agent. I have argued in (2) that this approach to rules of formation has unsuspected consequences for linguistic theory.

But we would also expect another formula to be available for "grasp", one such as:

"grasp"(action2)→ ((*HUM SUBJ)((SIGN OBJE)(TRUE THINK)))

In this sense, we have an action, preferably done by human beings to signs (which could be ideas, principles etc), namely of thinking them to be true, or adequate, signs.

The procedures of the basic mode always fit this last formula into a template structure for "He grasped the principles", and the other formula for "grasp" into the template for "the boy grasped the toy", by means of the preference and semantic density techniques described in earlier papers. These preferences for agent and object are part of the "superficial conceptual information" referred to earlier.

A few other rules will help to clarify the notion of "knowing our way round a formula " when interpreting it:

Agents are implicit (need not be specified by SUBJ case) unless(1)they occur at the top level in an action formula as described above, or they attach to the head of a formula , as in:

"patient"(item) → ((NOTPLEASE FEEL)(SUBJ MAN))

Here , the normal order, of agents being to the left of (= dependent on) the corresponding action, is violated, since MAN is the agent for FEEL, while at the same time being the head of the whole formula.

This violation of order in search is indicated by also violating the order restriction that normally makes the SUBJ case element the governor (right hand member) of the pair in which it occurs. The corresponding rule of analysis is "On encountering SUBJ as dependent, expect action for the agent to follow to the left".

Objects, however, are never implicit. Moreover, an object is considered an object of all actions to its right. This enables us to express the important notion of real and apparent agents of actions. So for example in:

"fire+at" (action)→

((*HUM SUBJ)((*ANI OBJE)((STRIK GOAL)((THING MOVE)CAUSE))))

This action (done preferably by human beings to animate beings) is one of causing a thing to move (the bullet) with the aim (GOAL case) of striking something. Since *ANI is the object of all actions to its right, it the object not only of CAUSE, but also of STRIK. Hence the striking is also of thesame animate being. Moreover, THING(the bullet) is internally the agent of MOVE, not the object of CAUSE, which is correct as far as the meaning of "fire+at" is concerned.

Cases

At present we operate with a distinction system of ten cases, which are listed below, together with (in capital letters) the semantic elements that represent them, the questions that define them, and examples of subformulas expressing them. Defining a case is a tricky matter, but the question method is reasonably adequate. Note that the subformula examples are of those parts of a formula that would express that notion AS PART OF THE MEANING OF A WORD. The subformulas are not , of course, how the system would express the quoted words if encountered in a text, when they would be represented by a template.

recipient: FOR — "for a woman" → ((FEM MAN)FOR)
what/who to? what/who for?

instrument: INST — "with a stick" → ((LINE THING)INST)
what with? by what means?

direction: *DIRE(see below), TO, FROM ,UP
"from the top" → ((UP POINT)FROM)
where to where from? at what? out of where? by what?

possessive: POSS — "owned by a man" → ((MAL MAN)POSS)
who owns the thing mentioned?

location: LOCA "at that time" → ((THIS(WHEN POINT))LOCA) when? where? where at? by what? in what time? near what? at what time? during when? before when?

containment: IN "in a glass" → (((((FLOW STUFF)OBJE)WRAP)THING)IN) in what?

source: SOUR "out of wood" → ((PLANT STUFF)SOUR) outof what? from what?

goal: GOAL
"so as to strike a woman"→
(((FEM MAN)OBJE)STRIK)GOAL)
to what end ? for what purpose?

accompaniment: WITH
"without a glass"→(((((FLOW STUFF)OBJE)WRAP)THING)NOTWITH)
accompanied by what/who? with what/whom? without what/whom?

subject: SUBJ who did this?

object: OBJE who/what was this done to?

Certain cases above have negative forms leading to additional elements NOTFOR, NOTPOSS, NOTIN, NOTWITH.

Case elements have two functions, and occur in two sorts of constructions: formulas and IREPs. In formulas they express part of the meaning of a word sense. Thus in

"drink" → (((WRAP THING)IN)(((FLOW STUFF)SOUR)THING))

we see that a drink has a liquid source (FLOW STUFF) , and is in a container (WRAP THING) The other function of these elements is, as already explained, the name of the tie between the template for some fragment and some part of another template.

*DIRE is the name of the class of direction case elements (TO and FROM) and it occurs only as the indicator of the case of a fragment, never in formulas. Conversely POSS occurs only in formulas, never as the indicator of a fragment case.

Case information is only included in a formula when it is specific: when we can say what aspect of the case is involved. In the formula for "pour", for example, we include a direction specification for downward ((NOTUP POINT)TO). However, inthe formula for "move" we do not include the element TO or FROM, even though movement must in fact be in some direction, since have no reasonable expectation about it as we do with "pour". Sentences containing "move" may very well go on to specify the direction involved, but its association with "move" is conceptually arbitrary and we cannot expect any confirmation of expectations that would, say, resolve ambiguities. In this respect the system differs from other systems that do create case expectations for wide classes of actions, which are essentially unspecific, as in this example, and so we would claim unhelpful semantically.

The IREP, or semantic block representation

What follows is an example of an IREP for a pair of English sentences. The format of the block is the list structure described earlier, as the result from the basic mode of operation. The only difference from that format is the presence in it of the stereotypes from which French is subsequently generated(see Wilks and Herskovits 4, Herskovits 5). The French, as generated from the block, is written above the print out of the block itself for diagnostic purposes. The appropriate context-sensitive stereotypes are drawn into the block during analysis, along with the formulas. The process of generation is then a recursive unwrapping of the block.

GIVE BANANAS TO THE MONKEYS ALTHOUGH THEY ARE NOT RIPE! THEY ARE VERY HUNGRY.

DONNEZ DES BANANES AUX SINGES BIEN QU' ELLES NE SOIENT PAS MURES ! ILS ONT TRES FAIM.

```
(((EX GIVE BANANAS) ((EX NIL NIL ((IMPCL)))) 4 (((THIS DTHIS)
DUMTHING) ((PRES ((*ENT OBJE) GIVE)) GIVE (DONNER)~ ) ((MUCH (((*ANI
SUBJ) ((TASTE SENSE) WANT)) (OBJE PLANT))) BANANAS (FEMI BANANE)) NIL
NIL NIL)) ((PTO THE MONKE~ YS) ((PTO (GIVE) RECI ((&PREOB A)))) 6
(((THIS DTHIS) DUMTHING) ((THIS PDO) PTO NIL) ((THE (MUCH ((MAN LIKE)
BEA~ ST))) MONKEYS (MASC SINGE)) NIL NIL NIL)) ((ALTHOUGH THEY ARE
NOT RIPE !) ((ALTHOUGH (GIVE) CONC (BIEN QUE (SUBC~ L)))) 1 (((MUCH
(((*ANI SUBJ) ((TASTE SENSE) WANT)) (OBJE PLANT))) (THEY BANANAS)
((PRON 6 MASC PLUR))) ((NPRES ~ (BE BE)) ARE ((IS_OBJECT HUNGRY)
AVOIR (DIROB Q FAIM)) ((IS_OBJECT THIRSTY) AVOIR (DIROB Q SOIF))
((IS_OBJECT AF~ RAID) AVOIR (DIROB Q PEUR)) (ETRE)) (((PLANT POSS)
((*ANI (CAN USE)) KIND)) RIPE (MUR)) NIL NIL NIL)) ((THEY ARE~ VERY
HUNGRY /. ) ((NIL NIL NIL ((INDCL)))) 1 (((THE (MUCH ((MAN LIKE)
BEAST))) (THEY MONKEYS) ((PRON 6 MASC PLUR~ ))) ((PRES (BE BE)) ARE
((IS_OBJECT HUNGRY) AVOIR (DIROB Q FAIM)) ((IS_OBJECT THIRSTY) AVOIR
(DIROB Q SOIF)) ((I~ S_OBJECT AFRAID) AVOIR (DIROB Q PEUR)) (ETRE))
(((*ANI POSS) ((((TASTE SENSE) WANT) STATE) KIND)) HUNGRY (AFFAME~ ))
NIL (((MUCH HOW) VERY (TRES))) NIL)))
```

THE IMPLEMENTATION OF THE EXTENDED MODE

There are three parts to the extended inference mode: the REPACK routine that takes the IREP block and repacks it; the EXTRACT routine which produces extractions, new knowledge not explicit in the text analysed; and INFER which tries to link an answer T-form to one expressing a problem variable, that is, a text pronoun giving trouble.

The REPACK routine.

This attempts to replace dummy nodes in the IREP wherever possible before handing the whole representation to the extraction procedure. This replacement is itself a complex form of inference, sometimes as complex as the inference routines on which we are concentrating here. However, there is no pretense that these procedures are mutually organized in any hierarchical fashion, so we may legitimately concentrate our attention one a single process in this way. The degree of dummy rewriting done by REPACK, in the construction of a new block IREPR from IREP, varies with particular action cases.

If we look back at the informal extractions done from the "John drank the whisky. . . " example , we will see that the new T-forms 21 and 41 are actually obtained by filling in a dummy agent in some template from a node in another template. Thus from [2 DIRE : DTHIS from a+glass] we obtained the new T-form, numbered 21, [whisky (DIRE from) a+glass]. This was done by filling the dummy agent node of the template for "with a glass" with the formula for "whisky", and shifting the direction case marker into the pseudo-action. This is a repacking, not an extraction proper, since the T-form obtained simply replaces a template "assertion" already in the representation. As we shall see, a true extraction is a new T-form altogether.

Let us now distinguish replaceable and unreplaceable cases roughly as follows The dummy agent in the second (instrumental) template for "He hit his father/with a club" cannot be replaced to yield any form equivalent to [father (INST with) a+club]. So we may say that the instrumental case in unreplaceable. But the dummy agent in the second(recipient) template of "He bought the flowers / for his mother" can be replaced to yield a form for [flowers (RECIP for) his+mother], and so recipient case is replaceable, and is replaced by the operation of REPACK.

At the top level REPACK can be written in LISP as:

REPACK(IREP)→ IREPR

The EXTRACT routine

EXTRACT takes each T-form, or template-like item, in IREPR in turn and replaces it, modified if necessary, in a new block IREPE, followed by T-forms extracted from it. At present, extractions are only made from templates that contain either one of the possible answers, or one of the variables of the problem. The former are templates containing a formula for a word on the list ANS, the latter are templates one of whose nodes is (QUERYMARK THIS). Any templates not contain either an ansuer or a variable are simply transferred unchanged from IREPR to IREPE.

Thus the general form of the extraction routine at the top level is

EXTRACT(IREPR ANS) → IREPE

Taking each template in turn, we first consider those processes that modify it, and then those which produce new T-forms from it. In the first category come manipulations to do with negation, and with the SUBJE and OBJE cases.

If an agent object formula is negated, the negative item in its formula is removed and the head of the corresponding action formula is negated, because all the subsequently applied inference rules canonically negate the action. Thus, in [] notation, we would achieve by this procedure the coded equivalent of

[John drank no+gin]→[John not+drink gin]

Each agent and object formula is then scrutinized by the question "does it satisfy the preference expressed by the corresponding action. If it does not, does any "of-phrase" qualifier of it do so instead". If so, replace the agent or object by that "of-phrase" qualifier as the true agent or object.

Thus:

[John drank a+glass+of+wine] → [John drank wine]

[A+group+of+women drank wine] → [Women drank wine]

The main phase of EXTRACT takes the action formula of a template and moves leftwards through it seeking case heads (other than SUBJ and OBJE). If it finds one, it asks is it replaceable, and , if it is, EXTRACT looks at subsequent fragments to see if REPACK has already replaced it. If it has been replaced it is forgotten, thus avoiding the same case information being extracted twice. It detects that REPACK has made such a replacement by finding the case name itself in the pseudo-action of a succeeding template, and a replaced dummy as the corresponding agent.

With the goal case, for example, the dependent of the case element becomes the action of the new T-form. In this case, as with every other, an attempt is made, on finding a potential agent or object for the new T-form at the top level of the action formula of the template under inspection, to identify it with the main agent or object formula of the template. If this can be done, the agent or object formula of the orginal template is used, as being more specific. For example, in extracting from the action formula in the template for "John fired+at the deer", we find the goal case in [fired+at], with dependent STRIK, which is the attempted action. The object of that action, found to its left, is *ANI which can be instantiated by the formula for "deer" in the main template., namely(THIS BEAST). So the latter is used as the object of the new, extracted, T form [John strikes deer], , since "deer" is more specific than "animate being".

For most other cases (recipient, direction, location, containment and source): the case element provides the new pseudo- action, and the new object itself is specified as the dependent of the case element. The new agent is found as follows: it is the highest level object actually available of the action that dominates the case (to its immediate right in the formula).

So in the formula for "pour" in "I pour the wine"

((*HUM SUBJ)(((FLOW STUFF)OBJE)((((WRAP THING)TO)MOVE)CAUSE)))

we encounter (moving leftwards through the the formula)the direction case in thesubformula ((WRAP THING)TO), implying that the (FLOW STUFF) , liquid, which is the highest level object in the formula, is moved in the direction of a container, or (WRAP THING). The case element TO is dominated by MOVE, whose highest level object is (FLOW STUFF), which would become the pseudo agent of the new extracted T-form, but since it can be identified with the object of the old template, namely "wine", it is, and that becomes the pseudo-agent of the new T-form, since it is more specific that "liquid". it is. (WRAP THING), the container, becomes the object of the new T-form and the direction case element becomes the action so we get an extractted form :

((((MAN OBJE)PLEASE)(FLOW STUFF)) (TO PDO) (WRAP THING))

which is [wine (DIRE to) some+container]

The INFER routine

This routine has access to the representation IREPE produced by EXTRACT, ANS and CSIR, the inventory of common sense inference rules. Its form at the top level is

INFER(IREPE ANS CSIR) → ANS'

where ANS' is either ANS or some sublist of it, preferably containing only a single item, the solution.

INFER first tries the zero-point strategy: trying to match some answer T-form with some variable T-form directly, with no use of CSIR rules. Matching here means that the two T-forms as arguments of a function MATCHT produce a non-nil result, which will be a list of the corresponding, but non-identical, nodes in the two matched T-forms. Thus the solution of the example "John drank the whisky. . . . . " is obtained by the zero-point strategy, and rests on the matching of the two T-forms:

[whisky (IN in) John+part]

[?it (IN in) stomach]

where MATCHT returns the list ((?it whisky)(John+part stomach)) containing the answer. No such match can be made for the alternative solution "glass".

The overall principle of inference at work is to select the shortest possible chain of inferences, on the assumption that an ambiguity of understanding of this sort should be solved in the most shallow way possible, lest the situation becomes intolerable for the understander. Thus a zero-point solution, if available, will always be the shortest possible chain of inference.

If the zero-point strategy fails, the CSIRs are called , stored as a list accessible by their action subformulas, and, moreover accessible from both "antecedent" and the "consequent" action subformula. At present we cope only with inferences of length one : those which require only a single CSIR for their solution. However, it should be possible to extend the present strategy to at least length two; and hopefully they will almost never be any longer.

Let us look once more at the example "The soldiers fired+at the women /and I saw /several fall". We have to resolve "several", which cannot be done by the basic mode since both soldiers and women can equally well fail. Let us set out the fragment representations and the extractions obtained as follows:

1 [soldiers fire+at women]

11 [soldiers fire THING]

12 [soldiers strike women]

2 [I saw DTHIS]

3 [?several fall DTHIS]

31 [?several (NOTUP BE) DTHIS]

The inventory of rules is searched for those containing any action subformula occuring in a T-form in the pool that also contains either an "answer" or a "problem variable". In this case we pull in a rule informally expressed:

[1 strike (*ANI 2)] ↔ [(*ANI 2) fall DTHIS]

Here variables are indicated by numbers ; *ANI expresses a restriction on the variable that any value of it must be animate, and the double ↔ indicates that this rule can be considered as running in either direction.

This CSIR form is of course a more perspicuous form of:

(1 (THIS STRIK) (*ANI 2)) ↔ ((*ANI 2) (NOTUP BE) DTHIS)

which would of course cover a wider class of activities than simply the English verb "strike". It would cover at least "hit" "batter" etc. as well.

Thus a chain of length one is established by the rule from T-forms 12 to 31, since the "animate condition " is satisfied and the variable ?several is identified by the rule with the formula for "women". It should be noted here that the inference rules are very weak in that the application of a rule like the present one is perfectly consistent with the description of a situation where an animate being is struck in some way but does not fall. And this weakness is wholly intentional.

One important inflection in CSIRs is whether or not negation is significant in them. The negation of the action in a T-form is normally significant. Consider "John drank no gin / in his martini / but it felt warm / in his stomach nonetheless". In the template for the first fragment, shifting the negation to the action, and extracting for the containment case from the formula for "drink", we shall obtain a T-form

[gin (NOTIN BE) container]

and another :

[gin (NOTIN BE) John]

Conversely we shall obtain , by the same method, from the second template [martini (IN BE) container] and [martini (IN BE) John]. In trying to tie only one of these drinks by matching to the extraction [?it (IN in) John], we shall, without the use of CSIRs be able to cut the ANS list down to a single member, namely [martini], since MATCHT will show us that [gin] cannot stay on ANS.

However, if we apply the same analyses to a sentence like "John wanted the car / in the window / and he knew /that he would get it", we know the correct referent of "it" is "car" and not "window" and we shall find ourselves linking the first and fourth T-forms with some rule such as [(*ANI 1) want 2] → [(*ANI 1) have 2]

But, and this is the point, if the same sentence had concluded ". . . but he knew he would not get it" we should have required the same rule and the same answer, this rule has its "consequent" action marked to show that negation of it is irrelevant to its application.

This inferential method can be seen to be non-deductive very clearly at this point since it could be said to be of the form $A \rightarrow [B \vee \neg B]$, which is not a rule with any content whatever in a deductive system, yet its role here is clear and necessary.

GENERAL DISCUSSION AND SOME PARTICULAR COMPARISONS

The system described cannot be considered in any way adequately tested , partly because no one has any very clear idea of what constitutes a test in this area. But even to qualify, the basic mode must be shown to be stable under a considerable vocabulary and range of senses for words, and the extended mode must be shown to be determinate with a decent sized inventory of CSIRs.

The strategies now employe in this extended mode will also be those employed in a general discourse ambiguity procedure to back up the basic mode's present ability to resolve ambiguity within a small context of a few fragments. Ambiguity over a context larger than that is rare in discourse, just as is ambiguity of the sort discussed in this paper, but we should be prepared for it in an adequate understander.

The work described has a strong , and possibly naive, psychological assumption, namely that chain length is a reasonable metric to establish preferred inferential interpretations. I think it is reasonable, and that the tension introduced into understanding by prolonged unresolved ambiguity has been overlooked.Notice here that chain-length means number of CSIRs employed, not counting extractions.Hence , different ways of writing down formulas will not affect chain length.

Moreover, I would justify the principle as being essentially an extension of what I called semantic preference (Wilks 2) used in setting up the basic representation. That preference was justified as an opting for the "semantically densest" interpretation which was, I claimed, the one "with the least meaning"(in the sense in which a string of random words carries the maximum possible information). Similarly, the shortest chain of inferences also minimises the information in play, and introduces the least extraneous inductive information into the system. It is clear that such a notion of information based choice is ultimately inadequate. We only have to consider a sentence like "I was named after my father" where it seems clear that we exclude one interpretation simply because it contains vitually no information. This alone shows there must be some qualification to a "minimising information" theory. However, the fact that all available theories are wrong, by no means puts them all in the same position. I think such hypotheses about the overall manner in which an understanding system endeavors to maintain its coherence are well worth making and testing, and that they represent an aspect of human language "competence" almost wholly ignored by current linguistics and artificial intelligence. One could make the point more precise as follows: virtually all the systems in those areas define "success", that is to say the success of a particular parsed representation with respect to a text. What they do not tell us is what to do when a number of success are registered, as is almost always the case in realistic practice. But human understanders do not just accept the many, or opt for the first they find, or pick one at random: they prefer one in particular on some principled basis.

It is for this reason that the subject investigated in this paper cannot be treated in isolation from an adequate linguistic base system, as some seem to think. The inferring of a correct interpretation is intimately related to the systematic exclusion of competing interpretations, and any system that does not allow realistic ambiguity of sense and structure in at the start can hardly appreciate this point because thed difficulty never arises there, but then neither does one essential aspect of natural language either. I have developed elsewhere (6) an abstract view of meaning along these lines: that to have meaning is essentially to have one meaning RATHER THAN ANOTHER. Or, put another way, having meaning essentially involves procedures for the exclusion of alternative interpretations. This, I believe, is the residual truth lurking beneath the "procedural view of meaning", a thesis which when taken a face value is patently false.

Let me mention a closely related shortcoming of the micro-world approach to natural language analysis: it concerns what I believe to be an endemic muddle in AI about the notion of "inference". Let me start by restating the obvious, not from dry motives of clarity, but because I believe the muddle has important practical consequences in the area of natural language understanding.

Here are some inferences, in the bare sense of that word, of transitions that people might make from one assertion to another.

(i)All Englishmen are untrustworthy and Cecil is an Englishman, SO he is untrustworthy.

(ii)Cecil is an Englishman, SO he is untrustworthy.

(iii)This is triangular, SO it is three-sided.

I take it that (i) is a deduction, true in all possible worlds, and quite independent of the meanings of the words "Englishman", "Cecil" and "untrustworthy".

(ii) is an inference , simply and solely, and certainly not valid, whether or not it happens to be true for some English Cecil.

(iii) is a valid inference, true in all possible worlds, as they say, because of the central meanings of "triangular" and "three sided", a fact that is sometimes expressed by saying that the premise missing, for this to be a deduction, namely "all triangular things are three sided" is analytically true.

What is the point for our purposes of all this dogmatic and semi-traditional classification? Simply this: the extra-conceptual (outside the semantic dictionary, that is) CSIR inferences of the sort we have discussed in this paper problems in text, are of type (ii). These inferences could function as part of a deductive system by the addition of sufficient inductively unreliable premisses to convert them to form (i). They could then function within established deductive machinery, such as first order PC, PLANNER in one of its modes of operation, etc.

But there may be no reward for doing that, at least in the case of natural language analysis, because the conclusions reached can be no more reliable than the dubious generalisations functioning as premisses, whatever the power of the deductive machinery intervening.

In this paper I have described how such weak information can fulfil a problem solving role in natural language analysis, in terms of a notion of "adequately confirmed" inference in context. But that does not require the deductive machinery at all.

My point will be clarified here by noting two research situations where, by contrast, the deductive machinery may pay its way: (1) in robots and (2)in simulated micro worlds.

In the case of a robot, really moving about in the world with deductively manipulated information and plans, the world itself can provide a clear sense of contradiction. If the robot's deductions tell it the door is open, but it bangs into the firmly closed door in fact, then the conclusion is contradicted and the preceding premisses can be reexamined, as would be the case with a scientific theory refuted by unsuccessful experiment. That is to say, the premisses may be unreliable, but because there can be contradiction of conclusions the deductive machinery can transfer the "not" back to some premise, as if refuting a scientific theory. (though the question of which premise it should be transferred back to is very difficult of course).

This situation I maintain is quite different from the analysis of continuous natural language where there is little or no expectation of contradiction: if, in understanding the text, the understander erroneously infers A, there is little or no chance of encountering the assertion −A in the text in the near future. A robot could in principle conduct Millian experiments, and in the case of a dialog in natural language, one can always step back and ask questions of the partner, but one understands texts without experimenting on them. The present system makes no claims at all to discover such contradiction in a general way and to backtrack having done so, to the next least preferred interpretation. And at the moment no system is in striking distance of such an ability.

The case of simulated micro-worlds is different. Here there is no contradiction at all, but there is no need for it since all premisses are, in effect, analytic, and no real information can ever enter the system. For example, after executing the command "Clear off the top of the red block" , it is clear by definition ,apart from the possibility of definitional contradiction. No lingering and sticky cigarette end can remain to imperil the stability of the house of bricks about to be built. It will be clear that such situations have little to do with the unreliable inductive information required for the analysis of natural language.

My main point here has been that if there is no payoff to be obtained from a strong deductive approach to natural language understanding, there is no strong case for pursuing it. This position is different from. though quite consistent with, the position that distinguishes between proofs and deductions in formal fields, and urges the pursuit of the investigation of the former: that is to say, the search for the actual principles at work in a field rather than a search for an axiomatisation of principles about whose content we are ignorant. An assumption of the deductive position is that their method does also provide principles of content, or human competence in this area, and not merely the formalisation of principles that could be expressed in some other way.

After those clarifications, some very brief comparisons follow between the work described here, and three other AI approaches to language understanding : those of Charniak(7), Schank(8) and Winograd(9). Detailed comparison and criticism of systems is not appropriate here, and I give only brief general remarks, in orderto contrast different systems along a number of dimensions; which are (a)the adequacy of the linguistic base constructed or proposed, in terms of application to everyday texts in English. (b)the degree of implementation and the definiteness of the task proposed as an explication of the elusive notion of "understanding". (c)the placing of the system within the inference-deduction opposition. and (d)the implementation of a preference system that both prefers certain interpretations to others on a reasoned principle.

(a)Charniak does not consider the linguistic base essential , and is not particularly interested in the ambiguities of sense of words. Schank's is the most adequate linguistic base of the three, and the one closest to the present system in general presuppositions. Winograd's system is restricted to unambiguous simple words, and a feature-type analysis of their meaning. Even if the words had only single senses, it is doubtful if the meaning of complex concepts , especially actions, could be expressed in that way.

(b)Charniak's system is not intensively implemented but has a very definite task defined for it, the resolution of the sorts of anaphora problem discussed in this paper. I think the strategies we advocate for the problem differ chiefly over whether or not the rules form an unmanageable totality if expressed at the word level, as he does, and as to whether one needs the facility to chain CSIRs, or what he calls "demons". Schank's system is on the verge of implementation through the conjunction of a number of large programs. However, at present it does not seem directed to a specific task in the inference field, since the drawing of inferences per se (as distinct from the drawing of them for the solution of some problem or performance of some task) is a no easy matter to assess. The "inferences to be drawn from x" is not a well defined notion, outside the pages of detective fiction. Winograd's system is completedly implemented within its original design and directed to a precise and assessable task. Its merit, as the implementation of an existing theory of grammar, is often said to be its interconnecting the syntax and the semantics (Minsky and Papert, 19). This is an odd remark, in an AI context, in that it presupposes a conventional and traditional distinction (semantics and syntax, that is) that much work in semantic analysis by computer (Schank and the present work for example) has found unnecessary in practice. The remark is interesting in view of the same authors' criticism of the perception-cognition distinction as over-commonsensical and dispensible.

(c)With inference and deduction I do not feel on sure ground because it is hard to assess the authors' work, since, naturally their expositions were not designed to answer this question of mine. Moreover, the situation is complicated by the fact that some of the authors use "deduction" to cover processes that are almost certainly not deductive at all. . My feeling is that Winograd's system is deductive and that Charniak's and Schank's, like the present one, are not, that is to say, the latter call in inference rules whose application is determined only by by the possibility of fitting them linguistically to the language context in hand. Any clarification from the authors on the relation of their work to this distinction will be gratefully received.

On point (d) my central one of preference and choice between interpretations I think nothing has been done by the authors , and perhaps they believe that there will always be one and only one interpretation successful in terms of their rules, or that the first found will do. Quillian, is, I think, the only worker in the field who has given any attention to this question.

REFERENCES

(1)Wilks, Y. An Artificial Intelligence approach to machine translation. Stanford AI Laboratory Memo #161, and in Schank, R., and Colby, K., (Eds.) Computer Models of Thought and Language, San Francisco, 1973 (in press).

(2)Wilks, Y. Preference semantics. Stanford AI Laboratory Memo #206, and in Keenan, E. (Ed.) Formal Semantics of Natural Language, Cambridge, 1974 (in press).

(3)Wilks, Y. Understanding without proofs, in Proc. Third. Internat. Joint Conf. on AI, 1973.

(4)Wilks, Y. and Herskovits, A. An intelligent analyser and generator for natural language. in Proc. Internat. Conf. on Computational Linguistics, Pisa, 1973.

(5)Herskovits, A. The generation of French from a semantic representation. Stanford AI Laboratory Memo #212.

(6)Wilks, Y. Decidability and natural language. Mind. 1971.

(7)Charniak, E. Jack and Janet in search of a theory of knowledge. (as 3)

(8)Schank, R., and Rieger, C. Inference and the computer understanding of natural language. Stanford AI Laboratory Memo #197.

(9)Winograd, T. Understanding Natural Language. Edinburgh. 1972.

(10)Minsky, M., and Papert, S. Artificial Intelligence Progress Report. A. I. Memo #252. MIT AI Laboratory.

(11)Quillian, R. Semantic memory. in Minsky, M. (Ed.) Semantic Information Processing. MIT. 1968.