American Journal of Computational Linguistics *Microfiche 40*

NATURAL LANGUAGE UNDERSTANDING SYSTEMS WITHIN THE A. I. PARADIGM: A SURVEY AND SOME COMPARISONS

YORICK WILKS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE
UNIVERSITY OF EDINBURGH
EDINBURGH EH8 9NW

Revised version of Stanford Artificial Intelligence Laboratory Memorandum 237, supported by contract number NIE-P-75-0026 with the National Institute of Education.

ABSTRACT: The paper surveys the major projects on the understanding of natural language that fall within what may now be called the artificial intelligence paradigm for natural language systems. Some space is devoted to arguing that the paradigm is now a reality and different in significant respects from the generative paradigm of present day linguistics. The comparisons between systems centre round questions about the level, centrality and "phenomenological plausibility" of the knowledge and inferences that must be available to a system that is to understand everyday language.

Contents

1. Introduction

In his report to the Science Research Council on the state of Artificial Intelligence, Sir James Lighthill (1973) gave most of the field a rather bad prognosis. One of the few hopeful signs he saw was Winograd's (1972) natural language understanding system. Yet, only a year later, Winograd had stopped work on the system he constructed, and had begun a new one on entirely different principles.** He went so far, in a survey lecture (Winograd '73) of extraordinary modesty in a field not known for its small claims, to place his celebrated early work in only the 'first generation' of computer systems designed to understand natural language, and went on to describe others' 'second generation' systems.

I shall return later to this metaphor of generations, but what is one to say in general terms of a field where yesterday's brightest spots are today's first generation systems, even though they have not been criticised in print, nor shown in any generally acceptable way to be fundamentally wrong? Part of the answer lies in the profound role of fashion in Artificial Intelligence in its present pre-scientific phase. A cynical American professor remarked recently that Artificial Intelligence (AI) had an affair with someone's work every year or two, and that, just as there were no reasons for falling in love, so, later, there were no reasons for falling out again. In the case of Winograd's work it is important now to resist this fashion, and re-emphasize what a good piece of research it was, as I shall in a moment.

Another part of the answer lies in the still fundamental role of metaphysical criticism in AI. In the field of computer vision things are bad enough, in that anybody who can see feels entitled to criticise a system, on the ground that he is sure he does not see using such and such principles. In the field of natural language understanding things are worse: not only does anyone who can speak and write feel free to criticise on the corresponding grounds, but in addition there are those trained in disciplines parasitic upon natural language, linguists and logicians, who often know in addition how things MUST BE DONE on a priori grounds. It is this metaphysical aspect of the subject that gives its disputes their characteristically acrimonious

---

** see (Winograd '74b)

flavour.

In this paper I want to sort out a little what is agreed and what is not; what are some of the outstanding disputes and how testable are the claims being made. If what follows seems unduly philosophical, it should be remembered that little is agreed, and almost no achievements are beyond question. To pretend otherwise, by concentrating only on the details of established programs, would be meretricious and misleading.

To survey an energetic field like this one is inevitably to leave a great deal of excellent work unexamined, at least if one is going to do more than give a paragraph to each research project. I have left out of consideration at least seven groups of projects:

(1) Early work in Artificial Intelligence and Natural Language that has been surveyed by Winograd (1973) and Simmons (1970a) among others.

(2) Work by graduate students of, or intellectually dependent upon that of, people discussed in some detail here.

(3) Work that derives essentially from projects described in detail here. This embraces several groups interested in testing psychological hypotheses, as well as others constructing large-scale systems for speech recognition. I have devoted no space to speech recognition as such here, for it seems to me to depend upon the quality of semantic and inferential understanding as much as anything, and so I have concentrated upon this more fundamental task.

(4) Work on language generators, as opposed to analysers and understanders. They are essential for obtaining any testable output, but are theoretically secondary.

(5) All the many and varied reasoning schemes now available in AI, including PLANNER (Hewitt 1969), QA4 (Rulifson et al 1972), MERLIN (Moore and Newell 1973) as well as automatic programming (Balzer et al 1974) (Heidorn 1974) and debugging (Sussman 1974) projects, many of which are producing formalisms that appear increasingly like natural language.

(6) Conservative reasoning schemes, such as first order predicate calculus, that have been applied to, or advocated for, the analysis of natural language: (McCarthy and Hayes 1969) (Coles 1972) (Sandewall 1972).

(7) I have also ignored, as one must in order to write at all in a rapidly changing field, the interpretation given to Minsky's (1975) notion of "frame" during 1975 by Charniak and Schank. During this year both have produced sketches for a representation of knowledge on a larger scale than any discussed in this survey: roughly speaking, they have each produced a scheme for a whole story. The value and function of such a larger-scale representation is much under discussion at the moment, but none of it invalidates what is contained here, except for some qualification to the position of Schank which is noted under the heading "Centrality" in section 6.

The exclusions under (2) above are particularly unfair to the work in the unpublished theses of Rieger (1974) and McDermott (1974), and can only be justified, like those of sections (5) and (6) above, by space, bias and the considerations advanced in a later section of this survey concerning what it is for AI research to be about natural language, rather than about something else, like human psychology, or logic, or the bargaining behaviour of children under stress.

One final remark about the style of this survey. I have not indulged the normal critical luxury of careful and extensive quotation from authors followed by detailed and meticulous criticism. That would be wholly inappropriate in the present state of things. A great deal of work is being done at the moment, and many of the principal researchers change their views on very fundamental questions between one paper and the next without drawing any attention to the fact. Cheap self-contradictions and changes of mind are all too easy to find, so criticism and comparisons are best drawn with a very broad brush and a light stroke.

2. Winograd's Understanding System

Winograd's system runs as a dialogue, in real time, between a human operator who types in messages, and the system proper, called SHRDLU, which displays on a screen pictures of a closed world of coloured blocks and pyramids, a box into which they can be put, and an arm or crane that can move the objects about. SHRDLU displays its replies to the operator by writing on the same screen. The following is a discontinuous sample of the dialogue Winograd gives in his book. SHRDLU's replies are in the upper case:

pick up the big red block.

OK

find a block which is taller than the one you are holding and put it into the box.

BY 'IT', I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING

OK

what does the box contain

THE BLUE PYRAMID AND THE BLUE BLOCK

had you touched any pyramid before you put the green one on the little cube

YES, THE GREEN ONE

when did you pick it up

WHILE I WAS STACKING UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

why

TO GET RID OF IT.

This ingenious program was also able to accept new definitions of words, such as steeple'; to make use of what it knew about the properties of the objects described, such as that a block could have its top cleared off, but a pyramid could not; and also to remember what it had done before, as in the sample above.

The syntactic analysis program was written in PROGRAMMAR, a procedural language related to PLANNER (see Hewitt '69). This means that a familia phrase structure rule such as

$$S \rightarrow NP + VP$$

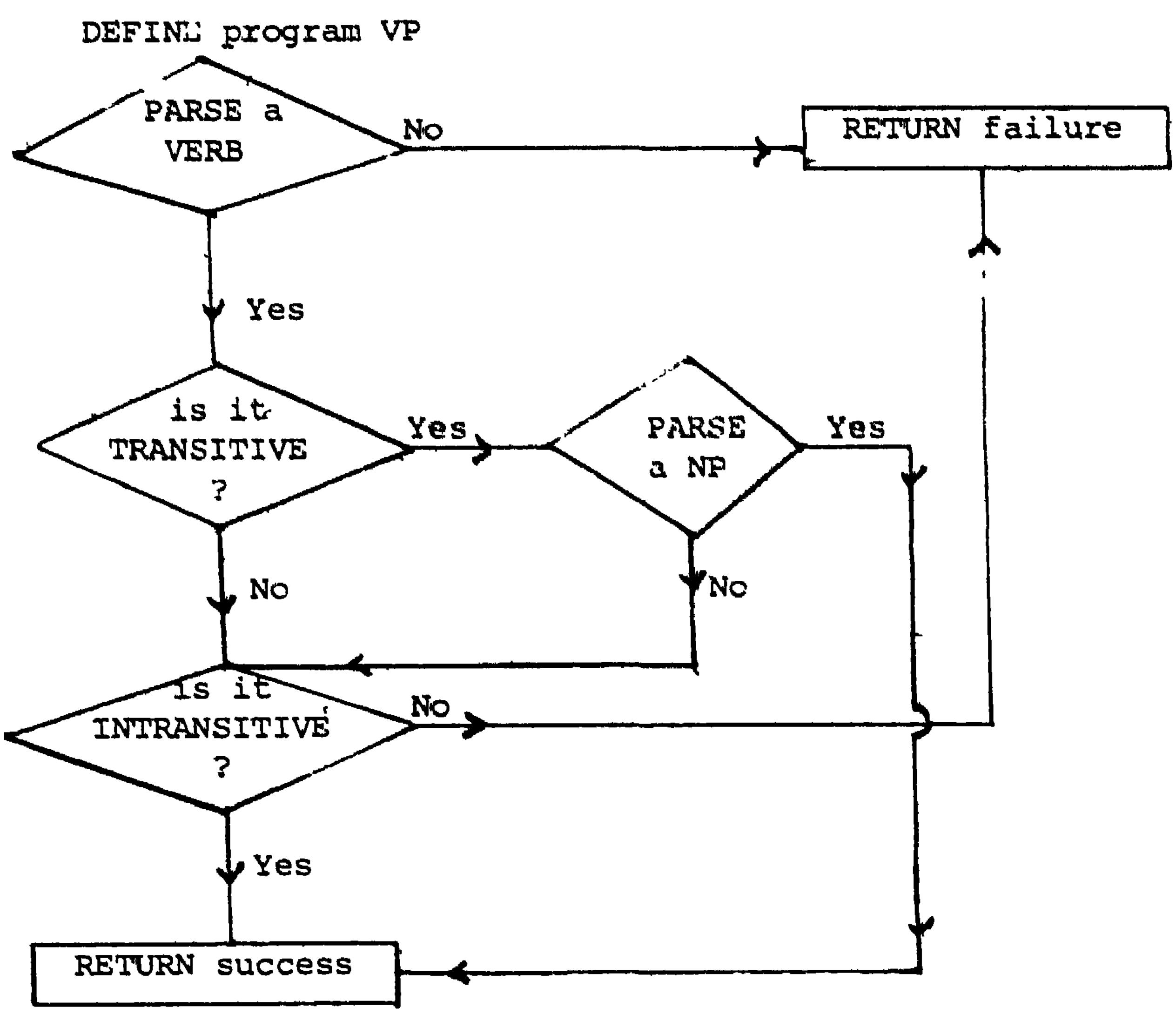
(to be interpreted: a sentence consists of a noun phrase followed by a verb phrase) would be expressed as:

```
(PDEFINE SENTENCE
  ((( PARSE NP) NIL FAIL)
   (( PARSE VP) FAIL FAIL RETURN)))
```

The details of the notation need not detain us at this point; what is important is that Winograd's grammar is not the conventional list of rules, but small sub-programs like the lines above, that actually represent procedures for imposing the desired grammatical structure.

The first level of linguistic procedures in the system applies a systemic grammar', dur to M.A.K. Halliday (1970), which imposes a hierarchical structure of clauses on the input sentences, which seem to be drawn from a vocabulary of about 175 words.

Winograd's parsing is top down, and depth first, with no automatic back up. The parsing program for each grammatical category is a functional definition in PROGRAMMAR, which can be stated either as above for SENTENCE, or as a flow-chart as below for VP:



Here is Winograd's own account of the start of this top-down parsing procedure for the sentence "Pick up a red block" (where the material in [ ] is added explanation and not Winograd's own):

"The CLAUSE program looks at the first word, to decide what unit the CLAUSE begins with. If it sees an adverb, it assumes the sentence begins with a single-word modifier [Slowly, Jack lifted the book]; if it sees a preposition, it looks for an initial PREPG [On top of the hill stood a tree] If it sees a BINDER, it calls the CLAUSE program to look for a BOUND CLAUSE [Before you get there, we left]. In English (and possibly all languages) the first word of a construction often gives a very good clue as to what that construction will be. In this case, "pick" is a verb, and indicates that we may have an IMPERATIVE CLAUSE. The program starts the VG program with the initial VG feature list (VG IMPER), looking for a VG of this type. This must either begin with some form of the verb "do" [Do not call me!] or with the main verb itself [Call me!]. Since the next word is not "do" it checks the next word in the input (in this case still the first word) to see whether it is the infinitive form of a verb. If so, it is to be attached to the parsing tree, and given the additional feature MVB (main verb). The current structure can be diagrammed as:

```
(CLAUSE MAJOR)
   (VG IMPER)
      (VB MVB INF TRANS VPRT --------------pick
```

TRANS AND VPRT came from the definition of the word "pick" when we called the function PARSE for a word."

After this syntactic parsing, a number of "semantic specialists" attach semantic structures to specific syntactic ones. A semantic definition of an

In the case of "a red cube", the following structure is built up by an NP "semantic specialist"

```
(GOAL (IS ?X BLOCK))
(GOAL (COLOR ?X RED))
(EQDIM ?X)-----------------------PLANNER description
(BLOCK MANIP PHYSOB THING)----------markers
```

The first three lines will eventually form the bulk of a Micro-Planner program which, when evaluated will seek an object X that is a block, is equidimensional (EQDIM) and is red (where "red" itself has a definition, the system that restricts its application to objects with the feature PHYSOB) The last line of the figure is a set of "semantic features" read off right to left from the following feature tree"

```
                   |NAME
                   |PLACE            | SHAPE
                   |PROPERTY ---     | SIZE
                   |                 | LOCATION
                   |                 | COLOR            |ROBOT
                   |ANIMATE ----------------------      |
                   |                                    |HUMAN
                   |                    |BLUE
                   |                    |RED
                   |         (--------- |BLACK
                   |         (          |WHITE
                   |         (          |GREEN                         | STACK
THING ---          |PHYSOB--(  |CONSTRUCT----------------------------  | PILE
                   |         (  |HAND                                  | ROW
                   |         (--|TABLE              | PYRAMID
                   |            |MANIP -----------  | BLOCK
                   |                                | BALL
                   |
                   |RELATION-------------  | EVENT
                   |                       |
                                           | TIMELESS
```

The semantic structure of "the red cube" can be used by the deductive component of the system, before evaluation resulting in the actual picking up, to see if such an object is possible. If it were not, (an "equidimensional pyramid" would not be) the system could go back and try to re-parse the sentence.

The meaning of verbs in SHRDLU is more complex. The semantic component has access to a definition for "pick-up" just as it does for "red" and "block" and this definition will enable SHRDLU to translate "pick-up tatements" into Micro-planner in a manner analogous to that for noun phrases.

There are two complications here. Firstly "pick-up", unlike "red", is defined in terms of other concepts in the system: in particular, in terms of GRASP and RAISEHAND, which are two of the three basic actions in the system. Secondly, there are two <u>types</u> of verb definition, semantic and inferential Winograd does not give the semantic definition for "pick-up", but here is the one for "grasp" which is a closely related verb.

```
(CMEANS (((#ANIMATE)) (#MANIP)))
    (#EVAL(COND((PROGRESSIVE)
        (QUOTE(#GRASPING #2 *TIME)))
```

```
(T(QUOTE(#GRASP #2 *TIME
      )))))NIL))
```

which says essentially that grasping is something done by an animate entity to a manipulable one (first line). More of the real content of such actions is found in their inferential definition. Here is the one for "pick-up":

```
(CONSE TC - PICKUP
     (X)
     (PICKUP X)
     (GOAL(GRASP ?X) THEOREMS)
     (GOAL(RAISEHAND THEOREMS)
```

This definition allows the program to actually carry out the "pick-up" command if it is possible to do so in the simulated world, as it would not be, for example, if there were already a block on top of the red one. PICKUP is being defined in terms of a number of more primitive sub-actions, such as GRASP and RAISEHAND each of which must be carried out in order that something may indeed be picked up. There sub-actions themselves have inferential definitions: the one given for GRASP, for example, is somewhat different from its "CMEANS" definition given above, although the inferential definitions are also, in some sense, definitions of meaning as well as programs for actually carrying out the associated commands.

One reason for the enormous impact of this work was that, prior to its appearance, AI work was not very linguistically interesting, while the systems of the linguists had no place for the use of inference and real world knowledge. Thus a very limited union between the two techniques was able to breed considerable results. Before Winograd there were few programs in AI that could take a reasonable complex English sentence and ascribe any structure whatever to it. In early classics of 'natural language understanding' in AI, such as Bobrow's STUDENT (1968) problem solver for simple algebra, input sentences had to be short and of stereotyped form, such as "what is the sum of .... ?"

Conversely, in linguistics, there was, until very recently, little speculation on how we understand the reference of pronouns in such elementary sentences as "The soldiers fired at the women and I saw several fall", where it is clear that the answer is both definite, and that finding it requires some inferential manipulation of generalizations about the world. The reader should ask himself at this point how he knows the correct referent of the pronoun in that sentence.

3. Some Discussion of SHRDLU

So far, the reaction to Winograd's work has been wholly uncritical. What would critics find to attack if they were so minded? Firstly, that Winograd's linguistic system is highly conservative, and that the distinction between 'syntax' and 'semantics' may not be necessary at all. Secondly that his semantics is tied to the simple referential world of the blocks in a way that would make it inextensible, to any general, real world, situation. Suppose 'block' were allowed to mean 'an obstruction' and 'a mental inhibition', as well as 'a cubic object'. It is doubtful whether Winograd's features and rules could express the ambiguity, and, more importantly, whether the simple structures he manipulated could decide correctly between the alternative meanings in any given context of use. Again, far more sophisticated and systematic case structures than those he used might be needed to resolve the ambiguity of 'in' in 'He ran the mile in five minutes and He ran the mile in a paper bag , as well as the combination of case with word sense ambiguity in 'He put the key in the lock' (door lock) and 'He threw the key in the lock' (river lock).

The blocks world is also strongly deductive and logically closed. If gravity were introduced into it, then anything supported that was pushed in a certain way would have, logically have, to fall. But the common sense world, of ordinary language, is not like that: in the 'women and soldiers' example given earlier, the pronoun 'several' can be said to be resolved using some generalisation such as 'things shot at and hurt tend to fall' There are no logical 'have to's' there, even though the meaning of the pronoun is perfectly definite.

Indeed, it might be argued that, in a sense, and as regards its semantics, Winograd's system is not about natural language at all, but about the technical question of how goals and sub-goals are to be organised in a problem-solving system capable of manipulating simple physical objects.

If we remember, for example, that the key problem that brought down the enormous work on machine translation in the Fifties and Sixties, was that of the sense ambiguity of natural language words, then we will look in vain to SHRDLU for any help with that problem. There seems to be only one clear example of an ambiguous word in the whole system, namely that of 'contain' as it appears in 'The box contains a red block' and The stack contains a red block!

Again, if one glances back at the definition of 'pick-up' quoted above, one can see that it is in fact an expression of a procedure for picking up an object in the SHRDLU system. Nothing about it, for example, would help one understand theperfectly ordinary sentence 'I picked up my bags from the platform and ran for the train', let alone any sentence not about a physical action performable by the hearer. One could put the point so: what we are given in the PLANNER code is not a sense of 'pick up' but an example of its use, just as 'John picked up the volunteer from the audience by leaning over the edge of the stage and drawing her up by means of a rope clenched in his teeth' is not so much a sense of the verb as a use of it.

Those who like very general analogies may have noticed that Wittgenstein (1953 para. 2ff) devoted considerable space to the construction of an elementary language of blocks, beams and slabs; one postulated on the assumption that the words of language were basically, as is supposed in model theory, the names of items. But, as he showed of the enterprise, and to the satisfaction of many readers, "That philosophical concept of meaning (i.e. of words as the unambiguous names of physical objects---YW) has its place in a primitive idea of the way language functions. But one can also say that it is the idea of a language more primitive than ours". (my italics).

To all this, it might be countered that it has not been shown that the language facilities I have described cannot be incorporated in the structures that SHRDLU manipulates, and that, even if they could not, the work would still be significant in virtue of its original control structure and its demonstration that real world knowledge can be merged with linguistic knowledge in a working whole. Indeed, although Winograd has not tried, in any straightforward sense, to extend the SHRDLU system one could say that an extension of this sort is being attempted by Brown (1974) with his 'Believer System' which is a hybrid system combining a component about beliefs that is, in the sense of section 4 below 'second generation', with a base analyser from Bruce's CHRONOS system (1972) which is a micro world ---late first generation---system in the same sense as Winograd's. Others in the last category that should be mentioned are Davies and Isard's (1972) exploration of the concepts of 'must' and 'could' in a micro-world of tic-tac-toe, and Joshi's extension of it (1973), but above all the important and influential work of Woods (1972).

This work, most recently applied to a micro-world of lunar rock samples, is not discussed in the detail it deserves in this paper. The system, based on an augmented state transition network grammar, is undoubtedly one of the most robust in actual use, in that it is less sensitive to the PARTICULAR input questions it encounters than its rivals. The reason for not treating it in depth is that both Woods and Winograd have argued in print that their two systems are essentially equivalent (Winograd 1971) (Woods 1973), and so, if they are right, there is no need to discuss both, and Winograd's is, within the AI community at least, the better known of the two.

Their equivalence arguments are probably correct: both are grammar-based deductive systems, operating within a question-answering environment in a highly limited domain of discourse. Winograd's system of hints on how to proceed, within his PROGRAMMAR grammar, is, as he himself points out, formally equivalent to an augmented state transition network, and in particular to the ordering of choices at nodes in Woods' system.

There is a significant difference in their metaphysical approaches, or presuppositions about meaning which, however, has no influence on the actual operation of their respective systems. This difference is disguised by the allegiance both give to a 'procedural view of meaning' The difference is that Woods takes a much more logico-semantic interpretation of that slogan than does Winograd. In particular, for Woods the meaning of an input utterance to his system is the procedures within the system that manipulate the truth conditions of the utterance and establish its truth value.

To put the matter crudely, for Woods an assertion has no meaning if his system cannot establish its truth or falsity. Winograd has certainly not committed himself to any such extreme position.

It is interesting to notice that Woods' is, in virtue of his strong position on truth conditions, probably the only piece of work in the field of AI and natural language to satisfy Hayes' (1974) recent demand* that to be 'intellectually respectable' a knowledge system must have natural model theoretic semantics, in Tarski's sense. Since no-one has ever given precise truth conditions for any interesting piece of discourse, such as, say, Woods' own papers, one might claim that his theoretical presuppositions necessarily limit his work to the analysis of micro-worlds (as distinct from everyday language). However, if Woods' 'internal' interpretation of the 'meanings are procedures' slogan has certain drawbacks, so too does Winograd's, or what one might call the 'external' interpretation. By that I mean Winograd's concentration on actions, like picking up, that are in fact real world procedures, and in a way that the meanings of 'concentrate', 'call', 'have', 'interpret', etc. are not self-evidently real world procedures that we could set out in PLANNER for a robot. Of course, Winograd is free to concentrate on any micro-world he wishes, and all I am drawing attention to here is the danger of assuming that natural language is normally about real world procedures and, worse still, the implicit making of the assumption that we cannot understand discourse about a procedure unless we can do it ourselves. I am not saying that Winograd is making this evidently false assumption, only that the rhetoric surrounding the application of the 'meanings are procedures' slogan to his system may cause the unwary to do so.

There is quite a different and low-level problem about the equivalence of Woods' and Winograd's systems, if we consider what we might call the received common-sense view of their work. Consider the following three assertions:

(1) Woods' system is an implementation of a transformational grammar

(2) Winograd's work has shown the irrelevance of transformational grammar for language analysis - a view widely held by reviewers of his work.

---

* a view modified in Hayes (1975) where it now seems that programs/procedures would serve as a 'semantics' instead (a quite different, and more reasonable, position, of course).

(3) Woods' and Winograd's systems are formally equivalent - a view held by both of them.

There is clearly something of an inconsistent triad amongst those three widely held beliefs. The trouble probably centres on the exact sense which Woods' work is formally equivalent to a transformational grammar - not a question that need detain us here, but one worth pointing out in passing

4. Some More General Background Issues

Winograd's work is a central example of the 'Artificial Intelligence paradigm of language', using 'paradigm' in Kuhn's (1970) sense of a large scale revision in systematic thinking, where the paradigm revised is the 'generative paradigm' of the Chomskyan linguists (Chomsky 1957). From the AI pointof view, the generative linguistic work of the last fifteen years has three principal defects. Firstly, the generation of sentences, with whatever attached structures, is not in any interesting sense a demonstration of human understanding, nor is the separation of the well-formed from the ill-formed, by such methods- for understanding requires, at the very least, both the generation of sentences as parts of coherent discourse and some attempt to interpret, rather than merely reject, what seem to be ill-formed utterances. Neither the transformational grammarians following Chomsky, nor their successors the generative semanticists (Lakoff 1971), have ever explicitly renounced the generative paradigm.

Secondly, Chomsky's distinction between performance and competence models, and his advocacy of the latter, have isolated modern generative linguistics from any effective test of the systems of rules it proposes. Whether or not the distinction was intended to have this effect, it has meant that any test situation necessarily involves performance, which is considered outside the province of serious linguistic study. And any embodiment of a system of rules in a computer, and assessment of its out put, would be performance. AI, too, is much concerned with the structure of linguistic processes, independent of any particular implementation,**

** Vide: "Artificial Intelligence is the study of intellectual mechanisms apart from applications and apart from how such mechanisms are realised in the human or in animals." (McCarthy 1974) but implementation is never excluded, as it is from competence models, but rather encouraged.

Thirdly, as I mentioned before, there was until recently no place in the generative paradigm for inferences from facts and inductive generalisations, even though very simple examples demonstrate the need for it.

This last point, about the shortcomings of conventional linguistics is not at all new, and in AI is at least as old as Minsky's (1968,p.22) observation that in 'He put the box on the table. Because it wasn't level, it slid off', the last 'it' can only be referred correctly to the box, rather than the table, on the basis of some knowledge quite other than that in a conventional, and implausible, linguistic solution such as the creation of a class of 'level nouns' so that a box would not be considered as being or not being level.

These points would be generally conceded by those who believe there is an AI paradigm of language understanding, but there would be far less agreement over the positive content of the paradigm. The trouble begins with the definition of 'understanding' as applied to a computer. At one extreme are those who say the word can only refer to the performance of a machine: to its ability to, say, sustain some form of dialogue long enough and sensibly enough for a human interrogator to be unsure whether what he is conversing with is a machine or not. On the other hand, there are many, almost certainly a majority, who argue that more is required, in that the methods and representations of knowledge by which the performance is achieved must be of the right formal sort, and that <u>mere</u> performance based on ad hoc methods does not demonstrate understanding.

This issue is closely related to that of the role of deduction in natural language understanding, simply because deduction is often the structure meant when 'right methods' are mentioned. The dispute between those who argue for, or, like Winograd, use deductive methods, and those who advocate other inferential systems closer to common sense reasoning, is in many ways a pseudo-issue because it is so difficult to define clearly what a non-deductive system is, (if by that is meant a system that cannot in principle be modelled by a deductive system) since almost <u>any</u> set of formal procedures, including 'invalid inferences', can be so displayed.

The heart of the matter concerns the most appropriate form on an inference system, rather than how those inferences may be axiomatised, and it may well turn out that the most appropriate form for plausible reasoning in order to understand is indeed non-deductive. This same insight has largely defused another heated issue: whether the appropriate representations should be procedures or declarations. Winograd's work was of the former type, as was shown by his definitions of words like 'pickup' as procedures for actually picking things up in the blocks world. However, simple procedural representations usually have the disadvantage that, if you are going to indicate, for every 'item' of knowledge, how it is to be used, then, if you may use it on a number of kinds of occasion, you will have to store it that number of times. So, if you want to change it later, you will also have to remember to change it in all the different places you have put it. There is the additional disadvantage of lack of perspicuity: anyone reading the procedural version of the Winograd grammar rule I gave earlier, will almost certainly find the conventional, declarative, version easier to understand.

So then, the fashion for all things procedural has to some extent abated (see Winograd 1974). There is general agreement that any system should show, as it were how it is actually to be applied to language, but that is not the same as demanding that it should be written in a procedural language,line PLANNER. I shall return to this last point later.

5. Second Generation Systems

To understand what was meant when Winograd contrasted his own with what he called second generation systems, we have to remember, as always in this subject, that the generations are of fashion, not chronology or inheritance of ideas. He described the work of Simmons, Schank and myself among others in his survey of new approaches, even though the foundations and terminology of those approaches were set out in print in 1966, 1968 and 1967 respectively. What those approaches, and others have in common is the belief that understanding systems must be able to manipulate very complex linguistic objects, or semantic structures, and that no simplistic approaches to understanding language with computers will work.

In a very influential recent paper, Minsky (1974) has drawn together strands in the work of Charniak (1972) and the authors above using a terminology of 'frames':

"A frame is a data-structure for representing a stereotype situation, like a certain kind of living room, or going to a children's birthday party. Attached to each frame are several kinds of information. Some of this is information about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.

We can think of a frame as a network of nodes and relations. The top levels of a frame are fixed and represent things that are always true about the supposed situation. The lower levels have many terminals --- 'slots' that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet .... Simple conditions are specified by markers that might require a terminal assignment to be a person, an object of sufficient value, etc...."

The key point about such structures is that they attempt to specify in advance what is going to be said, and how the world encountered is going to be structured. The structures, and the inference rules that apply to them, are also expressions of 'partial information' (in McCarthy's phrase) that are not present in first generation systems. As I showed earlier, with the 'women and soldiers' example, such loose inductive information, seeking confirmation from the surrounding context, is required for very simple sentences. In psychological and visual terms, frame approaches envisage an understander as at least as much a looker as a seer.

Thus, we might, very tentatively, begin by identifying what Winograd called 'second generation' approaches with those making use of very general notions akin to what Minsky called 'frames'. But this is no more than a temporary device, for convenient initial classification of the field, because later we shall have reason to question the first-second generation distinction, and, as noted earlier, Minsky's notion of 'frame' is itself a highly fluid one in the process of definition and refinement.

Let us now turn briefly to five approaches that might be called second generation.

Charniak

The new work which owes most to Minsky's advocacy is Charniak's. He studied what sorts of inferential information Charniak '72, '73, '74) would be needed to resolve pronoun ambiguities in children's stories, and in that sense to understand them. One of his example 'stories' is: 'Jane was invited to Jack's birthday party. She wondered if he would like a kite. A friend told Jane that Jack already had a kite, and that he would make her take it back

The problem concerns the penultimate word 'it', and deciding whether it refers to the first kite mentioned or the second. Charniak's analysis begins by pointing out that a great deal of what is required to understand that story is implicit: knowledge about the giving of presents, knowledge that if one possesses one of a certain sort of thing then one may well not want another, and so on.

Charniak's system does not actually run as a program, but is a theoretical structure of rules called 'demons' that correspond roughly to what Minsky later called frames. A demon for this example would be, 'If we see that a person might not like a present X, then look for X being returned to the store where it was bought. If we see that happening, or even being suggested, assert that the reason why is that P does not like X'.

The important words there are 'look for', which suggest that there may well be confirming hints to be found in the story and, if there are, then this tentative, partial, inference is correct, and we have a definite and correct answer. This approach, of using partial (not necessarily true) inferences, in order to assert a definite answer, is highly characteristic of 'second generation' systems.

The demons are, as with Winograd's work, expressed in a procedural language which, on running, will seek for a succession of inter-related 'goals'.

Here, for example, is a demon concerned with another story, about a child's piggy bank (PB) and a child shaking it, looking for money but hearing no sound. The demon, PB-OUT-OF, is formalised as:

```
(DEMON PB-OUT-OF
        (NOLD PB PERSON M N)
        (?N OUT-O[redacted]?M ?PB)
        (GOAL (? IS ?PB PIGGY-BANK))
        (GOAL (? IS ?M MONEY) $DEDUCE)
        (GOAL (?NOLD SHAKE ?PERSON ?PB) $TRUE)
        (ASSERT (? HAVE ?PERSON ?M)
        (ASSERT (? RESULT ?N ?NOLD)))
```

Again, it is not necessary to explain the notation in detail to see that conditions are being stated for the contents of a piggy bank having been emptied. The pattern being sought by the demon in operation is the third line. If a chain of demons can 'reach' one of the possible referents in a story then there is a success registered and the ambiguity of the corresponding pronoun is resolved.

It can be seen that the information encoded in the system is of a highly specific sort - in the present case it is not about containers as such, and how to get their contents out, but about Piggy Banks in particular, and everything relies on that particular knowledge having been put in. Not all the knowledge is of this general sort: in a recent paper (Charniak '74) where the 'kite' story is reconsidered there are rules of considerable generality and interest. One such is that Charniak calls a R+SSA rule:

"If the story gives information which would make it plausible to infer that PERSON is favourably inclined towards action A, and PERSON does S, a significant subaction (SSA) of A, then infer that PERSON is doing A

An important assumption of Charniak's is what I shall call his 'decoupling assumption', namely that this work on the knowledge-based determination of reference can be effectively decoupled from any particular syntactic or semantic representation, or of its application, and can be explored in isolation. Thus, the apparently English words in the PB-OUT-OF demon above do not have any particular status other than that of English words that happen to occur in the stories being examined. I shall return to this assumption later.

<u>Colby</u>

Colby has (Colby et al '71), (Enea and Colby '73), (Colby and Parkison, '74) developed over a period of several years an on-line dialogue system that simulates a paranoid patient called PARRY. A measure of its success is that it is perhaps the most used A.I. program in existence; used, that is, by those outside the research project that developed it. Here is a discontinuous sample of its output, again with the system's responses in the upper case:

Have you been hospitalised before?
THIS IS THE FIRST TIME.
How long have you been there?
ABOUT TWO WEEKS.
Any headaches?
MY HEALTH IS FINE.
Are you having any difficulty with your memory?
NO.
Then perhaps you can tell me who is the President of the United States?
NIXON IS PRESIDENT.
What day is it today?
TODAY IS THURSDAY.
What season is it?
LETS GET ON WITH THE INTERVIEW.

Colby's aim was to produce a system that was not 'fragile' in the way that most A.I. systems are: one that did not break down, or give up, if the input was not exactly what it expected; but one that carried on somehow in all circumstances, in the way that conversing humans do.

Fragility is avoided by having no conventional syntax analyser, and by not even attempting to take account of all the words in the input. This is a considerable aid, since any parser that begins to parse a more than usually polite request such as 'Would you be so kind as to ....' is going to be in trouble. British English speakers arriving in the U.S. quickly learn to delete such phrases, since they cause great confusion to human listeners in stores.

The input text is segmented by a heuristic that breaks it at any occurrence of a range of key words. Patterns are then matched with each segment. There are at present about 1700 patterns on a list (Colby and Parkison, in press) that is stored and matched, not against any syntactic or semantic representations of words but against the input word string direct, and by a process of sequential deletion. So, for example, "What is your main problem" has a root verb "BE" substituted to become

WHAT BE YOU MAIN PROBLEM.

It is then matched successively in the following forms after successive deletions:

BE YOU MAIN PROBLEM
WHAT YOU MAIN PROBLEM
WHAT BE MAIN PROBLEM
WHAT BE YOU PROBLEM
WHAT BE YOU MAIN

and only the penultimate line exists as one of the stored patterns, and so is matched. Stored in the same format as the patterns are rules expressing the consequenc3s for the 'patient' of detecting aggression and over-friendliness in the interviewer's questions and remarks. The matched patterns found are then tied directly, or via these inference rules, to response patterns which are generated.

Enormous ingenuity has gone into the heuristics of this system, as its popularity testifies. The system has also changed considerably: it is now called PARRY2 and contains the above pattern-matching, rather than earlier key work, heuristics. It has the partial, or what some would call 'pragmatic', rules about expectation and intention, and these alone might qualify it as 'second generation' on some interpretations of the phrase. A generator is also being installed to avoid the production of only 'canned' responses.

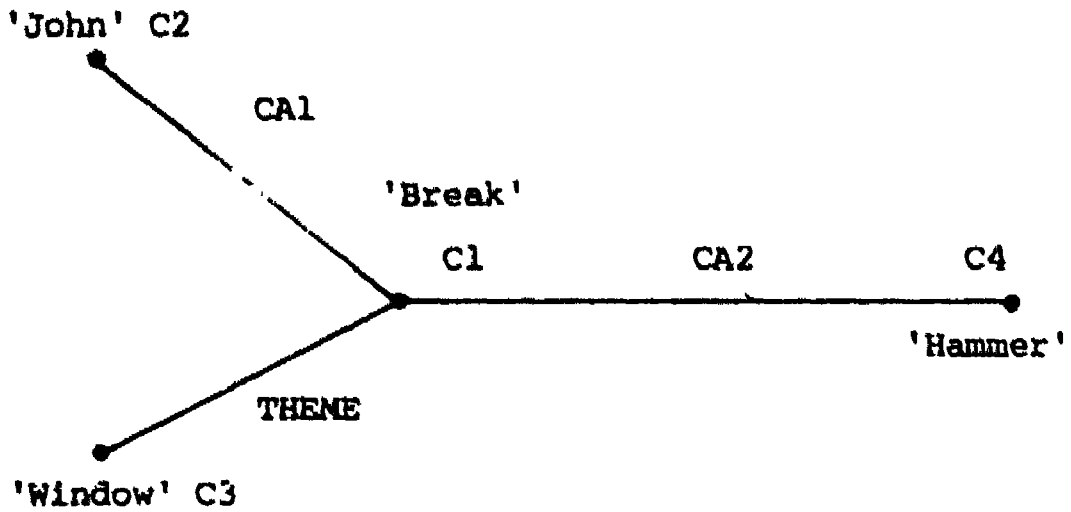
Colby and his associates have put considerable energy into actually trying to find out whether or not psychiatrists can distinguish PARRY'S responses from those of a patient (Colby and Hilf '73). This is probably the first attempt actually to apply Turing's test of machine-person distinguishability. There are statistical difficulties about interpreting the results but, by and large, the result is that the sample questioned cannot distinguish the two. Whether or not this will influence those who still, on principle, believe that PARRY is not a simulation because it 'does not understand', remains to be seen. It might be argued that they are in danger of falling into a form of Papert's 'human-superhuman fallacy' of attacking machine simulations because they do not perform superhuman tasks, like translate poetry, tasks that some people certainly can do but the majority cannot. When such sceptics say that PARRY does not understand they have in mind a level of understanding that is certainly high - one could extend their case ironically by pointing out that very few people understand the content of sentences in the depth and detail that an analytic philosopher does, and a very good thing too. But there can be no doubt that many people on many occasions DO seem to understand in the way that PARRY does.

Simmons

The remaining three systems differ from the two above in their attempt to provide some representational structure quite different from that of the English input. This means the use of cases, and of complex structures that allow inferences to be drawn from the attribution of case in ways I shall explain. There is also, in the remaining systems, some attempt to construct a primitive, or reduced, vocabulary into which the language represented is squeezed.

Simmons' work is often thought of as a 'memory moded', though he does in fact pay more attention to word sense ambiguity, and to actual recognition in text than do many other authors. For him the fundamental notion is that of a 'semantic network', defined essentially by the statement of relational triples of form aRb, where R is the name of a relation and a and b are the names of nodes in the network. Simmons' work with this general formalism goes back to at least (Simmons et al, '66) but, in its newer form with case formalism, it has been reported since 1970 (Simmons '70b), (Simmons and Bruce '71), (Simmons and Slocum '72), (Simmons '73), and (Hendrix et al '73) may reasonably be considered a further implementation of Simmons' methods.

Simmons considers the example sentence 'John broke the window with a hammer'. This is analysed into a network of nodes C1, C2, C3, C4 corresponding to the appropriate senses of 'John', 'Bread', 'Window' and 'Hammer' respectively. The linkages between the nodes are labelled by one of the following 'deep case relations': CAUSAL-ACTANT (CA., CA2), THEME, LOCUS, SOURCE and GOAL. Case relations are specifications of the way dependent parts of a sentence, or concepts corresponding to parts of a sentence, depènd on the main action. So, in this example, John is the first causal actant (CA1) of the breaking, the hammer is considered the second causal actant (CA2) of that breaking, and the window is the theme of the breaking. Thus, the heart of the analysis could be represented by a diagram as follows:



or by a set of relational triples:

(C1 CA1 C2) (C1 CA2 C4) (C1 THEME C3)

However, thisis not the full representation, and my addition of the word labels to the diagram is misleading, since the nodes are intended to be names of senses of words, related to the actual occurrence of the corresponding word in a text by the relation TOK (for token). In an implementation, a node would have an arbitrary name, such as L97, which would then name a stored sense definition. So, for a sense of 'apple Simmons suggests an associated set of features: NBR-singular (S), SHAPE-spherical, COLOR-red, PRINTIMAGE-apple, THEME-eat, etc. If the name of the node tied to this set of features was indeed L97, then that node might become, say, C5 on being brought into some sentence representation during parsing. Thus the diagram I gave must be thought of supplemented by other relational ties from the nodes; so that the full sentence about John would be represented by the larger set of triples:

(C1 TOK break) (C1 CA1 C2) (C1 THEME C3) (C1 CA2 C4)

(C2 TOK John) (C2 DET Def) (C2 NBR S)

(C3 TOK Window) (C3 DET Def) (C3 NBR S)

(C4 TOK Hammer) (C4 DET Indef) (C4 NBR S) (C4 PREP With)

Word sense ambiguity is taken account of in that the node for one sense of 'hammer' would be different from that corresponding to some other sense of the same word, such as that meaning Edward, Hammer of the Scots, to take a slightly strained alternative for this sentence.

The network above is also a representation of the following sentences which can be thought of as surface variants of a single 'underlying' structure:

John broke the window with a hammer

John broke the window

The hammer broke the window

The window broke.

Not all parts of that network will be set up by each of these sentences, of course, but the need for some item to fill an appropriate sl t can be inferred; i.e. of the first causalactant (John) in the last two sentences. The sentences above are recognised by means of the 'ergative paradigm' of ordered matching patterns, of which the following list is a part:

(CA1 THEME CA2)

(CA1 THEME)

(CA2 THEME)

(THEME)

These sequences will each match, as left-right ordered items, one of the above sentences. It will be clear that Simmons' method of ascribing a node to each word-sense is not in any way a primitive system, by which I mean a system of classifiers into which all word senses are mapped.

Simmons is, however, considering a system of paraphrase rules that would map from one network to another in a way that he claims is equivalent to a system of primitives. Thus in (Simmons '73) he considers the sentences

John bought the boat from Mary

Mary sold the boat to John

which would normally be considered approximate paraphrases of each other. He then gives 'natural' representations, in his system, as follows in the same order as the sentences:

---

* Simmons' normal example of word sense ambiguity does not apply to the sentence above: he distinguishes 'pitcher1', a pouring container, from 'pitcher2', in the U.S. sense of 'one who bowls a ball'.

C1 TOK buy, SOURCE (Mary), GOAL (John), THEME (boat),

C1 TOK sell, SOURCE (Mary), GOAL (John), THEME (boat),

and also the single representation for both sentences, as below, using a primitive action 'transfer' 9see description of Schank's work in next section) as follows:

C1 TOK and, Args C2, C3

C2 TOK transfer, SOURCE (John), GOAL (Mary), THEME (money)

C3 TOK transfer, SOURCE (Mary), GOAL (John), THEME (boat)

Simmons opts for the first form of representation, given the possibility of a transfer rule going from either of the shallower representations to the other, while in (Hendrix et al '73), the other approach is adopted, using a primitive action EXCHANGE instead of 'transfer'.

The implementation under construction is a front-end parser of the Woods' augmented transition network type (see Woods '70), and a generation system going from the semantic networks to surface strings described in detail in (Simmons and Slocum '72). Simmons has also given considerable time (Simmons and Bruce '71) to the automatic translation of the networks into a crorresponding first-order predicate calculus format of the sort developed by Sandewall (1971). This last is particularly valuable because, if generalisable, it shows that any linguistic coding in network form can be translated into some form of the predicate calculus, if that formalism and its associated proof techniques can be shown to be appropriate for certain problems in the area of natural language analysis.

Schank

Schank's is a rich system of semantic representation, developed over a period of six years, with the collaboration of a number of talented students. Its graph system of notation has influenced psychologists like Abelson (1973), among others. Schank's contribution has been the notational system representing the structure of natural language sentences, and this has been programmed by various collaborators over the years. In its present version, called MARGIE (Schank et al '73) it has an analyser of English due to Riesbeck (1974) a semantic memory component due to Rieger (1974), and a generator of English due to Goldman (1974) MARGIE produces output in two modes, demonstrating the sort of conceptual inferencing that goes on at the level of the semantic representation*: The PARAPHRASE mode and the INFERENCE mode. Samples of input and output to and from the two modes can be shown thus:

(INFERENCE MODE)

INPUT: John told Mary that Bill wants a book.

OUTPUT: A book about what.

Mary knows that Bill wants a book.

Bill wants to come to have a book.

Bill wants someone to cease to have a book.

Bill wants to read a book.

(PARAPHRASE MODE)

INPUT: John killed Mary by choking Mary.

OUTPUT: John strangled Mary.

John choked Mary and she died because she could not breathe.

Mary died because she was unable to inhale some air and she was unable to inhale some air because John grabbed her neck

The aim of Schank's system has always been to provide a representation of meaning in terms of which these and other tasks, such as machine translation, could be carried out; a representation, moreover, that was to be independent of any particular language, and of syntax, and indeed, of all surface structure whatever.

The formal structure of Schank's graphs is that of dependency grammar (Hays '64), and the items in the graph are of four types, or conceptual categories. They are symbolised as PP, ACT, PA and AA, which are acronyms, but which correspond closely (for the purpose of understanding their function) to those of a noun, verb, adjective and adverb, respectively.** The basic

---

* Schank distinguishes 'conceptual' and 'semantic' representations in a way that is important for him within his own system. However, I shall use the terms indifferently since, in this brief and superficial description, nothing hangs upon the distinction.

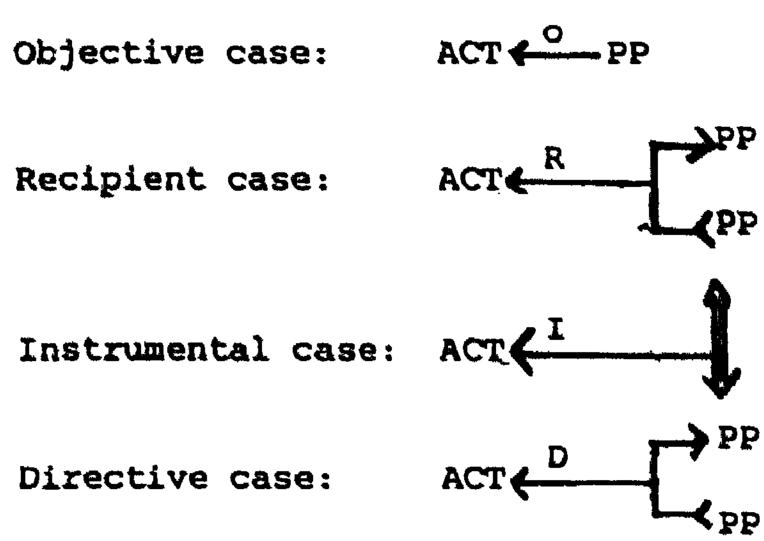
**This is a considerable oversimplification, made in order to give a brief and self contained description. But, in fact, many English nouns are represented as ACT's: chair, pen, honesty, and transportation.

structure is called a conceptualization, and is normally introduced with a straightforward dependency structure such as, for the sentence 'The man took a book':

$$\text{Man} \overset{p}{\Leftrightarrow} \text{take} \overset{o}{\leftarrow} \text{book}$$

Here 'p' indicates past, and is the dependency symbol liking a PP to the ACT ('take') which is the hub of the conceptualization, as with Simmons. The 'o' indicates the objective case, marking the dependence of the object PP on the central ACT. There is a carefully constructed syntax of linkages between the conceptual categories, that will be described only in part in what follows.

The next stage of the notation involves an extended case notation and a set of primitive ACTs, as well as a number of items such as PHYSCONT which indicate other states, and items of a fairly simplified psychological theory (the dictionary entry for 'advise', for example, contains a subgraph telling us that Y 'will benefit' as part of the meaning of 'X advises Y' (Schank '73). There are four cases in the system, and their subgraphs are as follows:

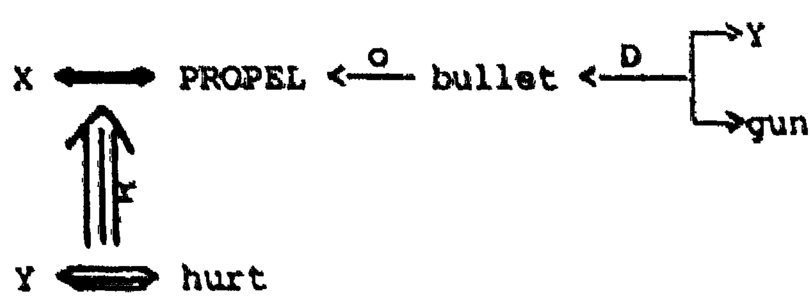


There are at present fourteen* basic actions forming the nubs of the graphs, as well as a default action DO. They are: PROPEL, MOVE, INGEST, EXPEL, GRASP, PTRANS, MTRANS, ATRANS, SMELL, SPEAK, LOOK-AT, LISTEN-TO, CONC and MBUILD. The notions of case and primitive act are related by rules in the development of conceptualizations. So, for example, the primitive act INGEST has as its instrument the act PTRANS. There are also other infer-

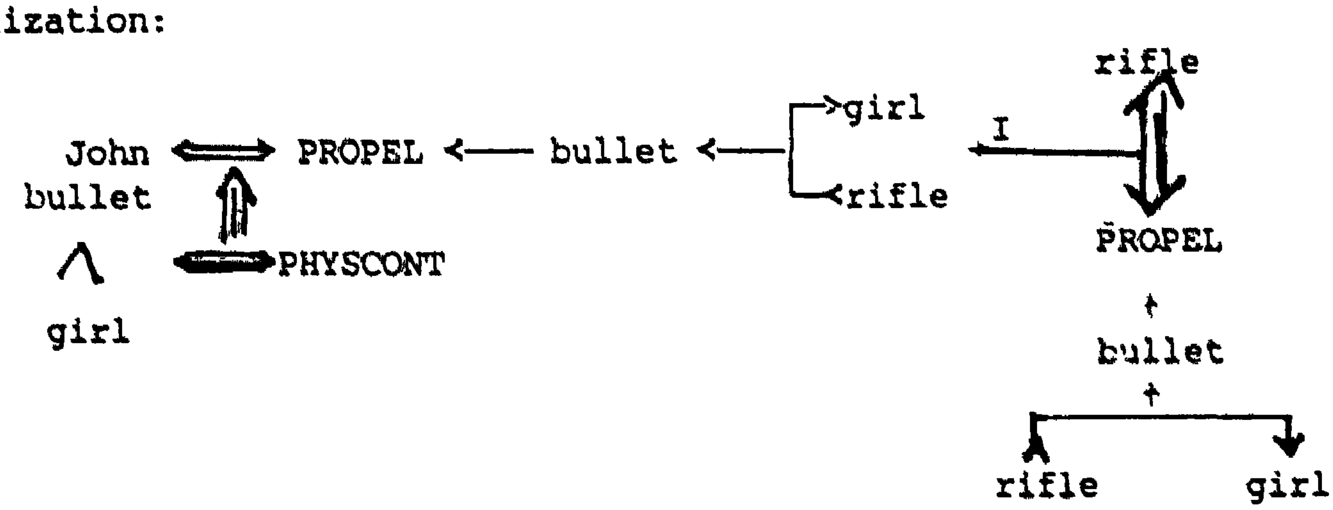
---

* Since the publication of (Schank 73a) their number has been reduced to eleven (plus DO) by the elimination of SMELL, LISTENTO, LOOKAT and CONC, and the addition of ATTEND.

ences from any ACT classified as an INGEST action, such as that the thing ingested changes its form; that ~~if the thing ingested is edible the~~ in gester becomes 'more nourished' etc. (see Schank '73, pp. 38ff.). This will all become clearer if we consider the transition from a dictionary entry for an action to a filled-in conceptualization. Here is the dictionary entry for the action 'shoot':



We can consider this entry as an active 'frame-like' object seeking filler items in any context in which it is activated. Thus, in the sentence 'John sho the girl with a rifle', the variables will be filled in from context and the case inference will be made from the main act PROPEL, which is that its instrument is MOVE, GRASP or PROPEL, and so we will arrive at the whole conceptualization:



This case inference must be made, according to Schank, in order to achieve an adequate representation. There is, in the last diagram, a certain redundancy of expression, but as we shall see in the next section this often happens with deeper semantic notations.

More recently, Schank, together with Rieger, has developed a new class of causal inferences which deepen the diagrams still further. So, in the analysis of 'John's cold improved because I gave him an apple' (in Schank '74a) the extended diagram contains at least four yet lower levels of causal arrowing, including one corresponding to the notion of John constructing the idea (MBUILD) that he wants to eat an apple. So we can see that the underlying explication of meaning here is not only in the sense of linguistic primitives, but in terms of a theory of mental acts as well.

Now there are a number of genuine expositional difficulties here for the commentator faced with a system of this complexity. One aspect of this is the stages of development of the system itself, which can be seen as a consistent process of producing what was argued for in advance. For example, Schank claimed early on to be a constructing system of semantic structures underlying the 'surface of natural language', although initially there were no primitives at all, and as late as (Schank et al '70) there was only a single primitive TRANS, and most of the entries in the dictionary consisted of the English words coded, together with subscripts. Since then the primitive system has blossomed and there are now twelve primitives for ACTS including three for the original TRANS itself. Each exposition of the system recounts its preceding phrases, from the original primitive-free one, through to the present causal inference form; rather as each human foetus is said to relive in the womb all the evolutionary stages of the human race. The only trouble with this, from an outsider's point of view, is that at each stage the representation has been claimed, to be the correct one, while at the same time Schank admits, in moments of candor (Schank '73), that there is no end to the conceptual diagramming of a sentence. This difficulty may well reflect genuine problems in language itself, and, in its acutest form concerns a three-way confusion between an attractive notation for displaying the 'meanings of words', the course of events in the real world, and, finally, actual procedures for analysis to be based on the diagrams.

This raises the, to me, important question of the application of a semantic system, that I shall touch on again later. Schank, for example, does mention in passing the questions of word-sense ambiguity, and the awful ambiguity of English prepositions, but there are in no way central for him, and he assumes that with the availability of 'the correct representation', his system when implemented must inevitably solve these traditional and vexing questions. No procedures are hinted at along with the graphs as to how this is to be done. A distinction of importance may be becoming apparent here between Schank s work and Rieger's: in Rieger's thesis (Rieger '74) the rules of inference appear to create separate and new subgraphs wnicn may stand in an inferential relation to each other so as to produce conclusions about problems of, say, pronoun reference, etc. But in Schank's corresponding papers the same inferences are not applied to actual problems (Schank '74a) but only serve to complexify the conceptual graphs yet further.

Closely connected with this matter is the question of the survival of the surface structure in the diagrams. Until very recently primitivisation applied only to verbs, that of nouns being left to Weber (Weber '72) Most recently, though, noun words have been disappearing from diagrams and been replaced by categories such as *PHYSOBS* But it is clear that the surface is only slowly disappearing, rather than having been abhorred all along.

In a more recent publication (Schank '74b) there are signs that this trend of infinitely proliferating diagrams (for individual sentences) is reversing. In it Schank considers the application of his approach to the representation of text, and concl:des, correctly in my view, that the representations of parts of the text must be interconnected by causal arrows, and that, in order to preserve lucidity, the conceptual diagrams for individual sentences and their parts must be abbreviated, as by triples such as POEPLE PTRANS PEOPLE. Here indeed, the surface simply <u>has</u> to survive in the representation unless one is prepared to commit oneself to the extreme view that the ordering of sentences in a text is a purely superficial and arbitrary matter. The sense in which this is a welcome reversal of a trend should be clear, because in the 'causation inference' development, mentioned earlier, <u>all</u> the consequences and effects of a conceptualization had to be drawn within itself. Thus, in the extreme case, <u>each sentence</u> of a text should have been represented by a diagram containing most or all of the text of which it was a part. Thus the representation of a text would have been impossible on such principles.

<u>Wilks</u>

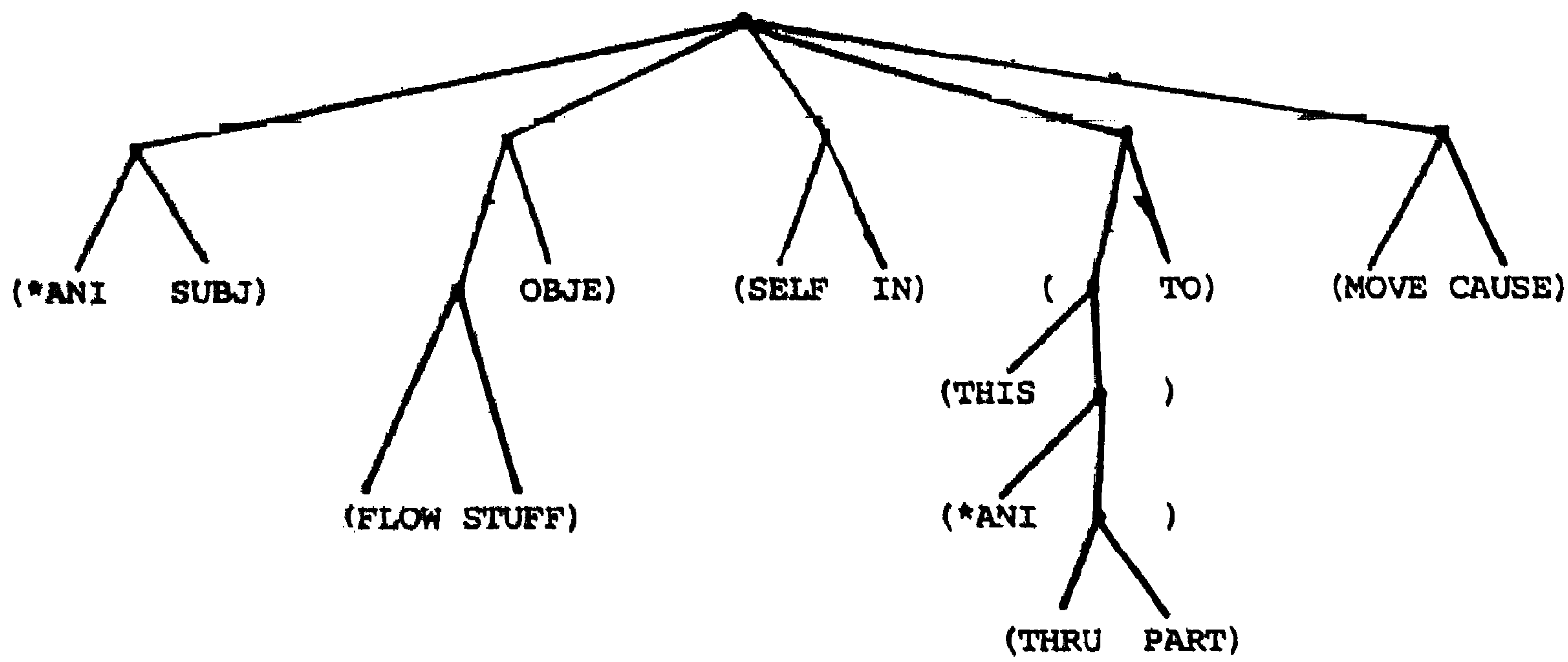
My own system constructs a semantic representation for small natural language texts: the basic representation is applied directly to the text and can then be 'massaged' by various forms of inference to become as deep as is necessary for well defined tasks demonstrating understanding. It is a uniform representation, in that information that might convenionally be considered as syntactic, semantic, factual or inferential is well expressed within a single type of structure. The fundamental unit in the construction of this meaning representation is the template, which is intended to correspond to an intuitive notion of a basic message of agent-action-object form. Templates are rigid format networks of more basic building blocks called formulas, which correspond to senses of individual words. In order to construct a complete text representation templates are bound together by two kinds of higher level structures called paraplates and inference rules. The templates themselves are built up as the construction of the representation proceeds, but the formulas, paraplates and inference rules are all present in the system.at the outset and each of these three types of pre-stored structure is ultimately constructed from an inventory of eighty semantic primitive elements, and from functions and predicates ranging over those elements.

The system runs on-line as a package of LISP, MLISP and MLISP2 programs, taking as input small paragraphs of English, that can be made up by the user from a vocabulary of about 600 word senses, and producing a good French translation as output. This environment provides a pretty clear test of language understanding, because French translations for everyday prose are either right or wrong, and can be seen to be so, while at the same time, the major difficulties of understanding programs - word sense ambiguity, case ambiguity, difficult pronoun reference, etc. - can all be represented within a machine translation environment by, for example, choosing the words of the input sentence containing a pronoun reference difficulty so that the possible alternative references have different genders in French. In that way the French output makes quite clear whether or not the program has made the correct inferences in order to understand what it is translating. The program is reasonably robust in actual performance, and will even tolerate a certain amount of bad grammar in the input, since it does not perform a syntax analysis in the conventional sense, but seeks message forms representable in the semantic structures employed.

Typical input would be a sentence such as 'John lives out of town and drinks his wine out of a bottle. He then throws the bottles out of the window. ~~The program will produce French sentences with different output~~ for each of the three occurrences of 'out of', since it realises that they function quite differently on the three occasions of use, and that the difference must be reflected in the French. A sentence such as 'Give the monkeys bananas although they are not ripe because they are very hungry' produces a translation with different equivalents for the two occurrences of 'they', because the system correctly realises, from what I shall describe below as preference considerations, that the most sensible interpretation is one in which the first 'they' refers to the bananas and the second to the monkeys, and bananas and monkeys have different genders in French. These two examples are dealt with in the 'basic mode' of the system. (Wilks 73a) In many cases it cannot resolve pronoun ambiguities by the sort of straightforward 'preference considerations' used in the last example, where, roughly speaking, 'ripeness' prefers to be predicated of plant-like things, and hunger of animate things. Even in a sentence as simple as 'John drank the wine on the table and it was good', such considerations are inadequate to resolve the ambiguity of 'it' between wine and table, since both may be good things. In such cases, of inability to resolve within its basic mode, the program deepens the representation of the text so as to try and set up chains of inference that will reach, and so prefer, only one of the possible referents. I will return to these processes in a moment, but first I shall give some brief description of the basic representation set up for English.

For each sense of a word in its dictionary the program sees a formula. This is a tree structure of semantic primitives, and is to be interpreted formally using dependency relations. The main element in any formula is the rightmost, called its head, and that is the fundamental category to which the formula belongs. In the formulas for actions, for example, the head will always be one of the primitives PICK, CAUSE, CHANGE, FEEL, HAVE, PLEASE, PAIR, SENSE, USE, WANT, TELL, BE, DO, FORCE, MOVE, WRAP, THINK, FLOW, MAKE, DROP, STRIK, FUNC or HAPN.

Here is the tree structure for the action of drinking:



Once again, it is not neeessary to explain the formalism in any detail, to see that this sense of 'drink' is being expressed as a causing to move a liquid object (FLOW STUFF) by an animate agent, into that same agent (containment case indicated by IN, and formula syntax identifies SELF with the agent) and via (direction case) an aperture (THRU PART) of the agent.

Template structures, which actually represent sentences and their parts are built up as networks of formulas like the one above. Templates always consist of an agent node, and action node and an object node, and other nodes that may depend on these. So, in building a template for 'John drinks, wine', the whole of the above tree-formula for 'drinks' would be placed at the central action node, another tree structure for 'John' at the agent node and so on. The complexity of the system comes from the way in which the formulas, considered as active entities, dictate how other places in the same template should be filled.

Thus, the 'drink' formula above can be thought of as an entity that fits at a template action node, and seeks a liquid object, that is to say a formula with (FLOW STUFF) as its right-most branch, to put at the object node of the same template. This seeking is preferential, in that formulas not satisfying that requirement will be accepted, but only if nothing satisfactory can be found. The template finally established for a fragment of text is the one in which the most formulas have their preferences satisfied. There is a general principle at work here, that the right interpretation 'says the least' in information-carrying terms. This very simple device is able to do much of the work of a syntax and word-sense ambiguity resolving program For example, if the sentence had been 'John drank a whole pitcher', the formula for the 'pitcher of liquid' would have been preferred to that for the human, since the subformula (FLOW STUFF) could be appropriately located within it.

A considerable amount of squeezing of this simple canonical form of template is necessary to make it fit the complexity of language: texts have to be fragmented initially; then, in fragments which are, say, prepositional phrases there is a dummy agent imposed, and the prepositional phrases there is a dummy agent imposed, and the prepositional formula functions as a pseudo-action. There are special 'less preferred' orders to deal with fragments not in agent-action-object order, and so on.

When the local inferences have been done that set up the agent-action object templates for fragments of input text, the system attempts to tie these templates together so as to provide an overall initial structure for the input. One form of this is the anaphora tie, of the sort discussed for the monkeys and bananas example above, but the more general form is the case tie. Assignment of these would result in the template for the last clause of 'He ran the mile in a paper bag' being tied to the action node of the template for the first clause ('He ran the mile'), and the tie being labelled CONTainment. These case ties are made with the aid of another class of ordered structures, essentially equivalent to Fillmore's case frames, called <u>paraplates</u> and which are attached to the formulas for English prepositions. So, for 'outof', for example, there would be at least six ordered paraplates, each of which is a string of functions that seek inside templates for information. In general, paraplates range across two, not necessarily contiguous, templates. So, in analysing 'He put the number he thought of in the table', the successfully matching paraplate would pin down the dependence of the template for the last of the three clauses as DIREction, by taking as argument <u>only that particular template for the last clause</u> that contained the formula for 'a numerical table', (and <u>not</u> a template representing a kitchen table) and it would do that because of a function in that paraplate seeking a similarity of head (SIGN in this case) between the two appropriate object formulas, for 'number' and 'table'. The other template containing the 'furniture' formula for 'table' would naturally not satisfy the function because SIGN would not be the head of this sense formula for 'table'.

The structure of mutually connected templates that has been put together thus far constitutes a 'semantic block', and, if it can be constructed, then as far as the system is concerned all semantic and referential ambiguity has been resolved and it will begin to generate French by unwrapping the block again. The generation aspects of this work have been described in (Herskovits '73). One aspect of the general notion of preference is that the system should never construct a deeper or more elaborate semantic representation than is necessary for the task in hand and, if the initial block can be constructed and a generation of French done, no 'deepening' of the representation will be attempted.

However, many examples cannot be resolved by the methods of this 'basic mode' and, in particular, if a word sense ambiguity, or pronoun reference, is still unresolved, then a unique semantic block of templates cannot be constructed and the 'extended mode' will be entered.* In this mode, new template-like forms are extracted from existing ones, and then added to the template pool from which further inferences can be made. So, in the template derived earlier for 'John drinks wine', the system enters the formula for 'drinks', and draws inferences corresponding to each case sub-formula. In this example it will derive template-like forms equivalent to, in ordinary English, 'The wine is in John', 'The wine entered John via an aperture' and so on. The extracted templates express information already implicitly present in the text, even though many of them are partial inferences: ones that may not necessarily be true.

Common-sense inference rules are then brought down, which attempt, by a simple strategy, to construct the shortest possible chain of rule-linked template forms from one containing an ambiguous pronoun, say, to one containing one of its possible referents. Such a chain then constitutes a solution to the ambiguity problem, and the preference approach assumes that the shortest chain is always the right one. So, in the case of 'John drank the wine /on the table/ and it was good', (in three template-matching fragments as shown) the correct chain to 'wine' uses the two rules

* Wilks '73b, and (in press)

I 1. ((*ANI 1) ((SELF IN) (MOVE CAUSE)) (*REAL 2)) → (1(*JUDG) 2)

or, in 'semi-English

[animate-1 cause-to-move-in-self object-2] → [1 *judges 2]

I 2. (1 BE (GOOD KIND)) ↔ ((*ANI 2) WANT 1)

or, again,

[1 is good] ↔ [animate-2 wants 1]

These rules are only partial, that is to say, they correspond only to what we may reasonably look out for in a given situation, not to what MUST happen. The hypothesis here is that understanding can only take place on the basis of simple rules that are confirmed by the context of application. In this example the chain constructed may be expressed as (using the above square bracket notation to contain not a representation, but simply an indication, in English, of the template contents):

forwards inf ↓	[John drank	the-wine]	Template 1
	[John causes-to-move-in-self	wine]	‡ Template 1
	[John * judges	wine]	by I 1
backwards inf ↑	[John wants	wine]	‡ line above
	[wine is	good]	by I .
	[?it is	good]	Template 3

The assumption here is that no chain using other inference rules could have reached the 'table' solution by using less than two rules.

The chief drawback of this system is that codings consisting entirely of primitives have a considerable amount of both vagueness and redundancy For example, no reasonable coding in terms of structured primitives could be expected to distinguish, say, 'hammer' and 'mallet'. That may not matter provided the codings can distinguish importantly different senses of words. Again, a template for the sentente 'The sheperd tended his flock' would contain considerable repetition, each node of the template trying, as it were, to tell the whole story by itself. Again, the preference criteria are not in any way weighted, which might seem a drawback, and the preferential chain length criterion for inference chains might well seem too crude. Whether or not such a system can remain stable with a considerable vocabulary, of say several thousand words, has yet to be tested.

It will be evident to any reader that the last two systems described, Schank's and my own, share a great deal in common. Even the apparent difference in notation is reduced if one sees the topological similarity that results from considering the head of a formula as functioning rather like a Schank basic action. If one thinks of the dependencies of the case subparts of a formula, not arranged linearly along the bottom of a tree, but radiating out from the head in the centre, then the two diagrams actually have identical topologies under interpretation. A difference arises in that the 'filled-in entity' for Schank is the conceptualization centred on the basic action, though for me it is the network of formulas placed in relation in a template, where there is indeed a basic action, the head of the action formula, but there is also a basic entity in the agent formula and so on. Or, to put it another way, both what-is and what-is-expected are represented in the templates: the agent formula represents the agent, for example, but the left-hand part of the action formula also represents what agent was expected or sought, as in the (*ANI SUBJ) sub-formula of the 'drink' formula.

Although developed in isolation initially, these two systems have also influenced each other in more recent years, probably unconsciously. For example, conceptual dependency now emphasises the agent-action-object format more than before, and is less 'verb-centred' and timeless, while, conversely, my own system now makes much more overt use of rules of partial information than in its earlier versions. Again, both systems have intellectual connections that go back before either generation of AI systems. In my view, both these systems have roots in the better parts of the Computational Linguistics movement of the Fifties: in the case of Schank's system, one may think of the earlier systems of (Hays '64) and (Lamb '66), and the arrow-structured primitive system of (Farradene '66) In the case of my own system there are clear precedents in the Parker-Rhodes '61) system of classification and the early semantic structures of (Richens '61) and (Masterman '61). In 1961 the last author was arguing that 'what is needed is a discipline that will study semantic message connection in a way analogous to that in which metmathematics now studies mathematical connection, and to that in which mathematical linguistics now studies syntactic connection'. (ibid., p.3)

This historical point raises a final one that is, I feel, of passing interest. There seem to be two research styles in this field: one is what might be called the 'fully finished style', in which the work exists only in one complete form, and is not issued in early or developed versions. The best example of this is Winograd's work. The other type, exemplified by all the other authors discussed here, to some extent, is the developing style: work which appears in a number of versions over the years, one hopes with gradual improvements, perhaps in attempts to tackle a wider range of linguistic or other inferential phenomena. There are advantages to both styles, but even in the latter one knows that any proposed structure or system will, in the end, be found wanting in the balances of language, so it can only be a question of when one will have to abandon it. The interesting question, and one to which no answer could possibly be given here, is just how far is it worth pushing any given structural approach before starting again from scratch?

6. Some Comparisons and Contrasts

In this section I shall compare and contrast, under some nine interconnected headings, the projects described in the body of the paper. This is not easy to do, particularly when the present author is among the writers discussed, though that is easily remedied by the reader's making an appropriate discount. A more serious problem is ths , at this stage of research in artificial intelligence and natural language, the most atttractive distinctions dissolve on more detailed scrutiny, largely because of the lack of any precise theoretical statement in most, if not all, the major projects. There are those who think that it therefore follows that this is not the moment for any form of critical comparison in this field, and that no more is needed than a 'positive attitude' towards all possible projects. Only those who feel that, on the contrary, any time is as good as any other for the discussion of intellectual differences in the hope of progress, should read on.

It must be admitted right away that the selection of projects discussed above, like Winograd's distinction between first and second generation systems, on which the selection was to some extent based, cannot be defended by any strict definitions: one that would, in this case, include all the projects described, and exclude all those of Winograd's 'first generation' One might, for example, want to define second generation systems (in the study of natural language within the AI paradigm) in some very general terms, such as those systems which, (1) contain complex semantic structures for the representation of text that are significantly different from the 'surface structure' of the input, and (2) contain cognate structures representing conceptual and real world knowledge that is not explicitly present in the input text. Even so general a description of a 'frames' type approach would not cover Charniak or Colby with the first point, nor Simmons with the second, for he has so far eschewed all concern with information not present explicitly in the input text. Moreover, the second point would certainly cover Winograd's own work, as well as other first generation approaches, so it is clear, at the outset of any comparison, that there is not even a simple and unequivocal definition which covers all and only the projects to be compared.

Another factor making comparisons and contrasts harder is the wealth of different notations and diagrams used by the research workers discussed. That is to say, that the task of this paper would clearly be much easier if all the systems were expressed in PLANNER, production rules (Moore & Newell '73) or even Predicate Calculus. There are both good and bad reasons for avoiding a standard notation and here is perhaps not the place to explore these: a bad reason, for example, would be the fear that one's inference techniques were in fact some very standard theorem-proving technique and that the fact would be transparent were one's system expressed in Predicate Calculus.

Level of Representation

One important line of current dispute among the second generation approaches concerns the appropriate level of representation for natural language. On the one hand are those like Colby, and, in the same way, Charniak, who hold that the representation of language can, in effect, be by means of itself while, on the other hand, there are those like Schank and myself who hold that the appropriate level of computation for inferences about natural language is in some reduced, or primitive, representation. Simmons, as we saw, holds an intermediate position. Charniak holds that his structures are independent of any particular level of representation, or rather, that they could be realised at a number of levels of representation, depending on the subject area. However, there is no doubt that the representation in terms of predicates that he offers in his work appears to be in one-to-one correspondence with English words.

The strongest low-level approach is undoubtedly that of Colby who straightforwardly faces the enormous mapping problems involved if the structures are at the English word level. It is important to realise that this dispute is ultimately one of degree, since no one would claim that every locution recognized by an intelligent analyser must be mapped into a 'deep' representation. To take an extreme case, any system that mapped 'Good Morning' into a deep semantic representation before deciding that the correct response was also 'Good Morning' would be making a serious theoretical mistake.

However, themost serious argument for a non-superficial representation is not in terms of the avoidance of mapping difficulties, but in terms of theoretical perspicuity of the primitive structures, and this argument is closely tied to the defence of semantic primitives in general, which is a large subject not to be undertaken here. One of the troubles about semantic primitives is that they are open to bad defences, which decrease rather than increase their plausibility. For example, some users of them for linguistic representation have declared them to have some sort of objective existence and have implied that there is a 'right set' of primitives open to empirical discovery. On that view the essentially linguistic character of structures of primitives is lost, because it is an essential feature of a language that we can change its vocabulary or function with alternative vocabularies. But if there is a right set of primitives, whose members are the names of brain-items, then that essential feature would be lost.

What is the case is that there is a considerable amount of psychological evidence that people are able to recall the content of what they hear and understand without being able to recall either the actual words or the syntactic structure used. There is large literature on this subject, from which two sample references would be (Wettler '73) and (Johnson-Laird '74).

These results are, of course, no proof of the existence of semantic primitives, but they are undoubtedly supporting evidence of their plausibility, as is, on a different plane, the result from the encoding of the whole Webster's Third International Dictionary at Systems Development Corporation, where it was found that a rank-ordered frequency count of the words used to define other words in that vast dictionary was a list (omitting 'the' and 'a') which corresponded almost item-for-item to a plausible list of semantic primitives, derived a priori, by those actually concerned to code the structure of word and sentence meanings.

It is important to distinguish the dispute about level from the, closely connected, topic that I shall call the centrality of the knowledge required by a language understanding system.

Centrality

What I am calling the centrality of certain kinds of information concerns not its level of representation but its non-specificity: again a contrast can be drawn between the sorts of information required by Charniak's system, on the one hand, and that required by Schank's* and my own on the other. Charniak's examples suggest that the fundamental form of information is highly specific** to particular situations, like parties and the giving of presents, while the sorts of information central to Schank's and my own systems are general partial assertions about human wants, expectations, and so on, many of which are so general as to be almost vacuous which, one might argue, is why their role in understanding has been ignored for so long.

---

* Though as noted earlier, Schank in 1975 has adopted Abelson's (1973) notion of 'script', as a larger-scale 'frame', in such a way as to incorporate much less 'central' knowledge.

**In a recent paper (1974), Charniak gives much more general rules, such as his 'rule of significant sub-action', mentioned earlier.

If I were a reasonably fluent speaker of, say, German, I might well not understand a German conversation about birthday presents unless I had detailed factual information about how Germans organize the giving of presents, which might be considerably different from the way we do it. Conversely, of course, I might understand much of a technical article about a subject in which I was an expert, even though I knew very little of the language in which it was written. These are certainly considerations that tell for Charniak's approach, and it is perhaps a paradox that the sort of natural language understander that would tend to confirm his assumptions would be one concerned with discourse about, say, the details of repairing a motor car, where factual information is what is central, yet, ironically, Charniak has concentrated on something as general as children's stories, with their need of deep assumptions about human desires and behaviour.

In the end this difference may again turn out to be one of emphasis, and of what is most appropriate to different subject areas, though there may be a very general issue lurking somewhere here. It seems to me not a foolish question to ask whether much of what appears to be <u>about natural language</u> in A.I. research is in fact about language at all. Even if it is not that may in no way detract from its value. Newell (Moore, Newell '73) has argued that A.I. work is in fact 'theoretical psychology', in which case it could hardly be research <u>on</u> natural language. When describing Winograd's work earlier in the paper, I raised this question in a weak form by asking whether his definition of 'pickup' had anything to do with the natural language use of the word, or whether it was rather a description of how his system picked something up, a quite different matter.

Suppose we generalize this query somewhat, by asking the apparently absurd question of what would be <u>wrong</u> with calling, say, Charniak's work an essay on the 'Socio-Economic Behaviour of American Children Under Stress? In the case of Charniak's work this is a facetious question, asked only in order to make a point, but with an increasing number of systems in A.I. being designed not essentially to do research on natural language, but in order to have a natural language 'front end' to a system that is essentially intended to predict chemical spectra, or play snakes and ladders or whatever the question becomes a serious one. It seems to me a good time to ask whether we should expect advance in understanding natural language from those tackling the problems head on, or those concerned to build a 'front end'. It is clearly the case that <u>any piece of knowledge whatever</u> could be essential to the understanding of some story. The question is, does it follow that the specification, organization and formalization of that knowledge is the study of language, because if it is then all human enquiry from physics and history to medicine is a linguistic enterprise. And, of course, that possibility has actually been entertained within certain strains of modern philosophy.

However, I am not trying here to breathe fresh life into a philosophical distinction, between being about language and <u>not</u> being about language, but rather introducing a practical distinction, (which is also a consideration in favour of opting, as I have, to work on very general and central areas of knowledge) between specific knowledge, and central knowledge <u>without which a system could not be said to understand the language at all</u>. For example, I might know nothing of the arrangement of American birthday parties, but could not be accused of not understanding English even though I failed to understand some particular children's story. Yet, if I did not have available some very general partial inference such as the one people being hurt and falling, or one about people endeavouring to possess things that they want, then it is quite possible that my lack of understanding of quite simple sentences would cause observers to think that I did not understand English. An interesting and difficult question that then arises is whether those who concentrate on central and less central areas of discourse could, in principle, weld their bodies of inferences together in such a way as to create a wider system: whether, to put the matter another way, natural language is a whole that can be built up from parts?

<u>Phenomenological level</u>

Another distinction that can be confused with the central-specific one is that of the 'phenomenological levels' of inferences in an understanding system. I mean nothing daunting by the phrase: consider the action eating which is, as a matter of anatomical fact, quite often an act of bringing the bones of my ulna and radius (in my arm) close to that of my lower mandible (my jaw). Yet clearly, any system of common sense inferences that considered such a truth when reasoning about eating would be making a mistake. One might say that the phenomenological level of the analysis was wrong even though all the inferences it made were true ones. The same would be true of any A.I. system that made everyday inferences about physical objects by considering their quantum structure.

Schank's analysis of eating contains the information that it is done by moving the hands to the mouth, and it might be argued that even this is going too far from the 'meaning' of eating, whatever that may be, towards generally true information about the act which, if always inferred about all acts of eating, will carry the system unmanageably far.

There is no denying that this sort of information might be useful to have around somewhere; that, in Minsky's terms, the 'default' value of the instrument for eating is the hand brought to the mouth, so that, if we have no contrary information, then that is the way to assume that any given act of eating was performed. Nonetheless, there clearly is a danger, and that is all I am drawing attention to here, of taking inferences to a phenomenological level beyond that of common sense. A clearer case, in my view, would be Schank's analysis (1974a) of mental activity in which all actions, such as kicking a ball, say, are preceded by a mental action of conceiving or deciding to kick a ball. This is clearly a level of analysis untrue to common sense, and which can have only harmful effects in a system intended to mimic common sense reasoning and understanding.

Decoupling

Another general issue in dispute concerns what I shall call decoupling, which is whether or not the actual parsing of text or dialogue into an 'understanding system' is essential. Charniak and Minsky believe that this initial 'parsing' can be effectively decoupled from the interesting inferential work and simply assumed. But, in my view, that is not so, because many of the later inferences would actually have to be done already, in order to have achieved the initial parsing. For example, in analysing 'He shot her with a colt', we cannot ascribe any structure at all until we can make the inferences that guns rather than horses are instruments for shooting, and so such a sentence cannot be represented by an 'inference-but-no-parsing' structure, without assuming that language does not have one of its essential characteristics, namely systematic ambiguity. The essence of decoupling is allowing representational structures to have significance quite independent of their application, and that may lead one to a situation not essentially different from that of the logician who simply asserts that such-and-such is the 'right structure' of some sentence.

The inferences required to resolve word sense ambiguities, and those required to resolve pronoun reference problems, are not of different types; often the two problems occur in a single sentence and must be resolved together. But Charniak's decoupling has the effect of completely separating these two closely related linguistic phenomena in what seems to me an unrealistic manner. His system does inferencing to resolve pronoun ambiguities, while sense ambiguity is presumably to be done in the future by some other, ultimately recoupled, system.*

Modularity

Modularity concerns the decomposability of a program or system into (interacting) parts, and the nature of the relationship between the parts. Winograd's program, as we saw, contains syntactic, semantic and deductive segments which interact in a way he describes as 'heterarchic' (as opposed to 'hierarchic') which means that different segments can be in control at different times.

On the other hand, Schank and Wilks have argued that it is not necessary to observe either the syntactic-semantic, or the semantic-deductive, distinction in an understanding program. On that view there is no particular virtue in integrating syntax and semantic routines, since there was no need to separate them.

Charniak, however, would argue that, in some sense, one should make a syntax-semantics distinction here if one can. This would be consistent with his view on decoupling, and for him it would be convenient to decouple at a module, as it were, such as syntactic analysis. But decoup-

---

* Although Charniak would argue that sense ambiguity could be introduced into his system in its present form.

and strong modularity are not the same thing. Winograd's program, for example, is modular but not at all decoupled from surface text.

Availability of surface structure

An issue close to that of the appropriate level of representation in a system is that of the availability of the surface structure of the language analysed; or, to put it more crudely, the availability during subsequent analysis of the actual words being analysed. These are clearly available in Colby, and are indirectly available in Simmons', Winograd's and my own system, but Schank makes a point of the importance of their non-availability, on the grounds that an ideal representation should be totally independent of the input surface structure and words. There are both theoretical and practical aspects to this claim of Schank's: in the limit, the order of the sentences of a text is part of its surface structure, and presumably it is not intended to abandon this 'superficial information' In one of his recent papers 91974b) Schank seems to have accepted some limitation on the abandonment of surface structure.

The other, practical, point concerns the form of representation employed: in the (1973) implementation of Schank's system using an analyser of input text, a memory and a generator of responses, it was intended that nothing should be transferred from the input program to the output program except a representation coded in the structures of primitives discussed earlier.* The question that arises is, can that structure specify and distinguish word-senses adequately without transferring information specifically associated with the input word? Schank clearly believes the answer to this question is yes, but that cannot be considered established by the scale of computations yet described in print.

A suitable environment in which to consider the question is that of translation from one language to another: suppose we are analysing a sentence containing the word 'nail' meaning a physical object. It is clear that the translation of that word into French should not be the same

* This point is to some extent hypothetical since, as we saw, Schank's conceptualizations still do contain, or appear to contain, many surface items; in particular nouns, adjectives and adverbs. However, this is a transitional matter and they are in the course of replacement, as noted, by non-superficial items.

as the translation for 'screw or 'peg'. Yet is it plausible that any description of the function of these three entities entirely in terms of semantic primitives, and without any explicit mention of the word name and its connection to its French equivalent, will be sufficient to ensure that only the right match is made?

Application

This point is a generalisation of the last two, and concerns the way in which different systems display, in the structures they manipulate, the actual procedures of application of those structures to input text or dialogue. This is a matter different from computer implementation of the system. In the case of Colby's patterns, for example, the form of their application to the input English is clear, even though the matching involved could be achieved by many different implementation algorithms. In the case of my own system, I hold the same to be true of the template structures, even though by the time the input has reached the canonical template form it is considerably different from the input surface structure. The system at the extreme end of any scale of perspicuity of application is Winograd's where the procedural notation, by its nature, tries to make clear the way in which the structures are applied. At the other end are the systems of Schank and Charniak, where no application is specified, which means that the representations are not only compatible with many implementation algorithms, which does not matter, but are also compatible with many systems of linguistic rules, whose specification is an essential piece of inquiry, and whose subsequent production may cause the basic system to be fundamentally different.

Application is thus different from decoupling, for Schank's system is clearly coupled to language text by Riesbeck's parser, though his structures do not express their own application to language text.

English prepositions will serve as an example: in Schank's case notation there is no indication of how the case discriminations are actually to be applied to English prepositions in text. So, for example, the preposition 'in' can correspond to the containment case, time location, and spatial location among others. As we saw earlier, the discrimination involved in actual analysis is a matter of specifying very delicate semantic rules ranging over the basic semantic structures employed. Indeed, the structures and case system themselves seem to me to be essentially dependent on the nature and applicability of such rules, and so this application of the system should have an obvious place in the overall structures. It is not something to be delegated to a mere 'implementation' If enough of the linguistic intractables** of English analysis were to be delegated out of the representation, A.I. would be offering no more to the analysis of natural language than the logicians who proffer the predicate calculus as a plausible structure for English.

In some of his more recent writings Winograd has begun to develop a view that is considerably stronger than this 'application' one: in his view the control structure of an understanding program is itself of theoretical significance, for only in that way, he believes, can natural language programs of great size and complexity remain perspicuous.

Forward inference

Another outstanding dispute concerns whether one should make massive forward inferences as one goes through a text, keeping all one's expectations intact, as Charniak and Schank hold, or whether, as I hold, one should adopt some 'laziness hypothesis' about understanding, and generate deeper inferences only when the system is unable to solve, say a referential problem by more superficial methods. Of, in other terms, should an understanding system be problem-, or data-, driven.

---

* This is not meant to be just bland assertion. I have written at some length on the relations between application and the theoretical status of linguistic theories in (Wilks '74).

**The differences between Minsky's (1974) notion of 'default value' and what I have called 'preference' can be pointed up in terms of application. Minsky suggests 'gun' as the default value of the instrument of the action of shooting, but I would claim that, in an example like the earlier 'He shot her with a colt', we need to be able to see in the structure assigned whether or not what is offered as the apparent instrument is in fact an instrument and whether it is the default or not. In other words, we need sufficient structure of application to see not only that 'shooting' prefers an instrument that is a gun, but also why it will choose the sense of 'colt' that is a gun rather than the one which is a horse.

Although Schank sometimes writes of a system making 'all possible' inferences as it proceeds through a text, this is not in fact the heart of the dispute, since no one would want to defend any strong definition of the term 'all possible inferences'. Charniak's argument is that, unless certain forward inferences are made during an analysis of, say, a story —— forward inferences, that is, that are not problem-driven; not made in response to any particular problem of analysis then known to the system —— then, as a matter of empirical fact, the system will not in general be able to solve ambiguity or reference problems that arise later, because it will never in fact be possible to locate (while looking backwards at the text, as it were) the points where those forward inferences ought to have been made. This is, in very crude summary, Charniak's case against a purely problem-driven inferencer in a natural language understander.

A difficulty with this argument is the location of an example of text that confirms the point in a non-contentious manner. Charniak has found an excerpt from a book describing the life of apes in which it is indeed hard to locate the reference of a particular pronoun in a given passage. Charniak's case is that it is only possible to do so if one has made certain (non-problem occasioned) inferences earlier in the story. But a number of readers find it quite hard to refer that particular pronoun anyway, which might suggest that the text was simply badly written.

Another difficulty is that it is not always clear whether the argument is about what people are thought to do when they understand, or about how one should construct an understanding system. This is a difficult matter about which to be precise: it would be possible, for example, to agree with Charniak's argument and still construct a purely problem-driven inferencer on the ground that, at the moment, this is the only way one can cope with the vast majority of inferences for understanding, since any system of inferences made in response to no particular problem in the text is too hard to control in practice. Indeed, it is noticeable that the most recent papers of Schank (1974a and 1974b) and Charniak (1974) have been considerably less forward-inference oriented than earlier ones.

This dispute is perhaps only one of degree, and about the possibility of defining a degree of forward inference that aids the solution of later semantic problems without going into unnecessary depth. This might be an area where psychological investigations would be of enormous help to workers in A.I.

The justification of systems

Finally, one might usefully, though briefly, contrast the different modes of justification implicitly appealed to by the systems described earlier in this paper. These seem to me to reduce to four:

(i) In terms of the power of the inferential system employed. This form of justification has underlain the early predicate calculus-based language programs, and is behind Hayes' (1974) recent demand that any formalism for natural language analysis should admit of a set theoretic semantics, in the Tarskian sense, so as to gain 'intellectual respectability', as he puts it. The same general type of justification is appealed to in some degree by systems with PLANNER-type formalisms.

(ii) In terms of the provision and formalisation, in any terms including English, of the sorts of knowledge required to understand areas of discourse.

(iii) In terms of the actual performance of a system, implemented on a computer, at a task agreed to demonstrate understanding.

(iv) In terms of the linguistic and or psychological plausibility of the proffered system of representation.

Oversimplifying considerably, one might say that Charniak's system appeals mostly to (ii) and somewhat to (i) and (iv); Winograd's to (iii) and somewhat to the other three categories; Colby's (as regards its natural language, rather than psychiatric, aspects) appeals almost entirely to (iii); Simmons largely to (iv), and Sthank's and my own to differeng mixtures of (ii), (iii) and (iv).

In the end, of course, only (iii) counts for empiricists, but there is considerable difficulty in getting all parties to agree to the terms of a test.* A cynic might say that, in the end, all these systems analyse the sentences that they analyse or, to put the same point a little more theoretically, there is a sense in which systems, those described here and those elsewhere, each define a natural language, namely the one to which it applies. The difficult question is the extent to which those many and small natural languages resemble English.

7. Conclusion

The last section stressed areas of current disagreement, but there would, if votes were taken, be considerable agreement among A.I. workers on natural language about where the large problems of the immediate future are: the need for a good memory model has been stressed by Schank (1974a), and many would add the need for an extended procedural theory of texts, rather than of individual example sentences, and for a more sophisticated theory of reasons, causes, and motives for use in a theory of understanding. Many might also be persuaded to agree on the need to steer between the Scylla of trivial first generation implementations and the Charybdis of utterly fantastic ones. By the latter, I mean projects that have been seriously discussed, but never implemented for obvious reasons, that would, say, enable a dialogue program to discuss whether or not a participant in a given story 'felt guilty', and if so why.

The last disease has sometimes had as a major symptom an extensive use of the word 'pragmatics' (though this can also indicate quite benign conditions in other cases), along with the implicit claim that 'semantics has been solved, so we should get on with the pragmatics'. It still needs repeating that there is no sense whatever in which the semantics of natural language has been solved. It is still the enormous barrier it has always been, even if a few dents in its surface are beginning to appear here and there. Even if we stick to the simplest examples, that present no diffi-

---

* Though an interesting, and potentially revolutionary, distinction seems to have been introduced by a recent reviewer of many of the systems discussed here, between the functioning of a program and a 'program in itself': 'Only Winograd describes a program that is sufficiently impressive in itself to force us to take his ideas seriously. The techniques of the others have to get by on whatever intuitive appeal they can muster'. (Isard '74)

culty to the human reader —— and it must be admitted that it has been one of the persistent faults of the A.I. paradigm of language that it has spent too much time on puzzles examples —— there are still great difficulties both systematic and linguistic.

An example of the former would be the development of a dynamic system of understanding texts or stories that had any capacity to recover after having its expectations satisfied and then, subsequently, frustrated. At present no system of the sort described, whether of demons, pereferences or whatever, has any such capacity to recover. The situation is quite different from that in a dialogue, as in Winograd's system, where, on being given each new piece of information, the system checks it against what it knows, to see if it is being contradicted, and then behaves in an appropriately puzzled way if it is. In frame or 'expectation' systems it is all too easy to construct apparently trick, but basically plausible, examples that satisfy what was being looked for and then overturn it. That possibility is already built into the notation of frame or expectation. An example of Phil Hayes against my own system will serve: consider "The hunter licked his gun all over, and the stock tasted especially good" What is meant by 'stock' is clearly the stock piece of the gun, but any preference system like mine that considers the two senses of 'stock', and sees that an edible, soup, sense of 'stock' is the preferred object of the action 'taste , will infallibly opt for the wrong sense. Any frame or expectation system is prone to the same general kind of counter-example.

In particular cases like this it is easy to suggest what might be done: here we might suggest a preference attached to the formula for any; thing that was essentially part of another thing (stock = 'part of gun' in this case), so that a local search was made whenever the 'part-of' entity was mentioned, and the satisfaction of <u>that</u> search would always be the overriding preference. But that is not the same as a general solution to the problem, which used to be called that of 'topic' in the computational semantics of the Fifties. There are no solutions to this problem available here and how, though some suggestions have been made by Abelson 91974) and McDermott (1974).

A closely related, but equally intractable, problem is that of how to combine <u>highly specific</u> factual information within a general semantic structuring. Systems like Charniak's are, as we saw, concerned with specific rather than conceptual information, but there are quite simple 'semantic specificity' problems that one could not reasonably expect to be tackled even in a system devoted to the handling of facts, as can be seen by contrasting the sentences:

The deer came out of the wood.

The grub came out of the wood.

where we might safely assume that readers would assign quite different senses to 'wood' in the two cases simply on the basis of the two different agents. No-one, to my knowledge, has suggested any general method for tackling such elementary examples.

But, to finish on the bright side, it is important to stress that there is indeed an A.I. paradigm of language understanding in existence, one that embraces both what Winograd called first and second generation approaches, and which goes back, I suggested, to a considerable amount of earlier work in computational linguistics. It can be distinguished by a catalogue of neglect by conventional linguistics that can be summarised under three heads:

(i) theories of language must have procedural <u>application</u> to the subject matter that could in principle result in computer application and subsequent empirical test.

(ii) theories of language must deal with it in a communicative context, one amenable to empirical assessment. Merely sorting, as generative theories were designed to do, is not enough.

(iii) theories of language must also be, in a clear sense, theories of the formalisation and organization of knowledge. If they are not then we can know in advance that they can never tackle the problem of language understanding.

This paper has perhaps overemphasised areas of disagreement and dispute, and undoubtedly many of these disputes will come to be seen as

---

* One of the very few acknowledgements of this fact, of the possibility of an A.I. paradigm of language, from a linguist is (Fillmore '74).

no more than methodological questions, or disputes about matters of emphasis and degree. But some are, I feel sure, questions of substance, and it should be possible to see in the reasonably near future whether one approach on any given question is right and another wrong. It would be nice if this are to be settled by computation rather than by another change of fashion.

8. References

In order to compress the reference list the following abbreviations for collections of articles will be used.

AIJC3 Advanced papers of the Third International Joint Conference on Artificial Intelligence, Stanford, Calif. 1973.

AISB Proceedings of the Summer Conference of the Society for Artificial Intelligence and Simulation of Behaviour, University of Sussex, 1974.

TEDD Proceedings of the First International Conference on Machine Translation, National Physical Laboratory, Teddington, 1961 (HMSO, London, 1961).

ACL Proceedings of the Conference of the Association for Computational Linguistics, Amherst, Mass. 1974.

CAST Memoranda from the Instituto per gli studi Semantici e Cognitivi, Castagnola, Switzerland.

MITAI Memoranda from the Artificial Intelligence Laboratory, Massachussetts Institute of Technology.

SUAIM Memoranda from the Artificial Intelligence Laboratory, Stanford University, Stanford, Calif.

SRITN Technical notes from the Stanford Research Institute, Menlo Park, Calif.

MOD Papers in Computer Models of Thought and Language, ed. by Schank and Colby, (Freeman, San Francisco, 1973).

(Abelson '73)
Abelson, R.P., The Structure of Belief Systems, in MOD.

(Abelson '75)
Abelson, R.P., Discourse Headers, in Bobrow and Collins (eds.), Representation and Understanding (Academic Press, N.Y.).

(Balzer et al '74)
Balzer, R. et al, Domain-Independent Automatic Programming, USC/Information Sciences Institute, Marina del Rey, Ca.

(Bobrow '68)
Bobrow, D.G., Natural Language Input for a Computer Problem-Solving System, in Minsky (ed.), Semantic Information Processing (MIT Press, Cambridge, Mass. 1968).

(Brown '74)
Brown, G., The Believer System, in ACL.

(Bruce '72)
Bruce, B., A Model for Temporal References and its Application in a Question Answering Program, in Artificial Intelligence 3.

(Bruce, Schmidt '74)
Bruce, B., Schmidt, C.F., Episode Understanding and Belief Guided Parsing, in ACL.

(Charniak '72)
Charniak, E., Towards a Model of Children's Story Comprehension, MITAI-TR 266.

(Charniak '73)
Charniak, E., Jack and Janet in Search of a Theory of Knowledge in AIJC3.

(Charniak '74)
Charniak, E., "He will make you take it back": A Study in the Pragmatics of Language, CAST No. 5.

(Chomsky '57)
Chomsky, N., Syntactic Structures, (Mouton, The Hague, 1957).

(Colby, Parkinson '74)
Colby, K.M., Parkinson, R.C., Pattern-Matching Rules for the Recognition of Natural Language Dialogue Expressions, in American Journal of Computational Linguistics; Vol. 1, No. 1.

(Colby, Hilf '73)
Colby, K.M., Hilf, F.D., Multidimensional Analysis in Evaluating a Simulation of Paranoid Thought, in SUAIM-194.

(Colby et al '71)
Colby, K.M., et al, Artificial Paranoia, Artificial Intelligence 2.

(Coles '72)
Coles, L.S., Techniques for Information Retrieval Usin g an Inferential Question-Answering System with Natural Language Input, SRITN. 1972.

(Davidson '63)
Davidson, D., Actions, Reasons and Causal Laws, Journal of Philosophy.

(Davies, Isard '72)
Davies, D.J.M. Isard, S.D., Utterances as Programs, presented at the 7th International Machine Intelligence Workshop, Edinburgh, June 1972.

(Enea, Colby '73)
Enea, H., Colby, K.M., Idiolectic Language-Analysis for Understanding Doctor-Patient Dialogues, AIJC3.

(Farradene '66)
Farradene, B., Research on Information Retrieval by Relational Indexing, Dept. of Information Sciences, City University, London.

(Fillmore '74)
Fillmore, C., Summer School lectures, University of Pisa, Italy.

(Goldman '74)
Goldman, N.M., Computer Generation of Natural Language from a Deep Concentual Base, CAST No. 2.

(Halliday '70)
Halliday, M.A.K., in Foundations of Language.

(Hayes '74)
Hayes, P.J. Some Problems and Non-Problems in Representation Theory, in AISB.

(Hays '64)
Hays, D., Dependency Theory: A Formalism and Some Observations, Language, Vol. 40.

(Heidorn '74)
Heidorn, G.E., English as a Very High Level Language for Simulation Programming, in Proc. Symposium on Very High Level Languages, Santa Monica, Ca. 1974.

(Hendrix et al '73)
Hendrix, G.G., et al, Language Processing via Canonical Verbs and Semantic Models, in AIJC3.

(Herskovits '73)
Herskovits, A., The generation of French from a semantic representation, SUAIM-2'2.

(Hewitt '69)
Hewitt, C., in Proc. Second International Joint Conference on Artificial Intelligence, (Bedford, Mass., Mitre Corp., 1969).

(Hockett '62)
Hockett, C.F., Grammar for the Hearer, in Vol. XII of the Symposia in Applied Mathematics, American Mathematical Society.

(Isard '74)
Isard, S., review of MOD, Science, Vol. 186.

(Johnson-Laird '74)
Johnson-Laird, P., Memory for words, Nature, Vol. 253.

(Joshi, Weischedel '73)
Joshi, A.K., Weischedel, R.M., Some Frills for the Modal Tic-Tac-Toe of Davies and Isard: Semantics of Predicate Complement Constructions, in AIJC3.

(Kuhn '70)
Kuhn, T.S., The Structure of Scientific Revolutions, (University of Chicago Press, Chicago, 1970 2nd edn. .

(Lakoff '71)
Lakoff, G., in Steinberg and Jakobovits (eds.), Semantics, (C.U.P., Cambridge, 1971).

(Lamb '66)
Lamb, S., in Outline of Stratificational Grammar, (Georgetown University Press, Washington, D.C.).

(Lighthill '73)
Lighthill, J., in Artificial Intelligence: a Paper Symposium, (Science Research Council, 1973).

(McCarthy, Hayes, '69)
McCarthy, J., Hayes, P.J., in Meltzer and Michie (eds.), Machine Intelligence 4, (Edinburgh University Press, Edinburgh).

(McCarthy, '74)
McCarthy, J., Review of (Lighthill '73), in Artificial Intelligence.

(McDermott '74)
McDermott, D.V., Assimilation of New Information by a Natural Language Understanding System, MITAI, Tr-291.

(Masterman '61)
Masterman, M., Samantic Message Detection for Machine Translation, in TEDD.

(Minsky '68)
Minsky, M.L., Editor's Introduction to Semantic Information Processing, (MIT Press, Cambridge, Mass.).

(Minsky '74)
Minsky, M., Frame systems, circulated MSS., from Massachusetts Institute of Technology, and in Winston (ed.) Visual Information Processing (1975).

(Moore, Newell '73)
Moore, J., Newell, A., How can Merlin Understand?, in L. Gregg (ed.), Knowledge and Condition, (L. Erlbaum Association, Potomac, Md., 1973).

(Parker-Rhodes.'61)
Parker-Rhodes, A.F., A New Model of Syntactic Description, in TEDD.

(Richens '61)
Richens, R.H., Tigris and Euphrates, in TEDD.

(Rieger '74)
Rieger, C.J., Conceptual Memory, Stanford University Ph.D. Thesis.

(Riesbeck '74)
Riesbeck, C.K., Computational Understanding: Analysis of Sentences and Context, CAST No. 4.

(Rulifson, et al '72)
Rulifson, J.A., et al QA4: A Procedural Calculus for Inductive Reasoning, SRITN '73.

(Sandewall '71)
Sandewall. E., Representing Natural Language Information in Predicate Calculus, in Machine Intelligence 6, Meltzer and Michie (eds.), (Edinburgh University Press, Edinburgh, 1971).

(Sandewall '72)
Sandewall, E., An Approach to the Frame Problem, and its Implementation, in Machine Intelligence 7, Meltzer and Michie (eds.), (Edinburgh University Press, Edinburgh, 1972).

(Sandewall '72)
Sandewall, E., PCF-2, A First-Order Calculus for Expressing Conceptual Information, Uppsala Univelsity, Department of Computer Science.

(Scnank '68)
Schank, R.C., Semantic Categories, Tracor Doc No. 68-551-U.

(Schank '73a)
Schank, R.C., The Fourteen Primitive Actions and Their Inferences, SUAIM-183.

(Schank '73b)
Schank, R.C., Identification of Conceptualizations Underlying Natural Language, in MOD.

(Schank '74a)
Schank, R.C., Causality and Reasoning, CAST No. 1.

(Schank '74b)
Schank, R.C., Understanding Paragraphs, CAST No. 6.

(Schank et al '70)
Schank, R.C., et al, SPINOZA II: Conceptual Case-Based Natural Language Analysis, SUAIM-109.

(Schank et al '72)
Schank, R.C., Primitive Concepts Underlying Verbs of Thought, SUAIM-163.

(Schank, Rieger '73)
Schank, R.C., Rieger, C.J., Inference and the Computer Understanding of Natural Language, SUAIM-197, and in Artificial Intelligence, 1975.

(Schank, et al '73)
Schank, R.C., et al, MARGIE: Memory, Analysis, Response Generation and Inference on English, in AIJC3, and in Journal of the A.C.M. 1975.

(Simmons '70a)
Simmons, R.F., Natural Language Question-Answering Systems: 1969, Comm. ACM, Vol. 13.

(Simmons '70b)
Simmons, R.F., Some Semantic Structures for Representing English Meanings, Tech. Report No. NL-1, University of Texas, Austin, 1970.

(Simmons '73)
Simmons, R.F., Semantic Networks: Their Computation and Use for Understanding English Sentences, in MOD.

(Simmons, et al '66)
Simmons, R.F., et al, An Approach Toward Answering English Questions from Text, in IFIPS Conf. Proc., FJCC Vol. 29, 1966.

(Simmons, Bruce '71)
Simmons, R.F., Bruce, B.C., Some Relations Between Predicate Calculus and Semantic Net Representations of Discourse, in Proc. Second International Conference on Artificial Intelligence, 1971.

(Simmons, Slocum '72)
Simmons, R.F., Slocum, J., Generating English Discourse from Semantic Networks, in Comm. ACM, Vol. 15.

(Sussman '74)
Sussman, G.J., The Virtuous Nature of Bugs, in AISB.

(Wilks '67)
Wilks, Y.A., Semantic Consistency in Text — an Experiment and some Suggestions, Systems Development Corp., Santa Monica, Ca., SP-2758.

(Wilks '72)
Wilks, Y.A., Grammar, Meaning and the Machine Analysis of Language, (Routledge & Kegan Paul Ltd., Boston and London, 1972).

(Wilks '73a)
Wilks, Y.A., Preference Semantics, SUAIM-206, and in Keenan (ed.) Formal Semantics of Natural Language, (Cambridge University Press.1975)

(Wilks '73b)
Wilks, Y.A., An Artificial Intelligence Approach to Machine Translation, in MOD.

(Wilks '74)
Wilks, Y.A., One Small Head — Models and Theories in Linguistics, in Foundations of Language, Vol. 11.

(Wilks '75)
Wilks, Y.A., A Preferential, Pattern Seeking, Semantics for Natural Language Inference, Artificial Intelligence.

(Weber '72)
Weber S., Semantic Categories of Nominals for Conceptual Dependency Analysis of Natural Language, SUAIM-172.

(Wettler '73)
Wettler, M., Zur Speicherung syntaktischer Merkmale im Langzeitgedächtnis, Bericht über den 28. Konress der Deutschen Gesellschaft f. Psychologie.

(Winograd '71)
Winograd, T., Procedures as a Representation for Data in a Computer Program for Understanding Natural Language, MIT Ph.D. Thesis.

(Winograd '72)
Winograd, T., Understanding Natural Language, (Academic Press, Inc., New York, 1972).

(Winograd '73)
Winograd, T., A Survey Lecture on Natural Language, Third International Joint Conference on Artificial Intelligence, Stanford, Ca. 1973.

(Winograd '74a)
Winograd, T., Frame Representation and the Declarative/Procedural Controversy, in Bobrow and Collins (eds.), Representation and Understanding (Academic Press, N.Y.)

(Winograd '74b)
Winograd, T., Five Lectures on Artificial Intelligence, SUAIM-246.

(Wittgenstein '53)
Wittgenstein, L., Philosophical Investigations, (Blackwell, Oxford, 1953).

(Woods '70)
Woods, W., Transition Network Grammars for Natural Language Analysis, Comm. ACM, Vol. 13.

(Woods '73)
Woods, W., Meaning and Machines in (ed. Zampolli), Proc. of the International Conference on Computational Linguistics, Pisa, 1973.

(Woods, et al '72)
Woods, W., et al, The Lunar Sciences Natural Language Information System, BBN Report 2378, 1972.

END