SEVEN THESES ON
ARTIFICIAL INTELLIGENCE
AND NATURAL LANGUAGE

Yorick Wilks

1975 No. 17

Istituto per gli Studi
Semantici e Cognitivi
6976 Castagnola
Switzerland
Preface

Many writers in the field modestly call their finely-honed pieces workpapers, but these really are. With the exception of chapters I and II, which appeared in earlier versions in the TINLAP (see references) position papers, all are first drafts.

The aim of the pieces is positive, to point out faults, and suggest possible improvements, in a language understanding system described elsewhere. But the chapters try not to presume upon prior acquaintance with that system, of preference semantics, which is most extensively described in a companion paper (ISSCO memo No. 16), co-authored with Margaret King.

Some readers may miss the positive thrust and see only disputatious arguments with, and comments on, the work of others, and especially that of Charniak and Schank. But that would be a mistake. It is true that I think dispute is helpful and stimulating in this field, but the fact that detailed disputes are possible is a sign, not of distance, but of an enormous area of shared assumptions and methods.

I am indebted here to conversations, some extensive, some brief, with Robert Abelson, Eugene Charniak, Ken Colby, Carl Hewitt, Maggie King, John Laski, Jacob Mey and Roger Schank, among others. The fatuities, as usual, are all my own.

Bibliographic note: Despite widespread misapprehension on distant shores, the newspaper referred to in connection with court cases really is called the The Times.
Seven theses

I  METHODOLOGY OF ARTIFICIAL INTELLIGENCE APPLIED TO
   NATURAL LANGUAGE UNDERSTANDING.......................... 3
   (AI is not a science and, in the field of language
   understanding, false consequences follow from
   assuming it is)

II  PRIMITIVES AND WORDS...........................................13
   (Words and semantic primitives are not ultimately of
   different types, although procedural benefits come
   from making the distinction)

III PROGRAMS AND TEXTS.............................................22
   (Plans stated in high level programming languages are
   most appropriately seen as texts, since they do not
   seem to be either real world procedures or procedures
   for handling natural language)

IV  EXPECTATION AND ATTENTION....................................29
   (Expectation, as a method of parsing natural language,
   is radically defective unless it has some general and
   systematic capacity for attending to what it is reading)

V  SMALLER CHUNKS OF KNOWLEDGE..................................46
   (Real world knowledge can be represented usefully at quite
   local levels in an understanding system, and moreover can
   function there as part of a general system of linguistic
   inference)

VI  WHAT SORT OF TAXONOMY OF CAUSATION DO WE NEED FOR LANGUAGE
   UNDERSTANDING....................................................51
   (The important distinction between causes and reasons, in
   the explanation of human behavior, should have procedural,
   and not only taxonomic, reflection in an understanding
   system)

VII FRAMES, SCRIPTS, STORIES AND FANTASIES.....................98
   (If the frames hypothesis is taken to be either the plot-line
   or do-it hypotheses (q.v.), then its application to
   natural language understanding has not yet been demonstrated)

REFERENCES..........................................................114
I Methodology of Artificial Intelligence applied to Natural Language Understanding

Are workers in AI and natural language a happy band of brothers marching with their various systems together towards the Promised Land (systems which in the view of many well disposed outsiders are only notational variants at bottom) or on the contrary are there serious methodological differences inherent in our various positions? I think there is in fact one central difference, and that it is a methodological reflection of a metaphysical difference about whether there is, or is not, a science of language. But it is not easy to tease this serious difference out from the skein of non-serious methodological discussions.

By "non-serious methodological etc." I mean such agreed points as that (i) it would be nicer to have an understanding system working with a vocabulary of Nk words rather than Mk, where N > M, and moreover, that the vocabularies should contain words of maximally different types: so that "house", "fish", "committee" and "testimonial" would be a better vocabulary than "house", "cottage", "palace" and "apartment block". And that, (ii) it would be nicer to have an understanding system that correctly understood N% of input sentences than one which understood M%. When I say non-serious here I do not mean unimportant, but only that nothing theoretical is in question, so that, for example, it could be only an arbitrary choice whether or not a system that understood correctly 95% of sentences from a 3000 word vocabulary was or was not better than one which understood 98% from a 1000 word vocabulary.

Indeed, the very sizes of the vocabularies and success rates in the example show that such a choice, however arbitrary, is not one we are likely to be called upon to make in the near future, so let us press a little deeper.

Consider the following three points, which I will name for ease of subsequent reference:
(1) **Theory and practice:** "Trying hard to make a system work is all very well, but it's too success oriented, what we need at the moment is more theoretical work".

(2) **AI and science:** "What we are after is the right set of rules, and expressions of real world knowledge, for understanding natural language: no approximate, 95%, solutions will do, just as they won't do in physics".

(3) **Where to start:** "Since difficult examples clearly require reasoning to be understood, we cannot even begin without such a theory because, without it, we could not know of even an apparently simple example that it did NOT require reasoning in order to be understood".

The above three positions are not intended to be a parody, and certainly not a parody of anyone in particular's views. I have not in fact heard all three from the same person, even though, in my view, they constitute a coherent position taken together: one which I believe to be not only wrong, and I will come to that, but also harmful. Let me deal with the sociology first, and in the form of a very crude historical generalisation.

It is clear that "natural language understanding" has come to occupy a less peripheral place in AI, and much of the credit for this must go to Winograd (1972). The position, expressed in (1), (2) and (3) above is in some ways a reaction to that, and in my view an excessive one. Behind the positions above lurks the suspicion that the success of Winograd's system was in part due to its oversimplifications and that we must now be wary, for a while at least, of applications, successful or otherwise: that we must, in short, emphasise how difficult it all is.

Now there is undoubtedly something in this, but it seems to me that the reaction may have the paradoxical effect of causing the study of natural language in AI to be given up altogether. In the last year or two a number of those who seemed to be concerned with the problems of
of natural language no longer seem to be so. There has been a subtle change: from the analysis of stories, or whatever, to the setting out of systems of plans which now seem to construct stories as they go along. It might then seem natural to move further: from the production of stories about tying one's shoe-laces, shopping in supermarkets, etc. to plans, for robots of course, that will actually shop in supermarkets, tie their own shoe-laces, play diplomacy or whatever. And then of course we are back where we started in AI: back to AI's old central interests, robots, problem-solving and the organization of plans.

All this would be a pity, not only because someone has, as always, to be left holding the baby of natural language analysis, but because it is too soon, and AI has not yet had the beneficial effect it is capable of having, and ought to have, on the study of natural language. There are at least four of these benefits: let me just remind you of them:

(i) emphasis on complex stored structures in a natural language understanding system; frames, if you like (Minsky 1974),

(ii) emphasis on the importance of real world, inductive knowledge, expressed in the structures of (i),

(iii) emphasis on the communicative function of sentences in context, i.e. the finding of the correct-in-context reading for a sentence, as opposed to the standard linguistic view, which is that the task is the finding of a range of possible readings, independent of context,

(iv) emphasis on the expression of rules, structures, and information within an operational/procedural/computational environment.

Conventional linguistics has still not appreciated the force of these points, which are of course commonplaces in AI.
Let me now turn to the position sketched out earlier under three headings, and set out some countervailing considerations. It should be made clear that in what follows I am making only methodological points about the assessment of systems in general. No attack on the content of anyone's system is intended.

First, to the theory and practice point. It seems to me worth emphasising again that there can be no other ultimate test of a system for understanding natural language than its success in doing some specific task, and that to pretend otherwise is to introduce enormous confusion. Considerations of logic or psychological plausibility may indeed be suggestive in the construction of AI language systems, but that is quite another matter from their ultimate accountability, which can only be whether or not they work. Suppose some system had all desirable logical properties, and had moreover been declared by every respected psychologist to be consistent with all known experiments on human reaction times and so on. Even so, none of this would matter a jot in its justification as a computational system for natural language.

In a similar vein, it seems to me highly misleading, to say the least, to describe the recent flowering of AI work on natural language inference, or whatever, as theoretical work. I would argue that it is on the contrary, as psychologists insist on reminding us, the expression in some more or less agreeable semi-formalism of intuitive, common-sense knowledge, revealed by introspection. I have set out in considerable detail (Wilks 1974) why such an activity can hardly be called "theoretical", in any strong sense, however worthwhile it may be. That it is worthwhile is not being questioned here. Nor could it be, since I am engaged in the same activity myself (Wilks 1975b). I am making a meta-, methodological, point that the activity does not become more valuable by being described in value-added terms. The worthwhileness, of course, is shown later by testing, not by the intuitive or aesthetic appeal of the knowledge represented or the formalism adopted.
Let me turn to position (2): *AI and Science*. It seems clear to me that our activity is an engineering, not a scientific, one and that attempts to draw analogies between science and AI work on language are not only overdignifying, as above, but are intellectually misleading. Conduct with me, if you will, the following Gedankenexperiment: suppose that tomorrow someone produces what appears to be the complete AI understanding system, including of course all the right inference rules to resolve all the pronoun references in English. We know in advance that many ingenious and industrious people would immediately sit down and think up examples of perfectly acceptable texts that were not covered by those rules. We know they would be able to do this just as surely as we know that if someone were to show us a boundary line to the universe and say "you cannot step over this", we would promptly do so.

Do not misunderstand my point here: it is not that I would consider the one who offered the rule system as refuted by such a counter example, particularly if the latter took time and ingenuity to construct. On the contrary, it is the counter-example methodology that is refuted, given that the proffered rules expressed large and interesting generalisations and covered a wide range of examples. For the simple methodology of refutation is the method of idealised science, where one awkward particle can overthrow a theory¹. In the study of language such a methodology is no more appropriate than it is to consider the definition of fish as something that swims and has fins as being "overthrown" by the discovery of a whale. Of course it is not, nor does the definition lose its power; we simply have special rules for whales.

The fact of the matter is surely that we cannot have a serious theory of natural language which requires that there be some boundary to the language, outside which utterances are too odd for consideration. Given sufficient context and explanation anything can be accommodated and understood: it is this basic human language competence that generative

¹ The bad influence may not come directly from science, but via "competence theory" in linguistics.
linguistics has systematically ignored and which an AI view of language should be able to deal with. We know in principle (see Wilks 1971) what it would be like to do so, even if no one has any concrete ideas about it at the moment$: it would be a system that could discover that some earlier inference it had made was inconsistent with what it found later in a text, and could return to try again to understand. And here, to be interesting, the backtracking would have to be more than simply the following of some branch of a parsing that had been ignored earlier: it would have to be something equivalent to postulating a new sense of a word, a new reference of a pronoun, or even a new rule of inference itself. It is surely these situations that the "AI paradigm of language understanding", and perhaps it alone, will be capable, in principle, of tackling, in the future, and it is these features of language, that require such manoeuvres, that show most clearly why the "100%-Scientific Rule"$$ picture does not fit language at all, and why time spent trying to make it fit may be a diversion of attention from really key areas like the heuristic of misunderstanding and contradiction.

Perhaps a moment's further dilation on the role of counter-examples is worthwhile here. Consider two counter-examples: one produced against the "expectation as basic mechanism of parsing" hypothesis of Riesbeck (Riesbeck 1974), and one against my own "preference as basic mechanism etc." (Wilks 1975c) hypothesis. Riesbeck considers sentences such as "John went hunting and shot a buck", where, putting it simply, the concept of

$: Winograd's thesis, of course, had a system for checking inferences and new information against all that it knew already, though it is not clear that such a direct method would extend to a wider world of texts. In (Wilks 1968) there was a very crude program for finding out that an assignment of sense, earlier in a text, had gone wrong, but it was almost certainly an inextensible method.

$$ The reader should be clear that "100%" here has no relation to the notion of approximation. What I am denying is that language performances form a closed set for explication in the way that the sentences of a logical language, or the data of an experiment, do.
hunting causes the system to expect more about hunting and so it resolves "buck" correctly as the animal and not the cash. One then immediately thinks of "John went hunting and lost fifty bucks".

Conversely, in my own system I make much of the preference of concepts for other concepts to play certain roles, so that for example in "John tasted the gin", "gin" will be resolved as the drink and not the trap, because of the preference of tasting for an edible or potable object like the liquid gin. Someone then, plausibly enough, comes up with "He licked the gun all over and the stock tasted good", where the preference on a small scale would get the wrong "soup" sense of "stock", and not the "gun part".

It should be clear that these counter-examples are to what appear to be, superficially, opposed theories of parsing. My point is that in neither case do the examples succeed in showing a theory useless, i.e. neither "preference is no good" nor "expectation is no good" follow from the production of the counter-examples. What is needed of course, and what in fact both parties are trying for, is some suitable mixture of the approaches. But, and here is the key point, there will not be any magic right mixture either. There can only be a combination that will itself go wrong with sufficiently ingenious examples. Only a recovery mechanism will save us, just as it saves people, who misunderstand all the time. There will never be, nor could there be, a RIGHT combination, in the way that $F = \frac{k m_1 m_2}{r^n}$ gives a right theory of gravitation when, and only when, $n = 2$.

One further aspect of the misapprehension that the construction of a "language understander" is a scientific task shows itself in the belief that such an understander would be predictive in the way that scientific theories are normally supposed to be. Supporters of this belief would argue that, if such an understander "models" human behaviour in any interesting sense then, given a story containing a pronoun, the understander should predict how a human understander will refer the pronoun when he reads the story, and hence predict which item in the story he will refer the pronoun to.
But, in this explanation, "prediction" has, in my view, taken on a very peculiar sense. After all, if the text is clearly written, we do not predict how people will refer its pronouns, because we already know. If we in fact understand the text, then predicting how people will refer its pronouns is as sensible a task as my predicting how men differ from women. In essence we already know, and prediction over such an area is to cause the word to lose its meaning of predicting what we don't know. The only exceptions would be where either (a) we were dealing with mentally incapacitated people who could not understand language properly and who might get perfectly normal pronoun reference wrong, or (b) if we were dealing with texts which were badly written and which referred pronouns unclearly, so that normal human understanders might properly differ about what was meant by such stories. But neither of these cases would support any general belief that a language understander is, in any normal sense of the word, predictive of human behaviour because it aspires to simulate human behaviour. Simulating and predicting are just not connected.

Finally, let me turn to the third aspect of the initial position, which I called where to start. This brings up the very difficult question about the relation of reasoning to natural language, and I have made some remarks on that in the paper in section 2 on "Primitives". Here I just want to try and counter, in a brief and inadequate manner, what I see as the bad effects of the where to start view.

The view is an alternative to a more simple-minded view which goes as follows: "we should now concentrate on difficult examples, requiring reasoning, when studying natural language understanding, because the basic semantics and syntax have been done, and we are therefore right to focus on the remainder". This view is simply historically false about what has been done, so let us leave that and turn to the much subtler where to start view which holds that, on the contrary, the basic semantics of natural language understanding have not been done and cannot even be started without a full theory of reasoning capable of tackling the most difficult examples, because, without such a theory, we can't know that it
isn't needed, even in the apparently simplest cases. The argument is like that against the employment of paramedical staff as a front line in community medicine: we cannot have a half-trained doctor treating even influenza, because unless he's fully trained he can't be sure it isn't pneumonia.

One obvious trouble with the argument, in both its linguistic and medical forms, is its openness to reductio ad absurdum replies. It follows from that position, if taken seriously as a theory of human understanding, that no one understands anything until they are capable at least of understanding everything. So, for example, a child could never properly be said to understand anything at all, nor perhaps could the overwhelming majority of the human race. There is clearly something untrue to our experience and common-sense there.

I am not treating this position with the seriousness it deserves in the space available here. In a weaker form, it might draw universal agreement. If, for example, it were put in the weaker form that it was not really worth starting machine translation in the way they did in the 1950's, because they knew they had no semantic mechanisms, and so without some ability to go further, it was not even worth starting there. In that weaker form the argument looks far more plausible.

What I am questioning here is its stronger form: and again the reply is the same, namely that the position is another version of the 100%-rule fallacy: that in science you have to have a complete theory to have any worthwhile theory at all. This is untrue to language and diverts our attention from application and from an extensible system that could misunderstand and recover.

Let me summarise this chapter: it is an attack on what I have called the 100%-rule fallacy, alias the use of scientific methodology and assessment in work on AI and natural language. In my view this position has four unfortunate aspects:
1) It requires holding, usually implicitly, the false metaphysical position that there is some boundary to natural language over which one cannot step.

2) It has a false view of the role of counter-examples as rejectors.

3) It encourages talk of theoretical advance in a non-theoretical area, and downgrades the engineering aspects of AI, and thus the notions of test and application, which are the only criteria of assessment we have or could have.

4) It distracts attention from the heuristics of misunderstanding which should be the key to further advance.
II Primitives and Words

We may usefully distinguish between internal and external questions when discussing the use of primitives for representing natural language content and doing related semantic computations. Here I shall give a few examples of internal questions; go on to explain why I shall turn immediately to external questions; and finally discuss two of the latter: the justification of primitives in general, and the distinction, if any, between primitives and words.

What I mean by "internal questions" about primitives are detailed considerations about what semantic primitives to choose, or how to insert them into larger structures in particular cases so as to represent some complex concept or conceptual relation, etc. These are questions that can only arise when the general notion of semantic primitive has already been accepted. It is not possible to discuss such internal questions while one is at the same time answering external questions, such as the justification of semantic primitives in general.

There are fairly straightforward internal criteria for the selection and maintenance of a primitive vocabulary: the vocabulary should not be obviously redundant, with two primitives covering the same, or nearly the same, range of meaning. If one could show of anyone's suggested set of primitives that this was so, it would follow that he did not have a good set. Secondly, a primitive vocabulary should not be obviously oriented towards a particular subject area, if it is at the same time claimed to be a general set. So, for example, if, in a proffered set of primitive actions, we saw a majority of primitives concerned with human bodily actions such as moving, expelling, ingesting, etc. we might well wonder how such a system would cope with the expression of general actions such as "divide", "separate", "specify", "undertake", "delay", etc.

One issue bridges the gap between internal and external questions in an interesting way, and will serve me here as a new point of departure. A point of difference between Schank's views and my own has always been over the appearance in semantic representations of what appeared to be
simply the surface words of the language. So, for example, in a representation of "John shot Mary" by Schank (1973b) there will normally appear the English word "gun". Whereas, in my representations, all one sees are structurings of the 30 primitives, and "gun" is not one of them. I have suggested that Schank's diagrams are therefore of mixed type, as between primitives and words. However, I now believe this criticism of mine to have been badly put, and to rest upon a confusion, namely that there is some clear distinction between primitives and non-primitives (i.e. surface vocabulary). Historically, what happened in Schank's system is clear: he wanted a surface-free semantic representation without words in it, and has gradually achieved this, most recently by the elimination of many noun-words in favour of Fodor and Katz type primitive lists, as developed by Weber. I am proposing now in my own system to reverse the process, as it were, and starting from an inventory of only primitives (about 54 in 1967, now grown to 80, see Wilks (1972)) to begin to insert non-primitives, i.e. what appear to be words, into the formula structures that define word senses, but with one important proviso that I shall now explain.

Let us consider the English action "fire at", as a single unit, and its expression in terms of more primitive concepts. I have expressed it until now (1975a) as the formula, or tree structure, below:

```
(*HUM SUBJ) (*ANI OBJE) (STRIK GOAL) (THING INST) (CAUSE)
   /
  /
(*THING MOVE)
```
The structure need not detain us except to note that the rightmost primitive CAUSE is the head, or principal primitive, of the action, and the rightmost THING that is caused-to-move is, of course, the bullet, while the leftmost THING, that is the INSTRUMENT, is the gun.

What I now propose to do is to insert English words into the formula: in particular the words "gun" and "bullet", as follows (in list form):

((HUM SUBJ)((ANI OBJE)((STRIK GOAL)((gun INST)((bullet MOVE)CAUSE))))

but to do this if and only if there are also formulas elsewhere in the system defining the meanings of "gun" and "bullet". Thus, in the decomposition of a formula, two kinds of entities may be encountered, primitives and "words", and if an atom is found that is not a primitive then it must be a word and so a formula must exist elsewhere for it.

Now, there is a danger of circularity in all this: in that, looking for the formula defining "gun" as that atom appeared in the formula for "fire at", we might find a formula that told us (in primitives) that a gun was an object used by human beings for firing at something. True, but since the head (rightmost element) of the formula for "gun" will be THING, this re-entrant method of formula construction cannot give less information than formulas co-sisting only of primitives, as mine have until now.

Now, does this method of re-entrant formulas — re-entrant in the sense of mentioning formulas inside each other (as that for "fire at" now mentions that for "gun" and "bullet") — force me to withdraw the "mixed-type" criticism I made in the past of Schank's descriptions.

Well, yes and no. No, in that Schank's use of, say, the word "gun" in the conceptual dependency representation for "shoot" was not re-entrant in the sense I have defined. There was, if I understand him, no formula elsewhere for "gun" in his example: there was the English lexeme and no more. Hence, if a program parsing his system searched in a text for "gun" as the instrument of "shoot", and found "bow and arrow", it would be helpless because it did not find exactly what it was looking for — it would have no descriptive formula for the sense of "gun" to help it know
when it had found roughly the same sort of thing. Just as in "John shot her with a colt", it would have no information with which to separate the horse and gun senses of "colt": it would either find the sought for English lexeme or, as in these cases, not.

But then again, on the other hand, the criticism is changed because my claim that there should not in principle be mixed type (word and primitive) semantic descriptions is implicitly withdrawn by the above proposal for re-entrant formulas. Now there need be no serious "theoretical" considerations involved in such a proposal. It can be seen as simply a notational convenience: in that "gun" in a formula for "shoot" is now just a shorthand form for the formula for "gun" existing elsewhere in the system. This makes the formulas easier to read for a human user by avoiding the insertion of too much repetitive material in terms of primitives into the body of the formula itself.

Yet, there is a more fundamental point buried in all this, very much an external question, and one which led me to mis-state my criticism of mixed type descriptions in the past. Let me now go into this a little: though it is not, as I shall show, intended to lead to conclusions that should give any comfort, either to those who have used mixed type descriptions in the past or to those who wish to avoid primitive descriptions altogether.

My basic error was to defend the use of primitives by implicitly assuming that they were in some way essentially different from natural language words and, therefore, that structures of primitives were in some way different from structures of natural language words, which is to say, assuming that a semantic description in terms of primitives was something other than a reduced micro-language, with all the normal weaknesses and vaguenesses of a natural language. I now see more clearly that that is not and cannot be so. Or, to put the message of this position paper in a slogan form, there is no escape from a natural language and certainly not via the primitive route.
Let me take first the justification aspect of this. Those of us who have used primitives for some time have all I think made use of the innocent notational convention of writing primitives in upper case letters thus: MAN, CAUSE, TRANS, etc. And I would not wish to change this; it is useful for indicating whether we are, at any given moment, describing a structure in the primitive or the surface language. The trouble is that such a usage does inevitably carry the suggestion that the upper case entities are something other than (over and above, or "deeper" than, but anyway different from) the ordinary words they look like. And this, of course, is error, pure and simple. There have certainly been claims from some quarters that these entities are quite other: that CAUSE has almost a halo round it, -CAUSE- and represents, or refers to, some other kind of entity in the brain or mind directly.

At its crudest, this is just the old referentialist fallacy moved up a level, as it were. The low level fallacy, still alive and well, is that the meanings of words are physical objects. "Chair", the story goes, signifies by referring to things like the one I am sitting on, and so therefore do "mind", "action", "friendship" and "cunning", though perhaps in a slightly more roundabout way. I do not want to discuss this view here, but only to point out that the view of primitives under discussion holds that primitives, like words, have their meaning/significance in the same sort of way, but in their cases by referring to certain ill-defined mental entities. I have argued in detail elsewhere (Wilks 1974) that this could not conceivably be so, or be known even if it were so.

But this comic-strip philosophy is not the heart of the matter: what this view of primitives does, in real terms, is to lead to research that attempts to justify particular sets of primitives directly in some way, and these days that usually means psychologically. But what I am

$\footnote{I cannot see that the notion of the "psychological justification of primitives" makes sense, though I would be happy to be shown. I do not of course refer here to work like (Johnson-Laird 1974) showing that humans seem not to store surface language. Such results are quite consistent with a "primitive hypothesis", but do not support it, since they are also consistent with the hypothesis that human semantic representation is not linguistic at all, but consist of, say, binary numbers!}$
arguing is that, if formulas, templates, conceptual dependency structures, etc. are simply usages of a language of primitives, then no direct justification of the vocabulary makes sense, and certainly not any justifications of a correct vocabulary, any more than it would make sense to try to establish the correct vocabulary of English or any other natural language. It follows from this that there can be a variety of primitive languages for semantic descriptions, no one necessarily better or worse than any other, any more than my vocabulary is better or worse than yours if I know 100 English words you don't, and you know 101 that I don't. In the case of each primitive vocabulary, the only ultimate test will be the success or failure of linguistic computations that make use of it.

Now there are limits to this sweeping parallel between primitive and natural languages. Clearly, and as I mentioned earlier, a primitive vocabulary should not have synonymous primitives, whereas in a natural language it is often a point of aesthetic pride to have as large a range of semi-synonyms as possible: "cemetery" and "graveyard" in English, for example. Again, there is at least one interesting piece of indirect justification of a primitive vocabulary, namely the project at Systems Development Corporation that put the whole of Webster's Third International Dictionary onto tape and counted the rank frequency list of words used in the definition of other words. That was, ignoring very frequent words like "a" and "the", down to the 80th in rank order, pretty close to my own list of primitives, and naturally I was pleased. That is what I, like anyone else, would have hoped for; since primitives are in fact used to define the senses of other words, we should be happy to be close to the list used unconsciously by the makers of a large and efficient dictionary.

$ Another and quite different demand for justification of primitives, in terms of a denotational or set-theoretic semantics, has been made by Hayes (1974) and, implicitly, by the whole movement in logic and linguistics associated with the name of Montague. Its view of semantic primitives as advocated here is expressed in David Lewis' term "markerese". I believe their arguments to be radically mistaken, and that this is capable of clear demonstration (see Potts 1975, and Wilks 1975c).
Nevertheless, and in spite of these two caveats, it seems to me that a primitive vocabulary is nothing other than a small natural language, and therefore not open to methods of justification unavailable to any other natural language. This brings to mind a mistake to pretend otherwise. To sum up, if words of a natural language like English are not justified rox gain significance by their direct reference to things, but only by their function within the overall language, then we may expect precisely the same to be true of a language of primitives, and no amount of writing in upper case letters is going to change that fact.

Now, if I take this to mean that there is no difference at bottom between primitives and other words, then I find that another criticism I have made implicitly or explicitly in the past must also be completely recast. This concerns the initial representation of sentences in predicate calculus and PLANNER-type formalisms, where it seems unnatural to represent "John is at the station" as (AN JOHN STATION).

I have been tempted to criticise such representations, to myself at least, on the grounds that they were simply the English words of the sentence (or something very like these) rearranged in some plausible way, and that therefore nothing had been shown or structured. But, as I said, if I take to heart the point that primitive vocabularies are not essentially different from those containing more obviously surface words, then I cannot maintain that criticism in the same way. Why shouldn't "John", "at" and "station" be somebody's primitive vocabulary?

Now the structural criticism of this method of coding up sentences is unaffected. That is to say: (1) Unless it relies on a method of expression, unless done in a very unsuitable manner, and it normally relies over much on our intuition of the structure of the original sentence, in such a way as to be not really the structuring of an example but a mere representation of the example itself. To see this, one only has to think of that in which the line beginning "Now the structural criticism...".
However, the point at least here is not this structural one but that of the status of the items in the description: as I raised the question, but did not answer it, why should those English words not be declared to be part of the primitive vocabulary? For, if as I have just argued no serious distinction of type can be maintained between words and other primitives, what could be wrong with that?

Well, it is easy to see what is wrong, given that one accepts one other principle: namely, that one's system, whatever it is, should be extensible in a non-trivial manner. What we are now discussing is the fallacy of the map, to adapt a philosophical cliché. We have a system claiming to represent the structure of natural language but which in fact represents it in the way a map would if its scale was one-mile-to-one-mile. There would be something wrong with such a map, that much is clear, and similarly there is something wrong with a system which is only extensible on the same scale as what it represents: it adds to its primitives (ordinary words in this case) at the same rate as it adds to the sentences covered. I am not trying to smuggle back any of the distinction between primitives and non-primitive words that I have abjured, but am only pointing out that, if one chooses surface words as one's primitives, there is nothing theoretically wrong, but just the practical (and in my view insuperable) difficulties of (1) inability to state significant semantic generalizations, and (2) the inability to extend one's coverage of the language in anything other than a mile-to-mile manner. It is for this reason that it still seems possible, to me, to give up believing in any difference in principle between primitives and other words but still advocate strongly the use of a sensible selection of words as a reduced, or primitive, sublanguage for semantic expression.

§ One way out I have not discussed here is for someone to argue that, for the representation of language, the mile-to-mile scale is all right, because, while one mile is not enough, two are.

One final point, which is not argument but merely the drawing of a bead on a more distant target. It is the case that the use of a Planner, or Predicate Calculus, type of representation in "quasi-English" that I have just discussed is intimately associated with the view that the study of reasoning is dissociable, or decouplable, from the study of the semantic structure of natural language, and can be pursued in isolation. In (Wilks 1975b) I have raised a number of doubts about that view, but here I want to add another which follows directly from the argument of this paper.

If it is true, as I have argued here, that there is no escape from a natural language into some other realm, and that a language of representation is just another natural language whether of primitives or of "quasi-English", then it follows that there is no special extra-terrestrial sphere for the examination of reasoning, but only translations into another natural language. Hence there is no reasoning about natural language separate from natural language $^5$, and all we can do is to choose the language in which we prefer to model the reasoning and over which we prefer to compute. Thus, to compress things somewhat, we have the choice between computing about reasoning in a primitive-like language, or one reducible to it by the "re-entrant" method I described, or in one like Planner quasi-English in which little is made explicit, and which would require another system to make its internal relationships explicit for any but the most trivial examples.

It may turn out that it is more sensible to say that language understanding depends on reasoning, rather than vice-versa. Everyone in AI seems to believe it without question, and I have done no more here than raise a few small doubts that it might, after all, turn out to be the other way round.

$^5$ I am not denying here that some non-linguistic forms may explicate our reasoning about, say, position in space or numerical relationships; nor am I denying that there is reasoning in the sphere of vision,(which is also possessed by dogs) and cannot therefore be linguistic. Granted all that, I am arguing that the sort of reasoning required to understand the argument of this paper, or of a standard newspaper editorial, must (a) be linguistic in nature, but (b) will not be explicated in an interesting and non-circular manner by a system using only the quasi-English of a programming language.
III Programs and Texts

Those who read papers on natural language understanding and on advanced programming languages like PLANNER must have noticed that they are becoming more alike, and not merely because discussion of Birthday parties now seems to play a prominent role in both. There are also, of course, an increasing number of papers with suggestive titles like "Programming in Natural English" and "English as a high-level programming language". All of which leads one to entertain the possibility that progress in understanding natural language may come, not from the many toilers in the vineyard of natural language itself, but from those who construct sophisticated programming languages progressively more and more like the real thing.

I believe that there may turn out, in the long run, to be a lot in that hope, but that in the short run there are causes for disquiet, as I tried to hint at the end of the last chapter. Let me indicate the source of that disquiet by starting from very simple considerations.

Many writers in the field of AI and natural language (NL) make use of the following convention in one form or another:

[John is in love with his teacher]

or

John-is-in-love-with-his-teacher

or

(IN-LOVE-WITH JOHN His-TEACHER)

is taken to be the name of the internal representation, semantic representation, or whatever, of the English sentence, "John is in love with his teacher".

A harmless enough convention, you may feel, and one almost impossible to avoid, since in actual linguistic systems the representation of that sentence, in semantic primitives, might be a very complicated object indeed, filling perhaps most of a page, and it would be very awkward not to have some shorthand name for it.
What then is the problem? Put simply, it is that much of the
time, in papers and discussion, the symbols above are used NOT as the
name of the representation (which is what they are claimed to be) but
as no more than alternative forms of the name of the English sentence
that they contain. Thus, what is being done in those exercises is in
danger of being more trivial than is generally realised. It could easily
become no more than degenerate linguistics or philosophy of language,
and of a rather amateur sort.

Let me try and pinpoint the trouble more precisely, and then
suggest what might constitute a remedy; one that may be no more than
expositional, since it would be a real advance if those who do what I've
described were aware of what they are doing, and no more.

The following is from (Hewitt 1975):

[((Marvin-attends-birthday-party-for-Seymour =
(ask (Eva which = present is-a-suitable-present-for Seymour
from Marvin)

(then:
(Marvin-gets-money (sufficient-to-buy-present)

(then:
(Marvin-buys present

(then:
(Marvin-goes-to (home-of Seymour)
at (time-of
(birthday-party-of Seymour)))))))]

It should be said right away that Hewitt does not claim to be analysing
NL. He is concerned with the "progressive refinement of plans", and the
quotation above is only one of several stages of the "Birthday Party Plan"
that he presents.

But what is the plan for? Since this question has to come up, it
means that we cannot totally avoid the well-known slogan "meanings are
procedures". Let us state three assertions that have some affinity to
this slogan:
1. The meaning of a program is essentially a symbol mapping procedure.

2. The meaning of some words in NL, like "unscrew", cannot be divorced from real-world procedures. (i.e. "unscrew" is hard to explain without showing how.)

3. The meanings of all NL words are ultimately procedural, however much common sense may deny this.

Point 1 is now a commonplace, as is 2. 3 is probably false, but there is a long history of belief in it. Marx, for example, probably held some form of it.

How do 1-2-3 help? Well if the quotation above is a procedure, what sort of procedure could it be? It must be the text of a procedure for a human, and is thus closest to 2. It cannot fall under 1, and be a procedure/program for a machine, except in the highly unlikely event of someone wanting to program a robot called Marvin to attend, and astonish, at birthday parties.

Suppose now, to be absurd, that some ordinary language philosopher, call him OLP, were to analyse for us the meaning of "Marvin attends a birthday party for Seymour", and that he gave us the following as the result of his researches:

"Marvin attends a birthday party for Seymour"

means not just an appearance of a person at an event but presupposes in ordinary language a series of concomitant actions, such as:

Marvin asks, say, Eva, what would be a suitable present from himself for Seymour,

then

Marvin gets sufficient money to buy a present

then

Marvin buys the present

then

Marvin goes to Seymour's home at the time of Seymour's party.
It seems obvious that OLP would not have told us anything very profound by doing that, yet what exactly would be the difference between the two forms written out above, given that the first is a text and not a machine program. And no one at this point should confuse the issue by saying, truly but irrelevantly, that "programs are always texts".

Here are three possible explanations of the difference:

Explanation (i): The second, OLP, form has no formal objects in it, like =,\(\exists\), (, ), etc. and they make a crucial difference. But surely that is not so, and the formal objects in the first form of the explanation have no obvious formal role. Consider the line:

(Marvin-gets-money (sufficient-to-buy present))

What is its structure, and why is it different from putting the hyphens and brackets it contains ANYWHERE ELSE in the line? And it is beside the point to object here that the syntax interpreter would object unless the hyphens were placed just as they are above. What I am asking for is some intuitive (and non-formal) justification of why the line is not:

(Marvin-gets-money (sufficient to-buy-present)).

Explanation (ii): The first version (let's call it CH for short) is not English, like the OLP one, but is the name of a formal representation of a plan.

But, on that view where is the representation of which this is the name? This explanation does seem to be the one closest to Hewitt's views, and he might argue here that this apparent English is just for ease of communication with the reader, and that the name of the representation corresponding to the quote above could just as easily be CH 5943, and only perspicuity would be lost by doing that.

But even if it were called CH 5943, what is the representation in fact, and what does it do when it runs. For surely it can't attend parties?
Explanations (iii): The CH version is not English, like OLP's, but is the name of procedures for attaching, or parsing, the representation to the surface English, or whatever.

But if this is so then the procedures are the heart of the matter, and we should be shown one. They are after all what all workers on representing NL are after, for it is non-trivial even to distinguish the occurrences in English of "is a suitable present for", "is a present suitable for" and "is a present for suitable", among other possibilities.

There is no way out by saying again that GNSYM 5943 is the name of the procedure that does that, because it's the procedure we want and not another name for it. There would be a danger, if the GNSYM line were pursued, of falling into the fallacy that one could solve the problem of representing English, say, without any "rich" representations, but simply by having a name for every possible English sentence, waiting for it on input, and manipulating structures associated with the name. It may be hard to put one's finger on just what the fallacy is in that, but that it is a fallacy is pretty clear.

Since the three possible explanations above are not convincing, one might have to conclude by default that the forms CH and OLP are NOT essentially different. This does not mean they are nothing, or worthless, but the CH form is then not what Hewitt seems to think it is. This error is, I think, in part caused by the general style of thinking that LISP programming induces: namely, that one always begins in LISP programming by writing a function equivalent to SOLUTION, or DOWHATIWant. In other words the answer is named before it is computed. But naming, in discursive argument at least, is not the same as computing.

What is the upshot of all this? In Chapter II above, I argue that any system claiming to represent the content of serious natural language, must in fact employ representations which do not have properties essentially different from those of natural languages. If they do have such properties they cannot represent the content. Thus, no language with a Tarski type semantics could represent the content of NL, and we can know that a priori.
Thus the choice, in representing NL, and so of the sort of knowledge that can only be expressed in NL, is not between using an NL for the job or not, for that is not a choice: we have to use an NL. It is whether or not we can reduce to a manageable NL, as in some of the approaches kicking around in AI-and-NL, or whether we represent NL by itself and claim it can't be reduced.

It seems clear to me that the CH form, if in fact indistinguishable from the OLP form, is the latter of these, and that although that position cannot be shown theoretically false, is clearly repugnant to common sense, in exactly the same way as the false but irrefutable view that we speak by rolling out prestored complete sentences from a list in our heads of all the sentences we will ever say in our lifetime. There is on that view no difficulty about how we relate THIS IS A PRESENT SUITABLE FOR SEYMOUR to THIS IS A SUITABLE PRESENT FOR SEYMOUR, because we just don't. They are wholes, recognised individually, and with no more relation than any two random sentences on this page.

The situation I am diagnosing can be put in terms of a very old and unfashionable form of argument called the dialectic: that of two apparently contradictory claims, both of which seem true (the thesis and antithesis), from whose opposition we are saved by a third claim, the synthesis. (I produce this piece of logical history with my tongue in my cheek, but someone will no doubt find AI merit in it, as in any other random dredging of the Ideengeschichte):

**THESIS:** there are no specially privileged languages for the expression of knowledge and any code with the property of being an NL will do, though nothing without that property will do.

**ANTITHESIS:** the thesis above leads to circularity from a computational point of view because we must have some way of analysing the content of whatever NL we choose to express knowledge in. (This was my argument against the CH form — it's just English really and how can a program understand THAT without another program to understand English?).
SYNTHESIS: the only way out is via a form of representation that is admittedly a NL, but reduced from one like English, and with DEFINITE PROCEDURES for going from the first NL, say English, to the reduced NL, say a semantic representation.

Hence, in a system like my own, the convention[ John likes driving ] is used as a name of the representation of "John likes driving", so as to refer to it easily. But what[ ...... ] is the name of is available and constructible from the information in papers and programs, and would be meaningless if that were not so.

The last matter is, of course, what everyone working in semantic representation of NL are after. But without that work what is the significance of "quasi-English" indistinguishable from OLP's? It is not worthless, but maybe only amateur linguistics, or philosophy, which we all do badly. The essential contribution is parsing, or attachment procedures, that make the whole thing non-trivial and non-circular.

If we ask CH where are the procedures that make the first quote above more than the obvious, there seems to be no answer at the moment. That need not remain so, but something more is needed in exposition, and some of the questions raised here need to be answered. It is perhaps a paradox that those who promulgate slogans about the centrality of procedures may be unwilling to hand them over at certain key moments.

What I am really getting at of course is that reasoning — interesting reasoning that is, not reasoning about whether we are looking at a hippopotamus or an old sock — is IN natural language, and therefore we need to concentrate on programs that understand NL rather than on programs that assume NL is understood. The alternative view, I'm hinting, leads inevitably to circularity or disguised triviality.
IV Expectation and attention, or where to put case information in a semantic analyser

Introduction

In this chapter I want to describe the way a preference semantics system handles case information and parses case structures, and to do so in a way that clarifies various inadequacies in previous presentations. At the end I shall draw a number of comparisons with a more recent system with similar overall aims, due to Riesbeck.

I set the scene for case, not historically, but with points of contemporary reference, by reminding the reader of the basic views of Fillmore and Schank on case: Fillmore constructs the notion of a case frame for surface verbs, in which any case may appear optionally or obligatorily.

Thus, if the horizontal line denotes the spot where the verb is to be substituted, Fillmore's frame for "murder" is (Fillmore '68, p.28):

(1)  \[ \begin{array}{c} \text{D (I) A} \end{array} \]

which means that that verb can take the Dative case, the Instrumental case and the Agentive case and, moreover, the brackets round the "I" mean that Instrumental is optional, in that it is not always necessary to accompany every proper use of "murder" with an explicit statement of what the murdering was done with. Lack of brackets however, round "A" and "D", shows that they must always be explicit, and so those cases are obligatory in the frame for "murder". I shall not say anything about the content of these case labels here, but will assume that the reader is familiar with at least the basic statement of case grammar and what it is intended to achieve.

By way of contrast, the Fillmore frame for "kill" is:

(2)  \[ \begin{array}{c} \text{D (I A)} \end{array} \]

where the overlapping brackets mean that one of those two cases so joined must appear, though it does not matter which.
In so far as case frame can be applied to Schank's views, it applies to his basic actions, or primitive acts; and he has declared that all cases that a primitive act takes, it takes obligatorily.

Thus, for example, Schank's primitive act TRANS expresses the underlying content of such actions as "buy", "sell" and "take", and he would begin the representation:

(3) The man took a book

as (Schank '73, p.196)

(4) \[
\begin{array}{c}
\text{man} \\
\text{TRANS} \\
\text{book} \\
\text{to} \\
\text{man} \\
\text{from} \\
\text{someone}
\end{array}
\]

where the arrows labelled "R" and "O" indicate Recipient and Objective case respectively, and the Agentive case is in fact indicated by the double arrow linking "man", the agent, to the act TRANS. The details here need not concern us, the point being that Schank is setting up case frames, not for surface verbs of English, like Fillmore, but for these primitive acts, of which he has about twelve.

From the point of view on case expressed in the system described in this chapter, both these strong positions have drawbacks, indeed they have complementary ones. In the first place, Fillmore's system, with the aid of which he wants to contrast verbs by means of their frames, only becomes significant if interpreted with the aid of some non-surface representation of actions. So, for example, Charniak has pointed out (personal communication) that the above contrast of the frames of "kill" and "murder" is only significant if there is some common, underlying, action that the two verbs share, and which can be thought of as being substituted for the horizontal line. For, if that is not so, then the contrast of the frames for "kill" and "murder" is of no more interest than the contrast between "kill" and "remember", or any other random verb. In other words, it is only because the two verbs already have something in common, over and above their case frames, that the comparison has point. Thus, the contrast of the case frames of surface verbs is, if unsupplemented, unsatisfactory.
Conversely, there may be trouble with Schank's attempt to both (a) relate surface verbs to underlying primitive actions, and then discuss only the latter, and (b) at the same time make all participants in the case frames obligatory. So, for example, Fillmore would express the case frame for "see" as ( —— O + D) and for "learn" ( —— O + A) (the "+" need not concern us here). While Schank (pp.220-1 ibid.) expresses both verbs by an underlying primitive MTRANS together with a case frame, for the primitive, containing at least A, O and R (Fillmore would call R by D). The individual letters for cases assigned by different authors need not detain us, nor need their contrasting interpretation of the case names, for my point is the perfectly general one that, whether or not Fillmore is right with this particular verb pair, it is highly likely that there are pairs of surface verbs like this one whose

(a) surface case frames are different
(b) Schankian primitive act is the same.

It follows from (b) that for Schank their "deep" case frame must therefore be the same too. Since, for him, every surface verb has a main act expressing it, there is clearly going to be a problem for Schank with this consequence of (b) unless he is prepared to say that there is no necessary relation at all between a verb's case frame and the case frame of its corresponding primitive act. It may well be possible to defend such a position within his theory, but he will still be left with the difficulty that verbs with quite different semantic behaviour (for Fillmorian case differences are not superficial) have identical behaviour in his system. There is bound to be a lack of discrimination consequent upon Schank's all-cases-are-obligatory view unless some careful avoiding action is taken, that he has not yet, to my knowledge, embarked upon. The point for what follows is that there are also difficulties about defining case frames only for underlying, or basic, actions.
Case in formulas

The preference semantics system has for some time tried to have it both ways and to say something, though not necessarily the same thing, about the case behaviour of both surface and underlying actions. Let me remind the reader of the essential points about case in semantic formulas: formulas are structured entities of semantic elements, and they represent and discriminate word senses. Some semantic elements like CAUSE, MOVE, etc. are act-like, many are not, viz: GOOD, THING, THRU, INST. There are about eighty elements in use altogether. A formula is normally written as a tree interpreted from the right to the left: thus

(5)

\[
(*\text{HUM SUBJ} \quad (*\text{ANI OBJ} \quad \text{THING INST}) \quad \text{STRIK})
\]

is for "to beat", and is an action preferably of STRIKing, normally done with an INSTREAMent that is a THING preferably to an ANInimate OBJECT, and preferably by a HUMAN agent. The tree structure is largely superficial so as to get a particular LISP representation and the same semantic entity could be written, as Simmons writes his Fillmorian case dependencies, thus:

(6)

\[
\begin{array}{c}
*\text{HUM} \\
\text{(or Agent)}
\end{array} \quad \text{STRIK} \quad \begin{array}{c}
\text{OBJ} \\
\text{INST} \\
\text{THING}
\end{array} \quad *\text{ANI}
\]
This form (6) emphasises that the three case parts at the "top level" of the formula are in some fundamental relation to the basic action, or head of the formula (always its right-most member and underlined here). However there are limitations to this re-picturing as with all picturings. For example, it makes little sense for formula trees representing noun senses. Again, the case subparts can be extended at lower levels in the formula, in which case the Fillmorian form above loses its force. Thus the formula for "sew" might well contain as instrument the subpart:

(7)

\[
\begin{array}{c}
\text{INST} \\
\text{WITH} \\
\text{THRU} \\
\end{array}
\begin{array}{c}
\text{LINE} \\
\text{THING} \\
\end{array}
\]

Here the specification on the instrument that sewing is done with (the needle that is) itself contains an accompaniment case specification: that the linear object has an aperture ((THRU PART) WITH). It is not to the point here whether other authors would allow an accompaniment case at this point or not. Specification in this system can be carried down to whatever level is thought appropriate by the formula maker, consistent with the syntax of well-formed formulas.

Note in passing that the specification of slot-fillers for cases is more complex than in either Schank or Fillmore, both of whom, in their different ways, restrict themselves to simple Fodor and Katz style individual unstructured markers for such specifications. Note too that the formula maker is not bound by any view that there is something in common between all specifications of agents in all formulas (and which would therefore constitute the "universal content of the notion of Agentive". It may well be that Fillmore holds some version of that assumption).
The formula maker here would be concerned only with specifying, as best as he could, the semantic preferences of the particular surface word in question. If it were a verb he would semantically specify the preferred agent of that action only, with no implicit reference to agenthood as such. The same goes for the other cases: he would attempt to put those cases into a formula that he thought necessary to specify the meaning of the action. Consider

(8) He lives in Lugano

and

(9) He drank some Barbera in Lugano.

It is logically true that one must, if one drinks, do it somewhere, but no one would hold that the notion of location was involved in explaining the meaning of drinking. However, one might well hold that it was impossible to explain the notion of living, in the sense of inhabiting, without making clear that it was done in some location. Thus in the formula for only one of these actions we would expect a location case subpart (......SLOCA).

This description of the insertion of case specification into the formula for some given surface verb does not correspond in any clear way to Fillmore's obligatory or optional category, though it seems clear that any formula should contain at least Fillmor's obligatory cases for that surface verb.

Formulas are not in any sense the representation of pieces of text. Text representations called templates are built up out of formulas, which remain entire within templates, and their initial procedural function is to try and dictate how the templates they form part of shall be constructed; thus a formula for "beat", as above, tries to ensure that any template containing it has a human agent, which is to say, has a formula placed at its agent node whose head (rightmost element) is MAN. The formula remains in the template whether or not its preferences were satisfied in building up the template\$, and at later stages the system can make

\$ as they would not be in "The monkey beat its offspring".
inferences guided by preferences unsatisfied at an earlier stage. The important point here is that the formulas are generally used in parsing procedures only as far as setting up templates, where a template is a connectivity of formulas with at least an Agent formula, an Action formula and an Object formula $$. We will write templates in short form not as triples of complex formulas, but as square brackets round the phrase or clause the template is for; the words being clustered within the brackets to represent the three nodes of the template.

Formulas are not used as parsing mechanisms to deal with superficial expression of case in English by means of (highly ambiguous) prepositions. This is done by other structures called paraplates, to which I will now turn, partly to improve on the account I have given of them in the past, particularly in (Wilks 1975).

**Case in paraplates**

Paraplates assert a connectivity between two templates, typically between a template representing a main clause and one representing a prepositional phrase. Thus, if we were representing

(10) John left his clothes at the cleaners

which would be represented initially in the system by two templates, the correct paraplate, when applied, would assert a spatial location case tie (SLOCA) between the two as follows:

(11) ![Graphical representation of SLOCA](image)

What is the shape of a paraplate in virtue of which it does this? Each paraplate corresponds to one of the cases in the inventory, which is the same inventory used inside formulas, in the way just described.

$$ The (6) way of writing formula (5) could be thought of as showing the sort of template (though with whole formulas where the elements of (6) are) that (5) is "trying" to be the central, action, node of.
Many paraphrases, however, may correspond to a single case. A paraphrase has the form of two \textit{template-types} connected by the case label of the paraphrase, where by template-type I mean an entity like a template except that, instead of a formula at each of its three nodes, it has a function ranging over formulas. Any template matching that paraphrase must have formulas that satisfy the functions in the corresponding part of the paraphrase. If the functions in both parts of a paraphrase are satisfied by a pair of templates (and the template for the prepositional phrase is normally considered to be the second or right-hand part, though this need not correspond to actual order of occurrence in text) then the case label of the paraphrase is asserted in the representation as holding between the two templates.

In earlier descriptions I have distinguished paraphrases sharply from inference rules, but in fact that can perfectly well be seen as a form of inference rules and, in that sense, I accept Schank's position who has argued for some time that case attachment is a form of inference. However, the essential role of paraphrases is as \textit{parsing structures} for prepositional phrases.

So, if we were representing "John picked up the statue made out of wood on the table after lunch" we would expect paraphrases for the various cases to end up asserting ties as follows:

(12) \[ \begin{array}{c}
\text{SOUR} \quad \text{made + out of wood} \\
\text{SLOCA} \quad \text{on the + table} \\
\text{TLOCA} \quad \text{after lunch}
\end{array} \]

where SOUR indicates source case, and TLOCA, time location.
The main point of interest in all this is the way in which paraplates are organized. In earlier versions, I wrote of them as an ordered stack, organized only under the name of the corresponding English preposition. This was misleading in two ways: (a) because they are also, and obviously, also organized under the basic action occurring in the left hand side of the paraplate. I have always considered this obvious but have not emphasized it in descriptions. (b) because the organization, even as occurring under an English preposition name and in connexion with a particular basic action, is still not a simple ordering, as I have certainly claimed. Or, concretely, even if we took all the paraplates in English for "by", and then selected out that subset called by an English main clause verb whose main semantic element was MOVE (say "go", "leave", "drive", "wander", "travel", etc.) we would still not have a simply ordered, but rather a partially ordered, stack of paraplates. An example of a set of paraplates should make this clear.

Let us consider "by", functioning in the following sentences, all of which may be considered to start, like (13), with "He left Comano by.....", where I have indicated the apparent case of the last clause at the right of each line:

(13) He left Comano by courtesy of the police, SOUR
(14) by the autostrada, DIRE (direction case)
(15) by car, INST
(16) by stealth, WAY (manner case)
(17) by Monday night, TLOCA
(18) by following the arrows, DIRE
(19) by stealing a boat, INST

Paraplates, as I said, are six-place entities, not all of whose places need be filled, corresponding to Agent-of-first-template, Action-of-first-template, Object-of-first-template, and the so on for the second template. Here are four paraplates that should match onto the templates for the sentences with corresponding numbers. Like the sentences, the paraplates will all have the same left-hand side, which is written only once.
These brackets containing formula parts are all to be interpreted as: matches onto corresponding part of a template if the latter has the mentioned subparts of formula. Thus (14)' matches (14) because the formula for "he" falls under/has as appropriate subpart *ANI left

- Lugano
- autostrada

and so on for the other correspondences of sentences and paraplates, which should assert the case label written at the right hand end of the sentence in each case (and on the corresponding arrow in the paraplate).

The paraplates above are some of those under "by" and accessed via the left hand main action MOVE. What I meant by (a) above is that this set of paraplates is clearly not ordered with respect to paraplates under "by" and accessed via any other basic action, such as BE as it occurs on the left of a paraplate matching onto "The painting is by Negri". The question I want to look at, and it is a central point of the chapter, is what are the ordering relations, if any, between the paraplates (14)'-(15)'-(18)'-(19)' . First, INST case is largely a default case for MOVE as it is cued in by "by", in the sense that almost any entity can be an instrument here if we have no reason to believe it is something else. Thus the more specific (14)' must be applied before (15)' in order to match direction case for (14), since, if the order were reversed (15)' might

$\square$ is a dummy place holder, *DO covers a wide class of actions, as does *REAL of entities

$$ though paraplates with differing left hand sides can be collapsed if their right hand sides are identical.
match with what "ought" to match with (14)'.' We could of course imagine something very specific in (15)' to match the formula for car, but that would risk missing "by cattle truck" which is not normally used for transporting people. A stronger but similar argument applies to putting (18)' above (19)' in a preference stack, given the very weak criterion embodied in (19)', that almost any action upon any physical object would satisfy in default, and one could extend this to putting a (13)' above a less specific (16)'.

The point here is as follows: it is clear that such paraplates must be applied in some order, but it is not clear that they are simply ordered. For example there is no reason why (18)' should be ordered with respect to (14)'. There is no formal trouble expressing a partial ordering of this sort procedurally. I argued in (Wilks 1975) that in many cases we might expect more than one template attached to a fragment at this stage and that we would "resolve the ambiguity" by preferring whichever template matched higher up such a preference stack of paraplates. Order, then, is important in such paraplate stacks, even if it turns out to be only partial order.

If such a stack is only partially ordered then we may, in the case of a prepositional phrase with two or more current competing templates for it, expect a draw at some stage: that is to say, two templates may match with paraplates that are not ordered with respect to each other. Such a situation might well correspond to an example such as

(20) He beat the girl with a withered arm

where the case dependence of the prepositional phrase could be either instrumental or a specification of which girl it was. In such cases one might imagine some expectational force to be drawn from the case content of the formulas. Thus, as we saw, "beat" contains an instrument in its formula, and so we can conceive of using this fact to decide the draw on the grounds that "beat" really "expects" an instrument, so why not give it one here, and settle the question. And there might indeed be some psychological grounds for doing that.

$ such as a formula expressing "thing used to move humans".
But, in general, it is clear that the paraplates to not in any sense function expectationally, except in the sense that by looking first at the paraplates most preferred in the stack (for a given preposition and basic action) one could be said to be expecting it. And that sense of expectation is perfectly consistent with trying more than one template, for a given phrase or clause, at the same time against the paraplate stack, as we might have to in a sentence like

(21) He beat the dog with a branch

where we would not only be resolving the case of the prepositional phrase, but also the word sense of "branch" (where the senses of "branch" might give rise to a corresponding number of templates).

If expectation means this, then we are dealing with an expectation system. However, as we shall see in a moment, Riesbeck uses that word to mean something stronger, and in particular something more depth- and less breadth-, first oriented than what I have described. In conclusion of this section let me point out that, just as the formulas yielded a Fillmorean verb frame analogue, so the paraplates yield an analogue with case frames for basic underlying actions. The paraplates under "by" whose left hand action is MOVE are part of the case frame of MOVE (as would also be all the paraplates under "in", "with" etc. that were accessed by MOVE). It will be clear that none of these are obligatory, since the paraplates link only templates already asserted by the text, and hence are not used in the way Schank fills the case slots of any primitive act "obligatorily".

$ In fact the only analogue of "obligatoriness" in the preference semantics system of cases lies in the later inferential function of formulas when the templates are "extracted" (see Wilks 1975, and chapter VII), which means inferences are drawn by unpacking every case subpart in the formulas of a template. So that, from a template for "He beat the dog", since, as we saw "beat" has an instrument that is normally headed THING, a new template would be extracted, not corresponding to anything on the surface sentence, equivalent to "a man uses a thing". This can be thought of as imposing obligatoriness on the case sub-parts of the formulas.
Expectation and attention: a digression

Within what one may now safely call the Artificial Intelligence paradigm of language understanding, there is broad agreement about the importance of densely structured knowledge-expressing entities, and these are not altogether misleadingly classified under Minsky's umbrella term "frame". In parsing terminology, frames are of their very nature top-down, that is to say that some bottom-level or surface item such as a word cues in a one or more of these very complex frames, and then the structure of the frame itself takes over the parsing process, seeking items to fill its slots etc. This is clearly different from bottom-up conceptions where entities associated with surface items are combined together to form a more complex structure.

However, frame-like structures are not so evidently committed in the other major choice of parsing strategy, depth-first versus breadth-first. Or, in common sense terms, following out several alternatives to see where they lead, or concentrating on one and following it till it succeeds or fails, and then later trying another. It is clear that depth-first parsing, whether semantic or syntactic, presumes upon a notion of failure and back up: that is to say, a notion of when a current structure has become inapplicable, and then of where, in the previous choices taken, to go back and try again.

But here we come to a slight tension in contemporary approaches: almost all of them are weak on failure and back up: that goes for Winograd, Riesbeck's implementation of Schank, and my own. Indeed part of point of top-down, frames, approach is to make backup much less needed. The tension comes in trying to combine this with a wholly depth-first parsing technique as Riesbeck (1974) has done, and in a moment I will consider contrasting examples in some detail, so as to show that a serious semantic analyser must have a greater breadth-first component than is generally realised. Or, to put what I am saying in slogan form, structures of expectation must be explored more than one at a time so as to simulate attention. Systems that do not attend are like people that do not attend but only expect what to hear. In neither case do they communicate or understand.
The approach I am criticising can be satirised by the example

(22) When John was at the races he saw a man streaking in jockey shorts

and a parser which reasons after the sixth word, "Ah, we're talking about the horse races", so "jockey" must refer to the men who ride the horses. One's reaction here may be, of course, no one could suggest a depth-first approach as crude as that, yet as I shall show in a minute the satire is not all that broad.

But a word is in order here on the source of the trouble about expectation and I will call it the phenomenological fallacy. It goes like this: when understanding language as human beings we are never conscious of alternative interpretations, the fact that a word we read in context has many senses out of context etc., therefore a semantic parser should not consider such alternatives either, for, if it has the right conceptual/semantic/preferential/frame structures in it, it will go directly in a depth-first manner to the correct reading and never consider any other.

The premise of this position is of course true, but the conclusion is totally false, and it is perhaps work setting out why.

Where it is wrong is in its assumption that the correct interpretation fits and the other possible interpretations do not fit at all. Hence the first path that can be followed will be the right one. The truth of the matter in semantics is that the right interpretation fits better than the others (or in the terminology I have used for some time, is more preferred) but to see that we necessarily have to see how well the other possibilities fit.

One detailed example Riesbeck discusses in the thesis is the parsing of

(23) John gave Mary the book.

The heart of the matter is that the action "give" generated a number of requests, one of which is that if what follows the verb immediately is human then assume it is RECIPIent case (p.98), otherwise assume it is
OBJECT case. There is nothing that I can see in the description he
gives that allows any back up to an alternative request should one fail
after appearing to succeed and this is consistent with the depth-first
no back-up position. I mention this because it seems so clearly weak
and is easier to follow than the preposition examples. So, for example,
both the following perfectly natural sentences would be wrongly analysed
on such a method:

(24) John gave Mary to the Sheik of X
(25) John gave his city his stamp collection.

There is nothing the least tricky or bizarre about such sentences, and
the example shows that simple unfettered expectation is not enough. Of
course, it may be that encountering "to" causes the existing structure to
be abandoned in favour of a more adequate one, and that would be one way
out for Riesbeck but he does not consider it. (He discounts such examples
on p.99).

This replacing of one structure by another on encountering a new
source of requests in the text is the strongest point in this system and
his key example is

(26) John gave Mary a beating,

where "give" sets up what one might call "conventional" structures which
are then overridden on encountering the requests generated by the represen-
tation of "beating". Thus all conventional structures, where one
person transfers possession of something to another, are abandoned as soon
as the particular expectations of "beating" are encountered, and a new
structure is set up in which one person hits another.

These two sorts of example are tackled rather differently in a pre-
ference system; let me mention only the RECIPIent example. The "John
gave Mary a book" sort of example is very easy to get right: "give"
has preference for a physical object or substance, but any real entity
immediately following it can be attached as its indirect object.
It is also an assumption (and this is the important point) that a representation with a filled-in object is a fuller representation than one without (the same would go for agents and actions, but objects are under discussion here). Thus in all the examples above the right structure will be obtained with these simple preference rules: in "John gave Mary the book", there is, of course, no problem; in "John gave Mary etc....", the system tries to attach Mary as a recipient but finds that that structure is therefore less full than one in which she is attached as the direct object and there is no recipient at that stage. Again with the "city and the stamp collection" there is no problem.

The simple method just described, although expectational, is prepared in each case to follow more than one path simultaneously (as to whether what follows the verb fits better as object or recipient, for example.)

**Riesbeck's parser**

I want to conclude with a few more details of Riesbeck's parser, not in a spirit of criticism, but simply because I feel that an important point, which may indeed be buried in the parser, needs bringing out if so, and emphasising in any case.

For each word in a left-right parse of a sentence, there are stored requests, and Riesbeck says specifically that these are not ordered (ibid. p.95). As I mentioned in connection with his best example, "John gave Mary a beating", different requests can take over and control the conceptual structure being built up. I shall confine myself here to a detail of prepositional structures since that is the centre of attention in this chapter.

Let us take the example

(27) John prevented Mary from leaving by locking her in.

There seem to be three phases here:
"prevent" generates requests (expectations) for FROM and BY1 (ibid. p.121). Let me call these objects preposition-forms, for the moment.

(iii) FROM itself has very complex requests associated with it (ibid. p.122) written as:

(28)

(TIMBED <= ((MOD QUOTE ((CANNOT))))
 (TIME CHOICE <= TIME)
 (SUBJ CHOICE OBJ))
 ((OR (EQ WORD (QUOTE BY))
   (BREAK_POINT))
 (RESET_ALL)NIL))

NIL)

This is clearly a complex object, but its role is clear, namely to tie together two conceptualizations: one for the preventing and one for the leaving, and in that sense its role is analogous to the paraplates described earlier. It is clear, too, that this request is seeking the preposition "by" to follow it (fourth line from bottom).

(iii) the requests of BY1 are an even more formidable object than that above, but Riesbeck makes clear that it is seeking forms of primitive act to fit a following "by" phrase should it occur, as it does in fact here.

Let me raise two preliminary questions that my reading does not settle in any straightforward way.

(a) what is a prepositional-form like FROM? Is it a case or not? At one point Riesbeck suggests that it is specific to "prevent" (ibid. p.122) which suggests that it cannot be a case in any general way, but is some other classification of preposition appearance, as post-verbs are, for example. On the other hand BY1, at least in this example, seems to be pretty much an introducer of a manner case, but then Riesbeck does say it is much wider than this example, and he doesn't give enough information on reading the formalism for one to guess whether or not it does correspond to a case.
(b) It is never made clear what happens if an expectation is not fulfilled, because they always seem to be. So it is never clear where, say, other FROMn's would come into play if FROM∅ failed to do the trick.

This last point is part of a very general question: is there anything corresponding to paraplate preference order in Riesbeck or not. I have argued, perhaps not convincingly, that there has to be some order imposed in case parsing, but he says specifically that the requests are not ordered. On the other hand, the strategy as described may imply an ordering: "prevent" calls FROM∅ and if that fails and there is a FROM1, say, available, then there is some ordering like that of application of the paraplates. But that might seem inconsistent with his claim that the requests are NOT ordered.

The same doubt continues about the "by" phrase: after all, in conceptual dependency "door" is, I believe, essentially an act, so if "by" is seeking an act, and "door" is an act, to establish the manner case, it will also presumably be happily matched by:

(29) John prevented Mary from leaving by the door

which seems intuitively to be direction, or something of that sort. Riesbeck may unconsciously accept another variant on what I called the phenomenalist fallacy, which is that people say what is expected of them, and so a preference order of structures is not necessary. But, alas, they all too often don't, and one has actually to attend to what they say.

Again, depending on the exact interpretation of prepositional-forms, the system is highly superficial in a strict sense: it is its verbs that seek prepositions (if FROM∅ is a preposition), rather than basic actions seeking cases. This surface method naturally makes it hard to state "significant semantic generalisations", and is particularly odd in a parser that claims to be based on Schank who abhors all processing based on surface correlations.

It should be clear that this final section is a request for clarification rather than a fundamental criticism. It is also in no way a criticism of Schank's system, since if my hunch is correct the system and the parser are entirely independent.
V Smaller Chunks of Knowledge

The title to this minute note is meant to provoke those who are busy hoisting flags in AI painted LARGER CHUNKS OF KNOWLEDGE. This hoisting has gone on in stages as everybody knows: one night distinguish the phase during which Winograd (1972) used a real world knowledge component in a parser to discuss examples like:

(1) He drove down the street in a car.

In that example "knowledge" is accessed to show that streets are not inside cars, in the normal course of events, and so the correct parsing shows us that the driving rather than the street is in the car.

In more recent discussions following upon Minsky's "frame proposals" (see chapter VII) the use of structured entities containing more information has been proposed, so that in this example, a frame apologist might argue, we get it right, not by deductive knowledge based on a fact, but by knowing where we are in some very large piece of data that gives us, in some form, the normal course of events when driving. In that way the fact that the person is in the car might hardly need to be parsed on to the sentence, but would follow simply from where we already were in a "driving frame" when the sentence (1) was encountered as part of a story.

These hypotheses may well turn out to be correct but, in the meantime, it still seems to me worth pointing out that there may be very general rules, having more linguistic content, which might settle a very wide range of such problems in a general way, and without accessing specific knowledge at all. Whether or not this view is true, everyone must admit that it would be nice if it was, because the general rules, or inference mechanisms if you like, would mean that we did not have to install all the specific common sense coding that now seems to loom before us until our retiring ages: you take supermarkets, and I'll take pubs and he'll take gardening and so on. If there are general rules, they would not only save space and time, but prove that the human being is at heart a generaliser and a lazy one at that, all of which I would like to believe. In
other words, it seems to me that much of this might be done by general representational strategies rather than ad hoc ones.

Let me make my point in a minute way by bringing up an old favourite that never completely disappeared: Bar Hillel's famous "disproof of machine translation". There is a moral advantage in mentioning it from time to time, because it stops AI workers thinking that they discovered the role of real world knowledge in language understanding.

In brief, Bar Hillel's (1958) example was the following "children's story":

LITTLE JOHN WAS LOOKING FOR HIS TOY BOX. FINALLY HE FOUND IT.
THE BOX WAS IN THE PEN. JOHN WAS VERY HAPPY.

Bar Hillel's focus is on the third sentence, "The box was in the pen", whose last word we naturally interpret in context as meaning "playpen" and not "writing pen". Bar Hillel argued persuasively that to resolve this correctly requires knowledge of the real world, in some clear sense: at least in the sense that the difficulty cannot be overcome in terms of some simpleminded "overlap of concepts", by arguing that the concepts of baby and playpen can be seen, by lexical decomposition of some sort, to be related in a way the concepts of baby and writing pen are not. Bar Hillel argues that that would not do, because the story would have been understood the same way if the third sentence had been "The inkstand was in the pen", where the "overlap of concepts" would now be between "inkstand" and "writing pen" which would yield the wrong answer on the same principles.

He was clearly right that the machine translation (MT) approaches of that time could not possibly have coped with the forms of inference that this example requires. Where I think he was wrong was to argue that no computable system of rules could, in principle, do so, and hence MT in the required sense was impossible. AI has shown how such examples can be tackled, and in more than one way.
A number of researchers in AI would certainly be prepared in principle to discuss the treatment of this example these days, and I choose to discuss it within my own only by way of illustration.

Some current systems would want to discuss the sentence "The box was in the pen" in terms of inference rules and structured "frames" for children's play situations. In my view the example can be dealt with at an earlier stage, during the establishment of the representation for the sentence itself. First, let us assume two sense formulas, one for each sense of "pen":

```
"playpen"

WRAP) (SUBJ THING)
  
  GOAL)

(childOBJE) (MOVE NOTLET)

"writing pen"

WRAP) (SUBJ)
  
  GOAL) (liquidOBJE) (LINE THING)

  MAKE)

(SIGNOBJE)
```
There is one very slight notational change here from the formulas as described earlier: I have inserted the "lower case" items "child" and "liquid". This in no way violates the rule that formulas consist wholly of semantic primitives, because these apparent non-primitive words are only shorthand, as it were, for other formulas. Thus, "liquid" occurs there only because there is, elsewhere in the system, another formula for the English word "liquid", and in any operation with the formula for "writing pen", the "liquid" formula would first be inserted at the appropriate point.

This relation, to allow formulas to be "reentrant" (see chapter II), makes them easier to read, but without allowing an uncontrolled proliferation of non-primitive words into the formulas.

Let us now consider the treatment of "The box is / in the pen". The sentence will have been fragmented at the stroke, and a template attached to each part: the first template having a dummy object place, and the second a dummy agent place, since a formula for "in" has become its pseudo-action, and that has no agent. Structures called paraplates, whose nature need not detain us (see chapter IV) then seek to link the two templates back together, the paraplate being in effect a case frame that resolves the "in" as the introducer of the CONTAINMENT case. The application of this paraplate allows the dummy agent of the second template to be "repacked" in this particular case frame by the agent of the first template and thus we obtain, by a "repacking inference", a template form equivalent to "box in pen", which is not, of course, an assertion explicitly present in the text. This "repacked template form" will have a formula for "box" at its agent node, and since we still have two formulas for "pen" in play, not having yet resolved between them, we shall in fact have two repacked templates at this point, both with a formula for "box" at their agent node, and with the two formulas for "pen" at their respective object nodes. Or, in "square bracket", shorthand form,

\[
\text{[box in pen]} = \\
\langle \text{---THING} \text{---PBE} \text{[playpen]} > \\
\text{and} \\
\langle \text{---THING} \text{---PBE} \text{[writing pen]} >
\text{.}
\]
where \([\text{playpen}]\) stands for the formula for "playpen" etc., and \((---\text{THING})\) stands for \([\text{box}]\), the formula for "box", whose head is THING etc. Next, after repacking, the two new template forms both try to expand themselves, just as a template for a text item would do at an earlier stage, to see how many preferences each can satisfy, and the "most satisfied" will be preferred. The extension to the previous descriptions of the system is this repeated expansion, performed at any post-inferential stage if competing representations are still to hand.

Inside the formula tree for "playpen" we see that playpens prefer to contain children, while writing pens prefer to contain liquids. And, since a box is a physical object (head of its formula is THING), and so is a child, while a liquid (head of its formula is STUFF) is not a physical object, it is clear that the template with the "playpen" formula is more satisfied than the other, and so is preferred. That is to say, given that the heads of formulas for "child", "box" and "liquid" are MAN, THING and STUFF, respectively, we can state this conceptual inclusion as

\[
\text{THING} \supset \text{MAN} \quad (\text{humans are objects})
\]

but \[
\text{STUFF} \supset \text{THING} \quad (\text{substances are not objects})
\]

This method is proof against the "Inkstand in pen" point of Bar Hillel's mentioned earlier. Furthermore, this solution is general, and the ambiguity is resolved by the application of the general rules of preference used in setting up the representation, and not in any sense by special rules for the example.
VI  What sort of taxonomy of causation do we need for
language understanding

This paper is in four sections as follows:

A. Introduction, which contains some general discussion of the dif-
ficult notions of CAUSE, REASON, MOTIVE, INTENTION, etc.

B. A proposal offers a suggestion as to how to introduce the notions
of cause and reason procedurally into a language understanding sys-
tem, and without any taxonomy of types of cause. It proposes a
hypothesis, one that is hopefully even refutable, and therefore not
vacuous; that one should prefer reason-like explanations of actions
for human like agents, if one can find them in what is being analysed;
whereas for other types of event one should prefer cause-like expla-
nations.

C. A little light machinery reminds readers of the existing state of
the preference semantics system, and of what changes would have to
be made to incorporate the above proposal, and to find examples to
test it.

D. Some comparisons, which is a very brief comparison with the sugges-
tions of others in the area of cause.

Anything interesting or novel in the paper is in sections B, and
more especially C, so a reader who dislikes general discursive discussion
should proceed directly to one of them.
A. Introduction

Any attempt to represent language, or to represent it indirectly via some logical formalism, inevitably requires a concept of causality. In this chapter I want to discuss the topic generally, then to indicate how the concept is dealt with in the preference semantics system of language understanding, and lastly to contrast the latter briefly with the methods adopted by others working within the artificial intelligence paradigm.

Some general discussion is unavoidable: partly because of the way that the notion of causality is entwined with other difficult concepts such as intention, reason, and motive; and partly because of the effect of the philosophical treatment of the notion. The key point in that tradition, from the point of view of those interested in the representation of knowledge, was Hume's denial that there was any knowledge of causation to be obtained from the real world by empirical methods, followed by Kant's attempt to refute Hume's view, but which shared with it the key assumption that causation is in some sense a category which we impose upon the real world, and without which we could not understand it.

To one concerned with the problem of representing the semantic content of natural language for some well-defined purpose, the philosophical tradition has had both good and bad effects. It has to some extent clarified what the content of our commonsense understanding of causality is, and that is a good thing if one accepts that it is this latter understanding that we ultimately are after. On the other hand, philosophical views of causality in degenerate forms have had a bad effect on AI writers, as I shall try to show, because they have led to the adoption, consciously or unconsciously, of philosophical positions that are at variance with commonsense. For example, there is a traditional position in philosophy that "reason" explanations of actions are in fact no more than disguised forms of "cause" explanations. The best known proponent of that view is Davidson (1963). It is not my intention to discuss the view here, only to point out that this philosophic view is certainly at variance with commonsense, where we do indeed make the distinction, as between, for example:
(1) John fell down the manhole when he slipped on a banana skin and

(2) John put his winnings on red because he lost three times on black.

In (1) we can ask "What caused John to fall down the manhole?", but we would not normally ask "Why did he do it?", or "What reason did he have". Of (2), however, we can sensibly ask why he did it. In ordinary speech we can say of (2) also, "What caused him to do that?", but the distinction between the situations seems to remain.

I am maintaining two things then, about the influence of philosophy on our subject matter:

(i) What we are after is the commonsense understanding of causality.

(ii) Some philosophical discussions may have done something to clarify the content of (i), but many have had the opposite effect, and some of these latter have been unconsciously absorbed by AI workers, who might be horrified to know that their views were philosophical positions.

One consequence of (i) is that we should not expect to solve any philosophical issues in the course of our task. Notice also, that (i) involves a very strong assumption that this commonsense understanding, whatever it may be, is indeed what is required by a language understanding program rather than technical knowledge, or philosophically clear views.

However, some discursive clarification of awkward terms like REASON, CAUSE, INTENTION, PURPOSE, MOTIVE is essential, so let us turn to that.

The difficulty about those terms, as everyone knows, is that they are simultaneously words of everyday commonsense discussion, AND of semi-technical fields like psychology, philosophy, etc. All one can hope to do is to sharpen them up a little without being untrue to their commonsense meanings.

(a) Two families of concepts

It seems fairly clear that we have two groups or families of concepts with which to explain human action. On the one hand is a family whose principal concept is cause, and on the other is the family of reason,
purpose, intention, and to some extent, motive. There is a tradition going back to Aristotle of describing the latter as types of the former — he called purposes "final causes" — but this seems in some ways as counter to commonsense as his decision to call the material of which a thing was made a material cause of its existence. Muddling of the two families is seen in the famous remark of Bismarck (who sought explanations for everything in terms of motive and intention) when he heard of the death of an enemy statesman: "I wonder what he meant by that?".

Roughly, the force behind the distinction is that cause is a "push concept" operating, as it were, on humans to produce whatever behaviour they exhibit, whereas the latter family are "pull concepts". They correspond, as it were, to future states of the world which draw behaviour out of human beings.

The peculiar member of the latter family, on that diagnosis, is motive. Motive is a way of describing reasons in the terminology of causes: as when a detective might ask "Who was motivated to do X?". I will not treat the interesting differences between the concepts of reason, motive$ and purpose here, but subsume them under a single notion that I shall denote by the familiar GOAL, which I have also used to denote the "purpose case" in semantic representation of natural language. Let us, instead, turn to wants, which have much in common with motives.

(b) Wants and volitions

The concept of wanting, or desiring, has two aspects: it may introduce states desired in a way indistinguishable from reasons or purposes. An answer to "Why did you kill your aunt?" may be "I wanted the top flat to be empty". However, the answer "Because I wanted to" to the same question is not an answer of the same type. The latter answer, in fact, tells us nothing over and above that the action was voluntary. It seems a moot point whether one should ascribe empty explanations of the

$ as when we speak of "unconscious motives" but not "unconscious reasons"
sort given by the last reply to the category of causes or reasons. The fact that causes of human action are normally thought of as excluding voluntariness, in so far as they are causes, suggests that it go into the class of reasons. The consideration that reasons and purposes are normally thought of as extending into the doer's future suggests that trivial, or "pushing", wanting go into the category of causes. Neither choice seems wholly satisfactory, and I shall opt here for the reason category, or GOAL.

It is worth while noticing noticing that the reason-cause category distinction has nothing necessarily to do with either voluntariness or prediction. So, for example, the rain may cause man to get wet, but he may well have wanted this cause to act on him in that way and deliberately gone outside without an umbrella so as to get wet. The man who takes an upper knowing it will cause him to be aggressive may well have wanted that (caused) behaviour.

Again with prediction, we naturally assume that actions are predicted by the doer, and a reasonable action may be predicted by the non-doer. However, purely caused behaviour may be equally predictable, just as one who falls off the cliff may be predicted to hit the rocks below. High grade legal minds (whose views are always relevant to these matters) can get mixed up here as when, in a recent Court of Appeal case (Times April 18, 1973), the judges considered a mental nurse who had had an overdose of insulin and, with the apparent collaboration of another nurse, had attacked a patient. They decided that the first nurse had a valid defence of automatism (rather than insanity) and that "If (the first nurse) acted without conscious volition, it was most unlikely that (the second nurse) would have known what he intended to do", and so the second nurse was also acquitted. It takes only a moment's reflection to see that the reasoning is quite wrong, and we can be quite aware of what a person, whose behaviour is wholly caused, is going to do: at a seance by a successful hypnotist we expect the (involuntary) behaviour of the subject to be whatever the hypnotist has ordered.
(c) Actions and behaviour

The demarcation of the two families of concepts above correlates with that between human actions and mere behaviour. By definition, human action has potential explanation in terms of reasons or purposes; and human behaviour that does not have such explanation is mere behaviour or automatism or insane activity.

So then, put very crudely, events in physical objects are generally caused, whether or not we know the causes. Objects do not normally have purposes or reasons, though we would not be astonished to read fairy stories in which rocks tried to drop on passers by.

People on the other hand may have their behaviour caused, as by drugs or gravity when falling over a cliff, but in normal common sense discussion their actions are explained by reasons and purposes, and we seek these in understanding their actions. In the case of persons, causal and reasoning explanations are not exclusive, and we quite properly say of someone that the drugs he is taking for some condition tend to make him aggressive, while at the same time that the fight he got into last night was to show his wife that he wasn't afraid of a mugger. It is only in metaphysical arguments that exclusivity of causes and reasons, or the apportioning of efficacy between them, becomes an issue.

Another point at which metaphysical aspirations creep in is over the notion of one person making another do something, or of causing them to do something, as in "She made her child put the toy down". No considerations about the abstract concept of freewill are involved in such common-sense doings, only that there was causation by person one on person two, and that is in no way compatible with persons one and two also having their reasons. $$

$\text{Though "behaviour" may not be a simple concept for which one can just lay down a use. In the High Court (\textit{Times}, December 14, 1973) a judge refused a man a divorce from his wife who was in pre-senile dementia in hospital, on the grounds that, in that state, she was not capable of "behaviour", and so not of behaviour that could lead to a breakdown of the marriage.}$

$$\text{As is "You made me love you; I didn't want to do it ......."}
I suggest that there is no place in commonsense understanding for determining causes of human behaviour, except in extreme cases. If someone slips and falls over a cliff, and let us suppose is killed at the bottom, we accept the laws of gravity as a complete explanation of what happened to him or her once contact with the cliff was lost.

My point is that the complete explanation of the fall by the laws of gravity and air resistance are the exception where human behaviour and activity are concerned. It is not only that cause and reason explanations are consistent and not contradictory, but that all our commonsense explanations normally presume upon an undefined and unenumerated universe of contributory factors that we do not and could not consider. There is no presumption that the causes and reasons we seize upon in explanation are all there are, or all that matter. So when we say, as before, that "She made her child put the toy down" we in no way exclude that the child wanted to please its mother, that it was bored with the toy anyway, although, and this is the difficult matter, we usually are denying that she did not intend to achieve that because she was only shouting at a passer-by in the street and her guilty child misinterpreted her.

(d) Intention is not a fundamental concept in this approach

Intention is a concept that has been much talked about in contemporary linguistics and more recently in artificial intelligence. My feeling is that it cannot bear the weight that has been put upon it. Like wanting, discussed earlier, the concept has a certain duality. The duality, I hesitate to call it ambiguity, is between intention as a concept close to reason or purpose, on the one hand, and intention as a concept to do with will power, on the other. The latter concerns when we say of someone that he intends to get up early the following day, but we are not sure whether he will make it. In that use we explicitly distinguish between intention and its achievement — it is a concept close to trying.
The more usual aspect of the concept in discussions like the present one is that of a notion closely related to purpose and reason, as when we say:

intention in
"His reason for starting the fight was to get the concert stopped."
purpose in

They are more or less synonymous, the difference being, I suggest, that "intention" carries as always a flavour of its will-power use, in that, more than the other two alternatives, it brings into question whether or not he succeeded in getting the concert stopped. This consideration is normally irrelevant to everyday explanation of events where we are told what did or did not happen. In the examples in the AI literature where it is introduced such as "What was his intention in giving me the pills to make me better", it is certainly not the purpose of such authors (Schank, for example) to highlight the possibility that he might not succeed. What is wanted is, I believe, adequately covered by GOAL, and the added specific overtones of "intention" are irrelevant and misleading. Again, another drawback of intention as a fundamental concept, from a procedural, language analysis and understanding point of view, is that it is essentially a subjective, state of mind of the doer, what-do-I-mean-by concept rather than an objective, what-is-to-be-understood-by, concept. This is why I believe most of the work from the "speech act school" has little to say to those concerned with language understanding programs.

I do not want to be misunderstood here: it is clear that a serious language understanding system may well need to cope with sentences containing "intend", and to grasp what is meant by: "When he fired the shot he intended to hit the President even though he hit his mother who was standing behind him". This use of "intend" is properly covered by an approach that copes with intentions as cognates of reasons or purposes, and more particularly those which do not succeed in being realised.
However, it is clear from that that intention is a derivative, or special case, concept. What I am objecting to is attempts to make it a central concept in language understanding. It seems to me that that could only be done by having a through-going subjective system, and those who make the concept of intention central do not appear to be prepared to do this, or even to think out what it would be to have a general language understander where all sentences were considered from the point of view of the doer rather than the reader. Even if it could be constructed it is not clear that such a system would be close to our own understanding of language.

I shall take it that, for our purposes, what is needed from the concept of intention can be stated in terms of GOAL: so we might put the usual "to intend X is not to intend all X's consequences, but to intend X is to intend all the means necessary to achieve X" as "having X as a GOAL does not entail having Y as a goal where Y is a causal consequence of X, but having X as a GOAL does entail having Y as a GOAL where Y is the normal cause of X."

(e) States, events and non-events

Many people starting out to discuss these matters at a more than discursive level find it necessary to begin by setting up a proto-logic of states, events, actions and whatever. This seems to me quite untrue to

$ There are classical taxonomies of causation in which such notions are central. Collingwood (1940), for example, takes causation in the sense of our causal ability, raise our own arms, say, as being one of the fundamental types of causation, and quite different from our knowledge of other people's ability to raise their arms. This is very much an "intentional" view of causality, but what would it be like to build a language understander based on it? Paradoxically, it would in fact be like those current understanding systems based on the premise that understanding is doing something with your hands, like the Winograd (1972) system. But there are troubles about that premise for other reasons, in particular because most understanding is not of that sort. Reading a newspaper for example is not much connected with abilities to influence the situations one is reading about in it.
commonsense, where distinction between such abstract entities is in reality quite impossible to pin down, and moreover such a taxonomy usually turns out to serve no particular purpose that cannot be stated more simply.

So, for example, a strong metaphysics of events and actions often leads to trouble in distinguishing, as separate events or actions, what are in fact the same event or action. But "John lifted his leg and kicked the door open" does not describe two actions or events but only one. Again, a metaphysics of events can find it hard to accommodate non-actions that are essential to understanding, just as sins of omission can be as serious as those of commission. In a case in the Court of Appeal (Times July 17, 1975) the judges had to consider the appeal, from a conviction of manslaughter, of a man who stabbed a girl who later died, but who was a Jehovah's witness and had refused a blood transfusion, an omission which had undoubtedly contributed to her death since the wound was not that serious. He argued that he was not guilty since she had broken the causal chain leading from his stabbing action; a chain that would be normally taken to include a transfusion. The judges rejected the appeal, creating in doing so the memorable phrase that "the violent must take their victims as they find them", and provided a classic example of the role of omission in a causal chain.

In what follows, the template structure representing, or derived for, a text will supersede any definitions in terms of events, states and actions. Template concatenations will be the units to which inferences apply, and whether or not the templates represent events or states or actions will be only a derivative question: to be decided, if at all, in terms of the semantic type of the head element of the formula in the agent position of the template, etc.
Summary of A

The discussion of this section can be summarised in the form of assertions as follows:

- REASON, PURPOSE, WANT, DESIRE, GOAL, MOTIVE and INTENTION form a family of explanatory concepts different from that of CAUSE.

- MOTIVE, DESIRE and WANT have somewhat peculiar behaviour but can still be contained within the first class.

- INTENTION has a particular overtone of ATTEMPT, and we will not use it as a basic category as some do, but assimilate its use to the overall one of GOAL.

- The first category concepts apply to and define human ACTION.

- The distinction of events, states, actions, etc. is for us merely a derivative one defined in terms of the template representations. It is not an a priori one on which anything explanatory rests.
B. A proposal

Discursive discussion of causality, and related notions, is of course not enough, and it is assumed that every structure we propose, and every coding of the notion of causality, must be at least relatable in principle to concrete needs of language analysis. Or, to put the matter the other way round, it is assumed that it is not enough at any point to say, "We need causal structure X because without it the system could not understand." In a serious AI enterprise the last word "understand" must be replaced by some concrete and non-trivial task. It should not be necessary to labour this point, were it not for a development in AI, that Dreyfus among others would no doubt be pleased to see, namely that of understanding becoming an undefined end in itself: an irreducible phenomenological primitive unconnected with any test or task.

Even question answering tasks can become trivial in this sense. So, for example, a question answering system might be developed that "understood" the nature of angels. When asked "How many angels can stand on the head of a pin?" it might correctly, and indeed infallibly, reply "15". Understanding has not been in any serious sense demonstrated of course because the questions are framed in the light of the "knowledge" contained and the whole business is circular.

This chapter also tries to separate off questions about the representation of causality from those concerned with the need for, and nature of, "higher level structures", "larger chunks of knowledge" or frames — these I discuss in chapter VII below. The assumption that the notions can be separated is a strong one, because in the view of some advocates of frame like structures the fixed sequence of events, say in an activity frame like Charniak's (1975) for shopping in a supermarket, itself explicates in part the concept of causality.

The theme of the chapter is simply this: we do not need any taxonomy of causation, and one primitive CAUSE will procedurally cover most, if not all, superficial taxonomies in the way that it has until now in the preference semantics system. However, the more sophisticated considerations
to do with causation and reasoning can be tackled by organized bodies of inference rules, operating on template structures, and distinguished as CAUSE and GOAL rules.

The procedural proposal, and it is not, I think, a self-evident one, is that we do well, as a heuristic, to seek first, or preferentially, for GOAL explanations of human-like behaviour even when a CAUSE one may be superficially indicated. Correspondingly, we should seek first for CAUSE explanations of non-human events and behaviour. Systems are at too primitive a stage as yet for any claim that this heuristic of preferential GOAL searching is in any strong sense more efficient, only in a weak sense with regard to the present system of organization of inference rules.

What do I mean by the phrase "taxonomy of causation?" Well, I take an assertion of the need for a taxonomy to be that there are essentially different types of causation and that they require different representation. The normal justification for a taxonomy in AI terms is that it limits inferencing (see Charniak (1975), Rieger (1975), and Schank (1975)) and I will return to discussion of that justification in section D.

Let us consider the sentences:

(1) John closed the door

and

(2) The wind closed the door.

(2) introduces the notion of natural forces, and the question about such examples within linguistics has normally been as to whether or not the wind was to be described as the agent of the proposition asserted, in the sense in which John is the agent of (1).
Let us add:

(3) The rock fell off the cliff and crushed John's lunch.

(4) Zeus, angry at the Athenians, loosed thunderbolts in the streets.

(5) ITT moved all their operations from Chile because they feared nationalisation.

(6) Fido closed the door.

(7) The PDP-10 has taken away half of my disc file area.

(8) The ants marched towards the sugar pile.

The view of causality and agency that I am suggesting here and will develop in section C below in terms of a semantic representation, is simply that in the case of human, domestic animals, corporations and gods, examples (1), (4), (5) and (6), we are not obliged to expand an event backwards by a causal explanation, even though in particular circumstances we may do so, but will, when inferring, prefer to expand by a reasoning explanation if we can. Whereas, in the case of natural forces, physical objects, and non-domestic animals, examples (2), (3) and (8), we prefer to expand by causal explanation if we can. I will leave machines (7) to the reader's taste.

This classification of the universe is oversimple and is intended to correspond to commonsense and no more. It is thus independent of questions as to whether or not sentences like (4) are only restatements of sentences like (2), or those like (5) are only a collection of those like (1).

In stating the proposal I wrote of preferring GOAL explanations to CAUSE explanations for human-like agents, and this must now be explained in procedural terms. In order to do that, I must remind the reader of the existing scheme of preference semantics representation.

The semantic matching programs, using preferential rules, match pieces of English into sequences of templates. An actual template, say for

(9) John shut the door

is a network of trees like this:
(10) "a particular male human"

Template
agent = (John)

(Template
action = (shut)

"by a human causes to move a thing in order that there will not be an aperture"

Template
object = (door)

"an object used by a human, and whose source is plant-stuff (wood), in order that animate things will not be able to use an aperture"

((PLANT STUFF) SOUR) (((THRU PART) OBJE) (NOTUSE *ANI)) GOAL) ((MAN USE) OBJE THING)
This simple template consists of three formulas (trees) of semantic primitives. The trees are in Agent, Action and Object positions respectively. I will not explain the syntax of trees here beyond the explanation of the meaning of each given in English at each template node. (The "the" should be indicated by being added as the first atom of the formula for door.)

It is clearly exhausting to write out each template on a page (and this is a simple one with no dependent nodes, as would occur in "the nice grocer shut the kitchen door quickly"), and so the whole object above will, in what follows, be written as:

(11) [John shut the+door]

Texts, as we all know, do not normally come in nice template-sized pieces. A more normal representation would be a tied sequence of templates, as follows for:

(12) John bought a car in the market and liked it immediately.

(13)

1.  [John bought a+car]
    1.1 [[John has a+car]]
    1.2 [[John use car ]]
    1.3 [[John paid money]]

2.  [John in market]

3.  [John immed-liked car$]

This representation is the result of three separate processes operating on the original three templates (Nos. 1, 2, 3) representing the text.

$ Ringing "car" here indicates that this is the resolved value of ?it after application of the inference rule.
I PARAPLAITE CASE TIES (curved template → template lines)

No. 2 has been tied to the central node of No. 1 by a tie labelled SLOC — the spatial location case. It is tied to the second node because intuitively it is the "buying" which is located at the market, not the car, as in "He bought the car in the market". This tie has been achieved by the paraplate structures (see chapter 4 above) whose form is <template> → <template> and which are organized under the names of the English prepositions which call them.

II EXTRACTIONS (double [ ] bracketted objects).

Nos. 1.1, 1.2 and 1.3 in double brackets are template-like forms which are added to a representation prior to the application of inference rules. They have been extracted from No. 1, as is shown by the dotted arrows connecting them. The two clearly defined types of extraction (see Wilks 1975) are:

(a) those that come from the unpacking of each case type (except AGENT and OBJECT) in the formulas of the source template (No. 1), and
(b) those which come from the violated preferences.

Thus, for the first type, we have in the formula for "bought" a GOAL subpart which says that the purpose of buying something is to use it, hence, from the template No. 1, 1.2 has been extracted. Similarly 1.3 follows by an extraction of the WAY (or manner) case in No. 1. 1.1 intuitively is a causal consequence extracted from No. 1, since "buy" is basically a GET action, its formula being (14):

(14)

(*HUM SUBJ) (*REAL OBJE) (USE GOAL) (((GET SIGN OBJE) (*HUM GIVE)) WAY) GET)
or, a getting-done by humans to real objects in order to use them, the manner being the human giving money = (GET SIGN). Thus, the three extraction arrows in the representation above could in fact be labelled CAUSE, GOAL and WAY respectively. We shall not concern ourselves here with those extractions that are "driven" by violated preferences since, although fundamental to the system as a whole, they are not of immediate relevance. The reader should note in passing that the role of extraction covers that assigned to the many and various notions of presupposition current in contemporary linguistics.

III  INFERENCES (square "template → template" links)

No. 3 has been tied to No. 1 in the representation above by an inference arrow (shown as a square line) on the basis of which the ?it in No. 3 has been correctly replaced by the formula for "car". Inference rules have the form <template> → <template> but always contain restricted variables. In this case the rule tying No. 1 to No. 3 would have been:

(15)  \[
\begin{align*}
&\text{animate 1 cause + self + have realobject 2} \\
\Rightarrow &\text{animate 1 *judges 2}
\end{align*}
\]

where "*judges" is represented by a general function satisfied by actions of liking, as in the present example, and the square brackets show that we are dealing with only the "readable shorthand" form of the rule. Extractions above are also inferences, of course, and this procedure (III) is distinguished by the application of a body of rules with variables applied to existing template, and template-like, forms to tie them together, i.e. these rules do not produce fresh template-like forms in the way extractions do. The tie arrow No. 1 → No. 3 in the figure is not in any strong sense part of the representation, since its whole content is exhausted by the correct filling in of the variable "?it".

It should be noted that after extraction phase II, and before the application of the rules phase III, the system attempts to resolve pronouns by matching template forms against each other directly, and in (Wilks 1975) this was called a zero-point strategy.
From the point of view of explicating causality and implementing the preference for GOAL links as described above, there are two principal defects in the present scheme of things.

(i) The causal extractions are not as well defined procedurally as those for case unpacking and for unfulfilled preferences. So, for example, since the link No.1 to 1.1 in (13) above is not labelled CAUSE, the causality remains implicit in that the action formula of No. 1 has as a matter of fact GET as its head, and that is considered to be a causal head. Yet it might well be argued that No. 1.3 (the paying of the money) causes No. 1.1 (the having of the car), yet this cannot be stated if all extraction links flow from the source template and are unlabelled.

This seems unsatisfactory if we want to set up what are intuitively thought of as causal chains: template/events joined together by CAUSE links, so as to represent causal linkage in the real world. The "if" is a big one, because as I argued earlier a causal chain is one of those items of degenerate metaphysical baggage that we might well beware of when explicating commonsense reasoning. In fact my conclusion at the end will be that the notion of causal chain does not need explicit installation in a language underlaver. But let us admit at the moment that its lack might well seem a defect.

(ii) To implement the preference for GOAL chains over CAUSE chains we shall need a taxonomy of inference rules that at least distinguishes these two types of rule, and in the system as described till now rules were not typed, but only organized under the head, or main action, of their consituent (template-matching) parts. There was no division into causal-, and reason-type rules.

In section C I shall make procedural suggestions for overcoming these defects. But first some discussion is needed of what sort of protocols need to be covered by such suggestions and by the general "GOAL-preference" suggestion.
Thus (3) above initially has templates:

1. [Rock fall - ]
2. DIRE [ - off cliff ]
3. [Rock crush John's + lunch ]

where the curved tie is inserted by case paraplates and the square tie by inference rules in the way described in (Wilks 1975).

The present suggestion requires that further tying of the first template should be preferably by other causal ties since it has a non-human superficial agent, even though we can imagine fairy tales where rocks decide, or try, to fall off cliffs.

What would require such an attempt to set up further causal links?

Suppose we have

(17) The rock fell off the cliff and crushed John's lunch. Peter pushed it.

We have to tie "?it" to either "cliff" or "lunch" and we can all see what the right answer is, and can imagine it established preferentially by an inference rule:

(18) [animate 1 push real 2] \(
\rightarrow
\) [2 fall. - ]

but of course we can also equally well imagine a rule

(19) [animate 1 push real 2] \(
\rightarrow
\) [2 crush real 3]

But (19) will be preferred since it matches more variables in the templates than the other one. What we cannot easily imagine is a rule that would lead to a match in which the lunch, rather than the rock, was pushed.

Let us contrast this causal chain development with a reasoning chain development. Let us take

(20) John closed the door to shut in the dog.

(21) John closed the door to keep the dog in. He did not want it muddy.

Let us set this out as a sequence of templates as follows:
Let us note several minor points in passing: the agent position in templates 2. and 3. are filled in by straightforward preference considerations, although the ?it in template 4. cannot be, since both dogs and doors can equally well be muddy. 2.1 (in double []) is the extraction mentioned earlier.

Let me now introduce two notions very rapidly (and delay their detailed explanation and justification until section C), so as to enable the one main point of this chapter to be stated in a diagram. Let us tie templates and extracted template forms together by ties curved G(OAL) and C(AUSE) in just the way that they are normally tied in a representation by case ties such as TLOC (time location), as in

\[
(23) \quad \text{TLOC} \left< \begin{array}{c}
\text{John} & \text{left} & \text{the + University} \\
\text{-} & \text{after} & \text{the + lecture}
\end{array} \right>
\]

Thus we have:

\[
(24) \quad \text{C} \left< \begin{array}{c}
\text{John} & \text{pushed} & \text{the + rock} \\
\text{rock} & \text{fell} & \text{-}
\end{array} \right>
\]

and

\[
(25) \quad \text{G} \left< \begin{array}{c}
\text{John} & \text{helped} & \text{drowning + man} \\
\text{John} & \text{knew} & \text{drowning + man}
\end{array} \right>
\]

Let us also delay to section C the question of what actual procedures attach these arrows, and what it is in the surface sentence or in our stored inference mechanism that causes us to attach them as we do. We just note for the moment, that, in a natural manner, both G- and C-type arrows lead to the template expressing the event or action to be explained.
Thus, to return to our example (21), we could insert arrows into (22) as follows:

(26) 1. \[\text{John} \quad \text{shut} \quad \text{door}\]
1.1 \[\text{[door} \quad \text{be} \quad \text{shut}]\]
2. \[\text{[John} \quad \text{keep+in} \quad \text{dog}\]
2.1 \[\text{[dog} \quad \text{not+be} \quad \text{outside}\]
3. \[\text{[John} \quad \text{want} \quad -\]
4. \[\text{[?it} \quad \text{not+be} \quad \text{muddy}\]

2. \(\xrightarrow{G} 1.\) is cued by the "to" in (21) and 2. \(\xrightarrow{G} 1.\) is cued by a natural rule that any extraction (like No. 2.1) of a goal template (like No. 2) is also a goal to the same target template (like No. 1.) There are no natural C ties: one might argue that 1 \(\xrightarrow{C} 2.\), if No. 2. actually happened, but we do not know that that state (the dog being kept in) has in fact been achieved, for, unbeknown to John, the dog might have been outside in the garden all the time. At this point in the story, the GOAL tie more properly expresses what is going on than would a CAUSE tie. In representing

(27) \[\text{[John} \quad \text{bought} \quad a+cake\]
\[\text{[to+please} \quad \text{his+mother}\]
\[\text{[she} \quad \text{loved} \quad ?it\]

however, we can insert a C-tie, because we are told the result.

Now to return to (26) we can see the point of all this, I hope. Cued by "John wants" we can tie templates 4 \(\xrightarrow{G} 2.1\) by actively seeking for a further immediate extension to the Goal chain since it is a human-initiated event at No. 1 that was our starting point. Then, given some inference rule, or in this case direct matching from 4 to 2.1, we can match ?it to the dog, and resolve "it".

The point of this chapter is to suggest, as a strategy for language understanding that the application of inference rules to tie (problem) variables be guided, if possible, by such heuristic considerations.
This does not mean that we never seek a causal chain to explain human-like behaviour, only that we prefer a reason (goal) chain if we can find one.

Thus, to make a crude contrast that will be dealt with more fully in section C, it would be wrong to try in (21) to expand cause chains leading to John's action — as do those understanding systems that would immediately begin to make "enablement inferences" about his moving his hand to the door-knob, or whatever. That is just not how commonsense understands human actions.

One other point has surfaced in the discussion of (21) that I will draw attention to here. Two things might be said about the slightly sleight-of-hand way the pronoun was resolved in that sentence:

(i) that the upshot said little more than does the "focus of attention" explanation of why "it" refers to the dog, rather than the door. I accept this, and claim that rules generalising the treatment of (21) make the notion of focus in action sentences more precise.

(ii) it might be said that matching 2.1 and 4. requires more low level real world knowledge tha , for example, "being outside makes you muddy", possibly installed in a KEEP-DOGS-CLEAN "frame" (Minsky 1975).

This I very much doubt, even if such a frame could be written and accessed. I discuss this question in far greater detail in chapter VII below, but I could go some way with this point and connect 4. and 2.1. by an inference rule such as:

\[ \text{[I be+in bad+state]} \rightarrow \text{[I be outside+living+space]} \]

Given suitable matching functions at the nodes of this template matching rule, it would serve to tie 4. and 2.1, given a sub-rule about simultaneous negation of both central (action) nodes not changing the effect of the rule. However, this piece of information seems to me an unnecessary complication, for the example would be understood in the same way with a dummy buzz-word in place of "muddy" — the function of the goal chain imposes sense on what is at "muddy", rather than vice-versa. Failure to recognize this essential feature of human understanding has been common in much AI work in this area.
C. A little light machinery

First in this section I propose two changes in the existing preference semantics scheme that will remedy the defects noted above and allow implementation of the "GOAL-preference for actions" heuristic of section B.

(1) In future actions of a wide class, on extraction, will go into what could be thought of as the Dowty-form (1972), rather than the McCawley-form, (which is in effect what we have at present — though it was not designed as such). Let us define as causal actions all those actions whose heads of formulas are CAUSE, CHANGE, FORCE, ASK, GIVE, GET, LET, MAKE, PICK, PLEASE, STRIK, USE, or WRAP. Instead of leaving the source template intact (the one containing a formula headed as above) and tying it to an extraction by an unlabelled tie, we now leave the source template intact, but produce two causally related extractions of which the first contains the dummy action DO (this corresponds to Dowly 1970). We then, before passing on, seek to replace the dummy template extraction if we can.

Thus, suppose we again have (9) "John shut the door" where "shut" has a formula with head CAUSE. We had, on extraction, as part of (22):

1. \[\text{[John shut door]}\]...
1.1 \[\text{[[door be shut]]}\]

We would now have

(28) 1. \[\text{[John shut door]}\]
1.1 \[\text{[[John DO X]]}\]
1.2 \[\text{[[door be shut]]}\]

In the case of the "car buying" example (13) the dummy could be filled immediately as we could go from:
because the dummy in 1.1 has been immediately replaced by another extraction, 1.4, on the basis of a general rule: "if the WAY (manner) of an extracted action (i.e. No. 1.4) has itself an action head (i.e. GIVE for "paid") then the WAY-based extraction can be the cause of any cause-tied extraction (i.e. No. 1.2 in (29)) coming from the same source template (i.e. No. 1)".

All this can be done in a general manner by the extractor program for causal actions (of which "bought" is one since its formula head is GET in (14)), without calling in a wider pool of "template \(\rightarrow\) template" inference rules.

Suggestion (i) then can be paraphrased as: "put CAUSE-extractions into a Dowty-dummy form, and then seek immediately to identify the dummy template with a sibling $^\$-extraction from the same parent, or source, template."

$^\$ The dummy identification search must in general be wider than sibling templates. In:

(31) John threw the lever to open the door and went in

a Dowty-dummy extraction from

\[\text{[John open door]}\]

will give a dummy template

\[\text{[John DO X ]}\]

as a CAUSE, and this should clearly be replaced by

\[\text{[John threw the+lever]}\]

which is not its sibling.
(ii) The second technical change is as follows: \langle template \rangle \rightarrow \langle template \rangle inference rules will now be classified functionally into GOAL, CAUSE, and IMPLIC rules, whereas until now they have all been pattern-seeking rules of a single type. As we shall see "functionally" here will usually mean only that we are dealing with different directional ties of the same rule.

Thus, in the present scheme of things, template-to-template inference rules equivalent to the following have not been distinguished as types of inference:

(32) "If A is part of B, look for B not being part of A".

(33) "If A strikes B, look for B falling down in some way".

(34) "If A wants some B, look for A seeking to obtain B".

Yet it is clear that (32) above has a wholly logical, non-causal, content of "A strikes B causes B to fall". (34), however, given our above inclusion of wantings as trivial reasons, rather than causes, has a GOAL content only.

The important point here for what follows is that (33) can equally well function in a GOAL or CAUSE chain, depending on how we view the directionality of the rule. For example, suppose we write a version of (33) as (in semi-English, given already as (18)):

(35) \[
\begin{array}{llll}
\text{animate} & 1 & \text{strike} & \text{real} & 2 \\
\rightarrow & & \begin{array}{ll}
2 & \text{fall} \\
\end{array}
\end{array}
\]

and we now consider the sentences

(36) John hit Bill and he fell.

(37) John hit Bill so that he would fall.

(38) becomes a representation of the form:

(38) \[
\begin{array}{llll}
\text{John} & \text{hit} & \text{Bill} \\
?\text{he} & \text{fell} \\
\end{array}
\]

which, on trying with the rule (35) above (considered as a CAUSE-labelled rule) becomes:
(39) \[ \begin{array}{c}
  \text{John} \\
  \text{hit} \\
  \text{Bill} \\
\end{array} \quad \Downarrow \quad C \\
\begin{array}{c}
  \text{Bill} \\
  \text{fell} \\
\end{array} \]

whereas (37) becomes:

(40) \[ \begin{array}{c}
  \text{John} \\
  \text{hit} \\
  \text{Bill} \\
\end{array} \quad \Downarrow \quad G \\
\begin{array}{c}
  \text{?he} \\
  \text{fell} \\
\end{array} \]

and then, with the same rule (35) cued in by "so that" as a GOAL-labelled rule:

(41) \[ \begin{array}{c}
  \text{John} \\
  \text{hit} \\
  \text{Bill} \\
\end{array} \quad \Downarrow \quad G \\
\begin{array}{c}
  \text{Bill} \\
  \text{fell} \\
\end{array} \]

The point here is clear, even though the notion of cueing-in requires more analysis: the same inference rule resolves the pronoun in both sentences, but in one example establishes a GOAL link, but in the other a CAUSE link. This fact is not surprising, because, of course, causal relations function normally in human plans. Indeed, we cannot be sure that there are any causal sequences that cannot function in GOAL links, except possibly such remote and immutable "facts" as that "the rotation of our galaxy causes the sun to move in space, with respect to the centre of the galaxy".

What is the actual difference in the examples (38) and (40) that results in the chains being labelled differently in the two cases?

I wrote that the GOAL-labelling of the representation of (37) was cued in by "so that", while "and so" or "and" in (36) is a normal cue for a CAUSE link. It might seem less redundant always to label links resulting from the application of the same rule in the same way. Following this suggestion through we might, by analogy with the earlier suggestion to prefer GOAL chain explanations of human behaviour to CAUSE chains, correspondingly prefer GOAL developments of human behaviour (John's in this case) to CAUSE-developments.

This is different from the earlier suggestion, which would only have preferred a G-link going into

\[ \begin{array}{c}
  \text{John} \\
  \text{hit} \\
  \text{Bill} \\
\end{array} \]
to a C-link. Here we are contrasting a G-link going into that template with a C-link going out of it, and so a causal result of John's action and not a causal explanation of it.

The latter suggestion would lead us to insert a \( \rightarrow \) link into (38) also, and infer that John's purpose was to knock Bill down. But this seems unfair to John's behaviour, until we know more. In (36) we have no reason to think that John's purpose or goal was to knock Bill down.

If we decline then to prefer GOAL to CAUSE developments of human action, and leave representations (39) and (41) as they are, we shall be left with what we might call a semi-taxonomy of rules. Some rules are purely IMPLIC, some (like the "want" rule) apparently purely GOAL, but the majority are GOAL or CAUSE depending on the direction in which they are applied.

This notion of "direction of application" clearly depends on another notion, that of the point of entry for link and chain constructions, and so on — a general procedural implementation of the notions discussed so far.

**How to implement the above**

The above suggestions do not require any general reorganization of the battery of template-to-template inference rules, either into larger scale structures (i.e. "frames"), nor into more exclusive classes of rules than IMPLIC and CAUSE/GOAL. In the present version of the system the rules are all stored under the action head element of both the left and right hand side of the rule. Thus for example, the rule

\[
(42) \quad ((\text{ANI 1}) \ \text{WANT} \ \text{*REALOBJ 2}) \rightarrow (\text{1 HAVE 2})
\]

would be stored under both elements HAVE and WANT. The functional distinction of the last section is then obtained by conventionally associating the storage under one direction as the GOAL direction, and, if it exists, storage under the other as the CAUSE direction of application of the rule.
Let us now see what this might do for us in practice by first summarising the strategy of the existing system and contrasting it with a strategy containing the proposals made so far.

At present the procedures called after encountering a problem pronoun, not resolvable by straightforward preference considerations, (see Wilks 1973) are:

(43)

i. EXTRACT all formulas in the paragraph of templates to produce template-like forms as part of the representation.

ii. Try to match directly from a problem template to a solution template, without inference rules (=Zero point strategy).

iii. Try all length one inference chains tying variables from a problem to a solution-template and vice-versa.

iv. Repeat last step for chain length 2

and so on.
We now have, following the suggestions above,

1. Extract all case ties as before (including goal case, and goal and cause ties due to surface cues like "so that", and indicated by curved lines in diagrams).

2. Extract all causal type actions and put into "Dowty-form-plus-cause-ties".

3. Attempt to identify template pairs to avoid redundancy, and assert the consequent cause ties.

4. As before apply zero point strategy if possible.

5. Identify entry points.

6. Attempt to construct goal links to entry points, if the latter are human actions: if they are non-human events attempt to construct cause links to them. In both cases the links should attempt to join only problem and solution points. These links are indicated by square (not curved) lines on diagrams.

7. If this fails try to construct cause links to the entry point if it is a human action: or if it is a non-human event attempt cause links from the event.
Notice in (44) that stage vii coming before stage vi constitutes the preference for GOAL (purpose) chain explanations of human action rather than CAUSE ones. Chains of CAUSE consequences of human action have not been left out of the picture, as might appear at first sight, because the duality of CAUSE and GOAL rules has ensured that the exploration of GOAL chains to human actions is simply the dual (reverse direction) of CAUSE chains from the same actions. Examples should now clarify that, as well as the notion of "entry point".

Let us return to our earlier example (12) "John bought a car in the market and liked it immediately", represented by the new strategy as discussed earlier, up to the entry point stage. We left the example represented as:

(45)

1. \[ \text{[John bought a+car]} \]

1.1 \[ \text{[[John paid money]]} \]

1.2 \[ \text{[[John has car]]} \]

1.3 \[ \text{[[John use car]]} \]

2. \[ \text{[John in market]} \]

3. \[ \text{[John immed+liked ?it]} \]

Where we have inserted the case ties, the extracted CAUSE tie from 1.1 to 1.2, and seen that no zero point strategy, of direct matching, will take us from the problem point in 3 (where the pronoun is) to the solution points in Nos. 1, 1.2, 1.3, 2 (templates where one of the surface substitution possibilities is).

The entry point question is as follows: what actions or events are those from which we should begin to construct GOAL or CAUSE chains? So, for example, in the (45) above, both templates No. 1 and 3 are human actions that could serve as entry points from which to try and start GOAL and then CAUSE chains.
Those who believe in systems that make inferences "forward", that is to say before encountering representational problems, have an easy answer here: they simply declare that representations of events should give rise to an unspecified number of consequential inferences, so that, in the terms in which I have stated the problem, every action or event representation is an entry point and so the question does not arise. I believe that this is a counsel of despair, and that indeed that way madness lies. It is certainly no substitute for intelligent heuristics as to how a system should proceed.

The system, in its existing mode, attempts to construct inference chains from problem templates to solution templates and vice versa, and it seems obvious that the entry point should be initially restricted to one of those. In the case of the above example the only templates that are either solution or problem templates are Nos. 1, 1.2, 1.3, 2 and 3. It seems fairly natural to propose the following three sub-rules for choosing entry points:

(46) R1: try problem and solution templates before others

(47) R2: try templates that represent events, actions and states (defined by head element of main action), before others, such as No. 2 that have a case head.

$\text{such as Charniak (1975), Schank (1975) and Rieger (1975). The dispute between "problem driven" and "data driven" inferencing systems has many of the characteristics of a pseudo-issue, and in any functioning system to analyse natural language, the above writers would almost certainly in practice have to make some decision about an entry point if their systems were not to drown in their own inferences. Conversely, the present system could easily become a forward inferencing system, since the extractions — which are now only done if a problem pronoun presents itself — could easily be done even without the appearance of problems. Indeed, since the problem with the "data-driven" position has always been "where do I cut off the inferencing", this would be a positive suggestion to those holding the "data-driven" position, since the extractions, in my view, would provide a natural level of such "forward inferencing" without adopting the ad hoc numerical cut off of Rieger's system, for example.}
(48) R3: try those earlier in representation before those later (unless reversed by a surface cue such as "before that")

R1, R2, R3 are meant to be in the order stated, i.e. R1 is the principal rule and R2 is not consulted until after R1, etc. All that they settle is where to start building GOAL or CAUSE chains from. In the case of representation (45), they tell us to start with template No. 1, which leads us to construct inference chains in the manner defined earlier, but subject to the following suggested constraints of preference expressing vi and vii of (44):

(49) R4: if the entry point is an action, begin by trying inference rules yielding GOAL links into that template. If entry point is an event or an apparent action by "natural forces" etc. seek inferences yielding CAUSE links leading into that template.

(50) R5: if R4 above fails and it is an action, seek CAUSE links leading into the entry point.

(51) R6: if R4 and R5 fail seek CAUSE links leading from the entry point.

Because of the duality of many CAUSE and GOAL rules, for human actions, the goal seeking of R4 will in fact have covered CAUSE productions from the action under R6, and the system must be blocked from repeating those rules that have already been tried in the reverse direction.

Let us conclude the car example (13) with the aid of these suggested heuristics. Template No. 1 is indeed an action (i.e. by a human or human like entity) so R4 will apply and the system will begin to apply GOAL labelled rules forward from No. 1 through the representation. We can well imagine a rule:

(52) ((*ANI 1) GET (*REAL 2)) \[\in\] (I HAVE 2)

that would match direct from Nos. 1 to 1.2, since the head of "bought" will be GET, all the variables in the rule would be satisfied. But of course this rule will not be called, because the overall purpose of the whole business is to solve pronoun problems (and more generally sense resolution etc.) and No. 1.2 poses no problem, hence at this stage at least
no GOAL ties will be put in by the system. I shall return to this point later where contrasting problem-solving with what I shall call more meta-physical approaches to representation and understanding. However, for those who believe that all possible warranted connections should be inserted into a representation, then the above rule could indeed tie No. 1 to No. 1.3 by a GOAL link on the ground that John might well have bought the car in order to have it. It may well be that I am in a weak position in declining to do this, since I have already inserted a CAUSE tie l.1 to l.2 in the extraction phase. However the difference is that the extraction phase has a well-motivated cut off, whereas to extend the present inference procedure so as to make all possible ties seems to me to have no such similar cut off built into it.

To continue, the tie l. to l.2 is not made by the above rule and the system continues until the rule

\[(53) \quad ((\text{*ANI } 1) \text{ GET } \text{ (\text{*REAL } 2)}) \quad \quad G \quad \quad (1 \text{ *JUDGE } 2)\]

ties No. 1 to No. 3, which latter is of course the problem point and so this link is asserted, and a (square) GOAL tie is put in upwards from No. 3 to No. 1, and in doing so "it" is correctly resolved to the car rather than the market. Heavy weather of a simple example you may feel, but the point made in doing it is, I believe, general and extensible. The last rule has a GOAL directionality from right to left because it was decided to make actions of liking, wanting, etc. (here grouped under the class \text{*JUDGE, of judging actions}) GOAL rather than CAUSE actions.

Notice in passing that if R4 had failed we could not have applied any CAUSE direction arrow under R5 (except in the presence of special surface cues) from No. 3 to No. 1, because causes do not precede their effects. And that would have been so, even if (53) had been a dual rule. It is not dual because we decided that wantings were reasons for what was wanted, and not in any sense their causes. However, (52) is a good candidate as a dual rule, and should be written

\[(54) \quad ((\text{*ANI } 1) \text{ GET } \text{ (\text{*REAL } 2)}) \quad \quad G \quad \quad (1 \text{ HAVE } 2)\]
since the getting may be said to cause the having, just as the having is a reason for the getting. If No. 1.2 had been a problem template, and (54) had asserted a GOAL tie from 1.2 to 1, this would have been in itself, by the rule duality, an expression of the CAUSE consequence from No. 1 to 1.2, and would not need labelling in the other direction as well by R6. However, because R6 only applies if R5 fails, there is no need to explicitly exclude repetition due to rule duality, because if a GOAL tie between two templates is found, (i.e. R5 is satisfied) a CAUSE one is never sought.

Further applications to the strategy and some further modifications to the heuristics.

I have concentrated on a single case in detail, but there are of course enormous numbers of cases not treated so far: in particular the development of causally-explained action in humans and the development of causally explained events in the inanimate world. At least one other modification to the heuristics must be explained too, what I shall call cue modification (I have hinted at the importance of this). I shall now introduce a number of examples in order to illustrate these points.

Let us consider

(55) John shut the doors wearing gloves so that they wouldn't touch the handle.

This example may be a little forced, and I will not discuss its representation, but even so it will serve for a couple of points. We need to resolve "they", and someone might perhaps claim that the pronoun refers to John's fingers, and that identification with neither "gloves" nor "doors" will do. If that is so, and some writers do argue for such kinds of cause, then one would not expect any amount of GOAL straightforward linking from the main action of shutting the door to resolve the pronoun. That is, GOAL chaining under R4 would fail and so we would have to seek CAUSE rules leading to the main action of shutting, and these would tie such extractions as

[[ John move fingers ]]

to the main action, as causes.
Examples of this sort have a forced quality, but many seem to feel that causal explanation of normal human action has a fundamental role (see Schank 1974, for example). More plausible is

(56) John's cold came during an exam and it made him drowsy where it is clear that a CAUSE explanation of his behaviour is indeed what is required in default, I would maintain, of a GOAL one.

Default is not required in the case of non-human activity and in

(57) The rock fell off the cliff and crushed John's lunch. Peter pushed it

it is clear that the "it" is the rock, not the cliff or the lunch, and that a causal link leading to the main event of falling will settle the matter. The rule in question will be of the general form

(58) ((*ANI 1) (MOVE CAUSE) (*REAL 2)) \(\rightarrow\) (2 MOVE ---)

which is clearly a dual GOAL & CAUSE rule of the sort described earlier. Then, ignoring extractions for this example, the representation after the application of the causal rule (left to right) (58) and the resolution of the corresponding pronoun "it", will be:

(59)

1. \[ \text{the+rock} \text{ fall } - \]
2. \[ \text{rock} \text{ off } \text{cliff} \]
3. \[ \text{rock} \text{ crushed } \text{John's+lunch} \]
4. \[ \text{Peter} \text{ pushed } \text{rock} \]

Template No. 4 will be tied to No. 1 by the rule when No. 1 is the entry point (under R1, 2 & 3), and under R4 we shall seek CAUSE rules leading to the entry point No. 1. The first success will reasonably enough be the one tying the problem template No. 4

\[ \text{Peter pushed ?it} \]

Notice that on this strategy none of the above would involve the assertion of a CAUSE link in the representation from

No. 1 \[ \text{the+rock} \text{ fall } - \]

to

No. 3 \[ \text{rock} \text{ crush } \text{John's+lunch} \]
Another important point should be noticed about this example. Earlier a modification was suggested to R5: namely that CAUSE rules should not run backwards in time, as it were, "up the representation". But that modification itself must clearly be subject to (in the present case) the past tense "pushed", represented in fact in No. 4 by a tense semantic element PASA (past active) which is made part of (the leftmost element) the formula for "push".

Adequate investigation of the role of "surface cues" as modifiers of the suggested heuristic rules R1 to R6 would require a separate paper. But one form of modification must be mentioned. Let us consider:

(60) The lightning startled John so that he dropped the stone he was carrying on his foot and crushed it.

In the course of identifying "it" with the foot rather than the stone we would almost certainly want to use the template for "(so that) John dropped the stone" as an entry point for inference. Thus, under R4 we would first seek GOAL links to it, which would be inappropriate because, although a template for someone dropping something is considered an action, in this example the "so that" makes clear that a CAUSE link is being asserted by the sentence into the template for the "dropping". Thus we might say that R4 must be modified by a GOAL-blocking consideration, in this case imposed by a surface cue. The alternative here is to let R6 apply (since R6, about CAUSES, has been asserted by the sentence and so blocked R4, about GOALS) and to seek rules linking causes of the dropping. A more extensive override-by-cue can be seen in:

(61) John fell from the cliff above his dog and crushed it.

Let us consider what inferential processes would be called in here. Since

\[
\begin{array}{c}
\text{John} \\
\text{fell} \\
\text{...}
\end{array}
\]

counts as an action template under the definition of R2, the point at issue will be what CAUSE or GOAL rules will properly link that template with the last one:

\[
\begin{array}{c}
\text{?someone} \\
\text{crushed} \\
\text{?it}
\end{array}
\]
Under R4 we should look initially for GOAL chains from the first template, and failing that CAUSE chains (under R5) leading from what caused John to fall to his falling. Neither of these strategies would produce any result here, in which case the system would operate R6 and seek causal consequences of John falling. In other words, the possible ties corresponding to R4, R5 and R6 are respectively:

Clearly the last is the one appropriate to the representation of (61) though the one under R4 would be appropriate to:

(62) John fell from the cliff above his dog so as to crush it.

Let us consider in a little more detail what the relevant rule is, how exactly and when it is applied and what relation this process has to the appearance of surface cues, such as "and" in (61) and "so as to" in (62).

At an earlier stage of this paper it was argued that surface GOAL and CAUSE cues result in the assertion of (curved line, like case ties) CAUSE and GOAL ties, at the same stage as the case ties themselves are asserted. These ties are unlike the (square line) inferential ties in that they do not result in the tying of variables. They remain part of the representation, and do not require reassertion by inference rules, though they may guide their application. Thus, in the present example (61), we would have expected the representation before the application of inference rules to have been (ignoring extractions):
where the curved tie No. 1 to No. 4 has been inserted in virtue of the "and" surface cue. (This itself is no trivial task because "and" may be merely logical conjunction, or mere consequence).

Now, let us state the relevant inference rules, in semi-English form:

(64)  [animate 1 fall ----] → C [1 break real 2]  
      → C [1 hurt ----]

where this format indicates that the left-hand side is not repeated in the second rule, so that we create a rule stack with identical left-hand sides. These rules are not dual (i.e. are marked left-right as CAUSE only) since we do not normally think of people falling so as to satisfy some goal purpose, though as we shall see, they may do so.

Now, back to (61), represented as above. R4 first attempts GOAL ties to No. 1 and fails because rules like (64) and (65) do not have a GOAL dual direction. Similarly R5 then fails to tie a CAUSE link up to No. 1, since there are no templates in the representation that will match with any rule asserting what caused John to fall, as there were when the text contained "Peter pushed him". Finally under R6, the rule (64) ties No. 1 to No. 4, and so reasserts the (curved) cue tie, by tying variables and solving the ?it. In pictorial terms, the representation above will then be changed from

\[ C \quad \text{to} \quad C \]
Now consider (62), where a cue tie has inserted a tie up from No. 4 to No. 1. As with (61) above R4 and R5 fail to assert a tie to No. 1 upwards in (63) because (64) and (65) are not dual. Eventually the same consequential cause tie is found under R6 as was found for (61) and the ?it is solved in the same way so that the final representation has the form:

(66)

1. \[ \text{John fell } \]

\[ \vdots \]

4. \[ \text{John crushed dog} \]

where G is the tie cued in by "so that" in (62). In other words the duality of the tie between No. 1 and No. 4 is explicitly asserted, but one part of the dual comes from the inference rule and one from the surface cue. The fact that such actions can in fact be done in furtherance of a GOAL is not stated by the rule.

Now consider (67).

(67) John dropped a rock from the cliff above his dog and crushed it

and we may add into consideration the rules:

(68) \[ \text{animate 1 drop real 2} \]

\[ \text{animate 3 hurt } \]

\[ G \rightarrow C \]

(69) \[ G \leftarrow C \]

\[ 2 \text{ break } \]

(70) \[ 2 \text{ break real 3} \]

Here the surface "and" will already have asserted a CAUSE tie from No. 1 to No. 4 (where the templates correspond to the earlier representations of the similar sub-sentences in (67)), but these rules are dual and so a GOAL tie will immediately be made from No. 4 to No. 1 under R4, the GOAL seeker, so that the final representation will be the inverse of that for (62) namely:

1. \[ \text{John dropped rock} \]

\[ \vdots \]

4. \[ \text{John crushed dog} \]
A point of interest here is that none of this differential inferential analysis of examples requires any explicit taxonomy of actions: that is to say actions like "drop" and pseudo-actions like "fall" do not have to be classified separately for treatment, for the difference between them will be reflected in the writing of the action-seeking functions in the inference rules. As we saw, the rule that matched onto "drop" (and which would have a main-action LET) was marked with a GOAL direction, whereas, by contrast, the rule matching onto "fall", with a main-action MOVE, had only a CAUSE directionality.

Are there decisive examples?

The hypothesis of this paper, as should be clear by now, is that we do not need any taxonomy of CAUSES beyond a distinction between CAUSE and GOAL rules, and that even that distinction is largely functional, and so not taxonomic in a true sense, because the CAUSE/GOAL distinction reduces in many cases to no more than the directionality of a rule. However, that directionality pays its way via the hypothesis that, in the case of human and human-like actions we should first explore the GOAL "explanation" of an action, whereas with non-human events we should first explore the CAUSE "explanation" of the event. The assumption behind the hypothesis is that acting in this way will not only be more consistent with our common-sense intuitions than any alternative but that it will also lead us to the solution of concrete sense and reference problems more efficiently.

The trouble with the hypothesis is that, like most hypotheses in this field, it would be difficult to test adequately, and its value will be determined by a number of informal criteria rather than a decisive test. Almost any real example/sentence to test it will be highly delicate, in that the shift of a single word would change everything. Suppose we set out to produce an example to test the hypothesis as follows:

(71) John visited his aunt. He had a cold at the time, and he hated her. It made him very rude to her.

Is "it" the cold or his dislike of his aunt? On the present hypothesis we ought to begin to look for GOAL-like explanations, such as hatred
of her, before *cause* ones such as his cold. The trouble with the example, which would make its analysis highly arbitrary on such a simple-minded set of assumptions, is that the normal understanding of the example is so dependent on the order in which things are stated: as can be seen by reversing the order of introduction of the cold and his dislike of the aunt. Juggling a little with the example will show that it is dependent on such considerations as that one does not normally have a cold all the time and so on. Like all real examples in natural language, it is highly delicate, even when made up explicitly to test a point. Suppose we try

(72) John's wounds hurt him as he threw the grenade at the snipers to dislodge them

 contrasted with

(73) John's wounds hurt him as he dug at the shrapnel splinters in his leg to dislodge them

or with

(74) John's wounds hurt him and as he bathed his feet he cried out with the pain of them.

We might well expect to reach "them" as "snipers" in (72) and as "splinters" in (73) by means of *goal* links, and to reach "wounds" in (74) via *cause* links, by default. But then, of course, if we had had

(75) John's wounds hurt him, and as he bathed his shattered feet he cried out with the pain of them.

Then, using the same techniques as (73) would again lead us to "wounds", even though the addition of "shattered" has plausibly made the right answer "feet" in (75) in a way very hard to specify formally. All this suggests that *decisive* examples for the hypothesis suggested here would be hard to find, though in no way impossible.
The main consideration throughout the chapter has been an appeal to the solution of concrete problems of language understanding, rather than to an undefined notion of "understanding". To that end, the only distinction suggested here (cause vs. reason) has been explicated in terms of a procedural difference, so as to void an idle taxonomy.

At various points in the chapter I indicated that a causal connection could be inserted into the representation at that point even though it solved no problem. Again, the question might arise as to whether or not the cause and goal ties asserted by the inference rules in the way described in the chapter should remain in the representation after they have tied variables satisfactorily. In some sense their function is exhausted by the tying of variables and after that their remaining is purely cosmetic. I am agnostic on the point.

A more interesting question is that of "causal chains". Should we expect the CAUSE and GOAL links inserted into a representation to form chains of causal links or not. Here, it seems to me, is a classic case of the opposition of procedural to metaphysical requirements. It seems clear even from the few examples discussed in this paper that the inferences required to solve problems of language analysis do not, in any convenient manner, form themselves up into causal chains extending forward in time.

Yet educated commonsense, though not, I suggest, naive commonsense, has come to believe (via 17th Century metaphysics filtered down to our own times) that there really are causal chains and that semantic representations should contain these. And this is believed even by researchers who would not be caught dead reading a book on 17th Century rationalism.

None of this is to deny that stories could be written that seemed to have a causal chain running through them, such as classics of the form "John left a nail out of the shoe of his horse, which caused it to stumble on the battle field at a crucial moment so that the front line fell back, so that ..." and so on. Of course, stories can always be written to demonstrate any view whatever, including that of events proceeding their causes no doubt. My point is that that is not the normal scheme of things,
and there is no reason to suppose that our everyday notion of causality for language understanding requires that the causal links inserted be chained up wherever possible above and beyond the solving of particular problems of analysis.

However, that is not to deny that some larger G and C linkages might be of interest when inferred into representations. For example, given any three templates with two CAUSE and one GOAL link as shown, the dotted GOAL link might be inserted, in line with the old principle of section (d): "Having X as a GOAL does entail having Y as a GOAL where Y is a normal CAUSE of X", thus:

(76)

A more difficult case, perhaps, is where we have an inferred causal consequence of an asserted purposed event, thus:

(77)

I suggested earlier, in section (d), that we most likely would not want to assume that persons having X as a GOAL also have Y as a GOAL, where Y is a causal consequence of X. But, as the law generally holds, if Y is a normal enough consequence, we might well want to assume they did. Consider the difficulty of establishing in court that one only shot to wound, when death is such a normal consequence of being shot. Whether or not we would want to insert the dotted GOAL link in (77) seems to me an interesting and open question.
D. Some comparisons

Some obvious questions of comparison arise between the suggestions of this chapter and its predecessors, and positions established by Charniak, Schank and Rieger.

In a clear sense this chapter is orthogonal, as it were, to Charniak’s work. In Charniak (1972) he did not advocate any strong taxonomy of causes and his general demon strategy is not essentially different from that of Wilks (1973 and 1975d) of which the present chapter is a refinement.

His more recent work on a taxonomy of what he calls "prerequisites" is highly interesting but, as I said, orthogonal to the present chapter because prerequisites are not essentially causal. What he calls "strict prerequisites" are probably equivalent to the (undiscussed) Implicit inferences mentioned here, but his other prerequisites cover much of the wide open spaces of presupposition, and may indeed go a long way towards clarifying in procedural terms a notion that has become progressively vaguer in the hands of the linguists. A point where I would differ from him would be in the wide area he calls "social prerequisites" which could be construed as causal inference rules in a wide sense. An example he gives is "One snaps one's fingers in order to call a waiter". My doubt about the function of this sort of rule, other than in the construction of a robot to call waiters of course, is whether or not a language understanding system would ever need a causal rule stated at this level of cultural detail. I would find it very hard to make up an example which depended on one's knowing this fact in order to resolve some problem in it. My suspicion is, and I develop this in chapter VII below, that any use of such a rule would be explicit in the surface text, as in: "John snapped his fingers but the waiter didn’t come", where we don't need the rule to understand it, but, on the contrary, as intelligent readers we can infer the rule from the example.

In Charniak (1975c) is an important argument which would probably be accepted by those, like Schank and Rieger, who do propose a wide taxonomy of causation: Schank's (1973c) proposals have essentially four types of cause (reason, result, enable, initiation) and any of these can be potential,
making eight in all; Rieger (1975) proposes a taxonomy of connections between events of which more than twenty are plausibly types of cause.

I will not discuss Charniak's argument here: it is essentially an argument that any psychologically justifiable inference system must make forward inferences ahead of need, that is to say, be "data" rather than "problem" driven. Charniak would argue therefore that a system that successfully solves linguistic problems is not enough. Apart from the truth or falsehood of the argument, it is clear that its acceptance justifies a wide taxonomy of the sort proposed by Schank and Rieger on the grounds that the taxonomy can limit forward inference, in that if sub-class of causes_1 are inferred, perhaps subclasses_2,...,n need not be.

If, on the other hand, one takes the view that forward inference should only be done on a very limited scale (such as defined by the extractions, for example) and is best avoided if possible$, then this justification of a cause taxonomy carries little weight.

It seems to me that their taxonomies (a) do not correlate with different procedures, i.e. pay their way, and (b) do not by their existence solve linguistic problems$$$. I admit this is a strong point in an area that is very much of the "fools rush in where angels etc.", but let me weaken it by saying that it does not seem clear to me that they have tried to show that their taxonomies are doing anything for them, and their justification is therefore very much an open matter. As I hinted earlier, my private suspicion is that such taxonomies are purely meta-physically motivated.

$ and especially in Schank's form, of drawing all the causal consequences of an action!

$$ Rieger's do not any longer seem even to claim to solve linguistic problems, since he is now (1975) interested in formalizing how people think that mechanisms like water closets work. But technical need is a long way from the needs dictated by commonsense understanding of language. Special forms of causality are needed to understand quantum physics, for example, which is little further from commonsense than the technicalities of plumbing.
Let me suggest a very tendentious table of desiderata and their presence or absence in the work of various people:

<table>
<thead>
<tr>
<th>DESIDERATUM</th>
<th>EC</th>
<th>CR</th>
<th>RS</th>
<th>YW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there a representation for sentences</td>
<td>X</td>
<td>n.a.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Is there a procedural inference scheme</td>
<td>Y</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>Is it problem or data driven</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>P</td>
</tr>
<tr>
<td>Is its forward inferencing naturally limited</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>n.a.</td>
</tr>
<tr>
<td>Is it actually trying to do things in the world rather than analyse language</td>
<td>Y</td>
<td>Y</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Does it have preferential procedures explicating and corresponding to its CAUSE/GOAL distinction</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>Y</td>
</tr>
<tr>
<td>Has it a CAUSE taxonomy (beyond &quot;reasons&quot;)</td>
<td>X</td>
<td>Y</td>
<td>Y</td>
<td>X</td>
</tr>
<tr>
<td>Topic or primitive driven</td>
<td>T</td>
<td>?</td>
<td>was P</td>
<td>P</td>
</tr>
</tbody>
</table>

With this rather crude picture of one view of the state of play I rest my case, but of the main principle I remain convinced; we need to employ only taxonomies that are reflected by procedures that themselves function naturally within a system for assigning representations to pieces of natural language.
VII Frames, Scripts, Stories and Fantasies

Holmes: Watson, I think our plans are about to change. I expect a visitor in this room very soon.

Watson: Who is it to be, Holmes?

H: You shall see, but let me tell you now to expect a pronounced American accent and a silver topped stick. Our visitor will beg our pardon for intruding on us unannounced, but will assure us that he has something extraordinary to impart.

W: I find all this very hard to believe.

H: In addition, he will warn us that Mrs. Larrabee's fate depends on our cooperation and attention.

W: Look here, Holmes...

(there is a loud knock and a stranger enters without waiting. He has a silver topped stick and speaks with an American accent)

Stranger: You must be Mr. Sherlock Holmes.

(Holmes, who has moved to the mantleshelf; nods)

You must forgive this intrusion, gentlemen, but what I have to tell you is so astounding, so impossible of belief...

W: Holmes, will you see this man or shall I call...

H: Do not disturb yourself, my dear doctor.

S: Every moment is vital, 'Mr. Holmes, vital to Mrs. Larrabee, whom I think you know. Your attention to what I have to say may be instrumental in saving her life.

H: I think, sir, you had better take a seat.

Minsky's frame proposals (1974 & 1975) have stimulated enormous discussion, both within and without AI, and are producing an almost certainly helpful unification of insights. Many people have been caused to see that they have been constructing frames, unknowingly, for many years, while others have gone away disappointed that they have not been told precisely what a frame is. With an idea as general and abstract as this one, it is probably right not to press too soon for precision, but to allow the field to advance while sleepwalking, in the approved Koestler manner. In this chapter, I in part follow this trend, and produce a frame-like structure in my own system of semantic representation, partly because frames, like Everest, are there, and partly because flexibility is all in this field, and no one wants it to be thought that their system cannot adapt to some change in the rhythm of progress.

However, and as will become clear, the frame structure presented here is not wholly serious, for the main thrust of this chapter is a preliminary attempt to refine what are the precise claims being made in the application of frame analysis to natural language understanding.

I shall begin with a widely accepted distinction between static and dynamic themes in Minsky's paper; and, in the case of the latter, examine a strong form of claim associated with it, which I shall call the plot-line hypothesis. I shall compare this with what I shall call the do-it hypothesis. Both of these are, in a sense, frame hypotheses, and I shall argue that the latter, in its connection with natural language, is probably false, while the former is, in the Scots phrase, not proven.

Static and dynamic

I will not repeat the by now familiar Minsky quotation that gives his initial definition of frame: let us say just that in order to understand anything interesting, an AI system must have access to complex structures representing familiar knowledge. Minsky defines four types of frames, of which two will concern us: the scenario frame (or stereotyped event, such as a birthday party) and the communication event representation (concerning the way stories are told).
The former specifies that party dress should normally be of a special sort, that the present should please the host, and so on. These specifications are what Minsky calls defaults: that if we have no reason to think otherwise we assume in our representational structure that the "dress slot" will be filled by "best clothes".

The latter frame type concerns sequential events. It is clear that the former is more static in nature, and the latter more dynamic. I say "more", because as Charniak (Charniak & Wilks 1976) has pointed out, Minsky gives two accounts of a scenario frame for a birthday party, one more static in nature and the other more dynamic in that it is, in effect, a series of instructions for someone attending a birthday party.

The dynamic frame

Two recent examples of frame construction in connection with natural language understanding are Schank's "scripts" (1975) and Charniak's frames (1975a). Let me outline them very briefly.

Schank defines a script as "a predetermined causal chain of conceptualizations that describe a normal sequence of things in a familiar situation" (ibid. p.117). He sketches a restaurant script, from the customer's point of view, as follows:

"script : restaurant
roles : customer; waitress; chef; cashier.
reason : to get food so as to go down in hunger and up in pleasure.

scene 1 entering

PTRANS - go into restaurant
MBUILD - find table
PTRANS - go to table
MOVE - sit down
scene 2 ordering

ATRANS - receive menu
ATTEND - look at it
MBUILD - decide on order
MTRANS - tell order to waitress."

and so on for (3) eating and (4) exiting. It should not be necessary to explain the associated primitive actions on the left hand side for the reader to get the general idea. Schank also has a program which will take a paragraph length restaurant story and produce a slightly longer story with the "missing bits" filled in from the script above.

Charniak has produced a frame for "normal shopping" in a supermarket as follows (ibid. p.47):

"a) Goal: SHOPPER owns PURCHASE-ITEMS.
b) SHOPPER decide if to use basket, if so set up cart carry FI.
c) SHOPPER obtain BASKET * cart-carry.
d) SHOPPER obtain PURCHASE-ITEMS.
e) method-suggested
f) do for all ITEM PURCHASE-ITEMS.
g) SHOPPER choose ITEM.
PURCHASE - ITEMS - DONE
h) SHOPPER at ITEM.
i) side-condition DONE at ITEM
   also
j) method-suggested
k) cart-carry (SHOPPER, BASKET, DONE, ITEM)
l) SHOPPER hold ITEM.
m) ITEM in BASKET *cart-carry".

and so on till the shopper exits from the supermarket at line (u). Again the general idea should be clear, even without description of what Charniak means by terms like FI (= frame image) in line (b).

Here is another one, more exotic perhaps, and so the "fantasy" of the chapter title.
script : male puberty rite.
roles : male child, village elder, helpers, crowd.
reason : placing ritual incisions on back of child.

a) Goal: CHILD is tattooed.
b) HELPERS hold CHILD (by both arms).
c) ELDER obtain TOOLS.
d) ELDER exhorts CROWD (on proper behavior).
e) (general condition)
   $bad behavior by CROWD \rightarrow activity halted.
f) ELDER checks if CHILD properly purified.
g) (special condition)
   CHILD not purified \rightarrow activity halted.
h) ELDER mark CHILD's back
   method suggested
   \rightarrow do for all CUT-MARKS

and so on. Again the general idea is clear, and the choice of a remote, and imaginary, culture is not accidental, as I shall try to show.

I have written the above in a mixture of Schank's and Charniak's notation for ease of communication, but the same puberty frame could be expressed in terms of the template-matching inference rules of the last chapters, simply by allowing the rules to "stack up", as it were, in "sorites" form, not just as P \rightarrow Q but as P \rightarrow Q \rightarrow R \rightarrow S and so on. Thus:

b) [helpers hold child]
c) \rightarrow [elder get tools]
d) \rightarrow [elder exhort crowd]
e) [crowd $badly+behave - ] \rightarrow [action halt - ]
f) \rightarrow [elder check child]

and so on, where this is in short-hand readable form; though as before, the English words should in fact be replaced by functions seeking semi-formulas of primitives, so as to produce real rules matching onto a text representation. I shall return to this point later.
My worry about frames like the last is just what problem of language understanding they could or would solve. And I am taking it as axiomatic that there must be some concrete problem they do solve, such as word-sense or pronoun reference ambiguity, otherwise they are not essentially connected with the understanding of natural language. That is to say that representation of knowledge as such, and independent of any purpose or problem, is not a task for AI. Anyway, in so far as it is a task, we have a perfectly good language for doing it in, namely English, and have no need to construct semi-formulisms. This point, too, will surface again.

Suppose we have the "story sentence"

(1) Little Kimathi's mother accidentally touched his arm during the puberty rite. The crowd drew back in horror.

If we wish to "understand" this sentence, do we need the frame above to do it? The frame in some sense covers (1) with line (e), given some adequate definition list of bad behaviors in the frame (to which we are directed by "$", let us imagine).

And yet it is clear that we understand (1) perfectly well without the frame, just as we did the "waiter" prerequisite example in chapter V, and as Feldman (1975) argued that we would understand his unusual sentence about a Victorian house in India.

In commonsense terms, we could say that we infer from (1) that the mother touching Kimathi during the ceremony was a bad thing; we do not need it to understand.

In the terms I used to construct my frame, we could say that the two appropriate templates in the representation of (1) are to be tied by some 'template $\rightarrow$ template' inference rule of the general form

$$[	ext{human(s) display alarm}] \leftrightarrow [\text{other+human(s) performed bad+action}]$$

I think the above commonsense analysis applies to many of the current "John came home from the hospital. His face was sad" examples. An earth-frameless Martian would understand these so long as he understood the constituent words. Hence, the choice of a puberty-rite rather
than a supermarket for my frame, for we are Martians where puberty rites are concerned. **If we do understand** (1) it cannot be from our associated frame because, presumably, we do not have one.

I am not saying that frames of this sort do not have a function in language understanding, but only asking that their proponents clarify what claims they are making.

The plot-line hypothesis

One possible strong claim is as follows (the plot-line hypothesis): "in order to understand a story we need to know how **basic stories** (i.e. frames) of that type go; that is, we only understand, and can only understand a particular story by judging how it follows, or diverges from, the normal story of that type".

This thesis is almost explicit in Schank, and is, I think, being made use of implicitly by all those who now refer in this connection to Bartlett's work on memory. A better general reference might be (Trankell 1972, brought to my attention by Ken Colby), where there is some wonderful description of the sequential oral transmission of stories whose telling shifts, with the number of repetitions, towards a "normal story".

Linguistic evidence supporting this plot-line hypothesis (PLH), in connection with Charniak's frame, say, would be some story such as

(2) John went to get some things at the supermarket. He took the last basket that was left.

where it might be argued that we only understand "took" as "grasped" or "got hold of" rather than "took along with him" (and the difference might be crucial in some language) because we have access to the supermarket frame.

One very general, and not very persuasive, worry about the strong PLH above is that it might seem to lead to a regress: we only understand parts of stories because we know the normal story, but perhaps we only understand the normal story because we understand some normal, but larger, sequence, leading on up to the normal life. This regressed conclusion,
interestingly enough, is one that Dreyfus (1972) might be prepared to accept, because he believes that understanding is only possible within the context of a whole human life (which, of course, automata don't have).

But more plausible worries can be seen at less general levels, those of detailed organization. These are so enormous, and so little understood (though writers like Charniak emphasise them frequently) that they might outweigh all psychological evidence gleaned from Bartlett and Trankell, and even make it worth getting examples like (2) wrong if one could think of no "lower level rule" to get (2) right.

These organizational problems concern the swapping in of enormous frame structures, which to follow when a number competes, and when to abandon one. Consideration of a sentence like:

(3) John stopped off at the supermarket on his way to dine at the Orpheus restaurant. He couldn't get out of his mind the documentary on African puberty rites he had just seen on television.

Plausible enough, and already at least three huge frame structures have been brought in, plus perhaps one more for viewing television. These difficulties are obvious, and there is no point in dwelling on them, except perhaps to notice in passing that, when expounding frames originally, Minsky argued that existing structures were "too local". But there is clearly some sense in which, for everyday vague talk at least, the sorts of frames being proposed may be "too local" too.

Let me take a different tack at this point. One could argue that a strong consideration in favor of applying frames to natural language problems is that it might help with the problem of "topic", which is, very roughly, how can we get word senses right by knowing what area of discourse we are in. In Riesbeck's example "John went hunting and shot a buck", we feel we know "buck" means "deer" not "dollar" because we are talking about hunting, not money; or we could say, because we are in a hunting frame.
It is worth recalling here, momentarily, a hypothesis about topic-determination popular in the Fifties and Sixties, that the necessary tool was a thesaurus. A thesaurus is, as it were, a dictionary in reverse, in which words were classified under a thousand or so very general heads, each of which could be considered as defining an area of discourse. So, in the above example, looking up "hunt" and "buck" would reveal a list of heads for each, but it was hoped that the intersection of the list (in this case, say, some head for "pursuit of animals for sport") would reveal the appropriate area of discourse, and so the right word senses.

The trouble with the hypothesis was the same as the trouble with Riesbeck's example, namely sentences like

(4) John went hunting and lost fifty bucks

that I mentioned in chapter I, where it is plausibly money that is lost. One obvious worry here is whether a naive use of frames will do any better with (4) than a thesaurus.

What is needed is clear: an integration of mechanisms operating at different levels, and not only at some frame level, or what Sloman has called (1975) the way different levels "cooperate in disambiguating themselves".

Local preference considerations deal with (4) but perhaps fail with (Phil Hayes' example)

(5) John licked the gun all over and the stock tasted particularly good

if you feel that "stock" means "gun-part" and not "soup". For any particular example, like (5), solutions can be thought up, such as declaring that part-whole preferences, where a formula for "stock" told you it was a gun-part, should take preference over all others. But it is not clear that there can be a general solution, because there will be yet more ingenious (and intuitively peculiar sounding) exceptions to that declaration too.
Another, oblique, way of accepting the PLH is to simply ignore or discount divergences from the anticipated. This is profitable in CAI programs (which clearly have a frame) but less so in approaches like that of the Wayne State University Professor of Education who advises against correcting child readers who "read in" what they expect rather than what is on the page in front of them (San Francisco Chronicle, July 24, 1975).

"Thus in the example ['The boy jumped on the horse and rode off'] the student predicted on the basis of the words 'jumped on' that 'pony' might follow, and when he got to the verb 'rode' saw no reason to change it".

At this point, a defender of the PLH might object that he is not claiming that things always turn out as anticipated, but only that knowing the norm is essential to understanding the non-norm. Those who accept this argument at a lower level, as I would for preference considerations expressed earlier, therefore have to be clear why they might not want to push the "power of the norm" argument up this far, that is to say, as far as normal event sequences.

The do-it hypothesis

It is not easy to tease out the PLH from other arguments being used, implicitly or explicitly, to defend the use of frame structures. Among the explicit ones is Charniak's ingenious argument (1975a) for his frame structures concerned with the way in which they solve certain internal organizational difficulties he encountered with his earlier (1972) demon structures. That argument is beyond the scope of this chapter, and I shall turn to the implicit but pervasive hypothesis I shall call the do-it hypothesis which says, roughly speaking, that, in representing knowledge, we should concentrate on the representation of human activities that we know how to perform: stacking blocks, eating in restaurants, shopping in supermarkets, cooking food, or, in Rieber's case (1975), the slightly different matter of how a water closet performs its task. The relation to the PLH is clear: sequential tasks also have a plot line: first you do this, then you do that, and so on.
Here, very briefly stated, are three arguments for the relevance of the do-it hypothesis to natural language understanding:

1. that it is hard to express in natural language how we perform many ordinary physical activities: C. S. Lewis' favorite example to make the point was asking them to describe how a pair of scissors worked. Constructors of frames often refer to the difficulty of specifying simple human activities in detail.

But the mere fact that it is hard to specify such activities, whether in language or frame formalisms, in no way shows that such specifications are required for understanding language about the activity. On the contrary the much emphasised difficulty indicates distance between the activity and language rather than the reverse.

2. that one cannot understand language about human activities unless one has, or can, perform them. The core of this argument is rather like the one that young school children in class cannot understand the discussions of love in the Shakespeare plays they have to study because they have not experienced the corresponding feeling.

There is clearly something in this argument, but the trouble comes with specifying just what it is that the non-performer does not understand, given that one could answer questions as if one did understand. I can answer a lot of questions about skiing, and certainly appear to understand newspaper stories about it, even though I cannot ski.

3. that the relevance of an activity frame to natural language is bound up with the teachability of the activity.

This ingenious argument (put forward by Charniak) has the advantage of avoiding asserting the FLH directly, as Schank seems to do, but is weakened by the fact that the frame activities in question (eating in restaurants, shopping in supermarkets) are not, in any obvious sense, taught at all, certainly not in the sense that using scissors or tying one's shoe laces are taught: they are activities on too large a scale for that.
In conclusion, it seems to me that there is a false argument in
the air that goes as follows:

(A) We need representations of knowledge to understand natural
language.

(B) Two of the things we surely know are how to perform simple every-
day activities, and how normal stories go.

(C) Therefore, we need representations of those two forms of knowledge
to understand natural language.

The premises are surely true; the argument per se false, and the con-
clusion is, in my view, probably false as regards the do-it hypothesis,
and not proven as regards the PLE.

In the case of the PLE, simpler hypotheses, already in the field
and proposing more loosely organized inference rules and text structures
built up as we proceed (rather than imposed in advance), need to be
tested to destruction also, rather than being simply abandoned untested.

If, on the other hand, frame advocates take the other tack and just
give up the claim that frames are required to understand natural language
(and some are tempted) then natural language workers can just get on
with their work. But they may hang around long enough to go on asking
what then frames are for.

Many answers come to mind. One application to language that is
definite is the understanding of parody and satire. You cannot under-
stand a parody unless you know the normal story it is a parody of.

Another, linguistically related, and undoubtedly a frame-using
situation is when we are conversing in a language we do not understand
well. When we reason that X has just come back from a party and is
probably telling me about the wonderful girls he's met, so I'll try and
interpret what he's saying along those lines. However, this situation
is essentially one in which we don't really understand, and is not exten-
sible in any obvious way to ordinary language use where we do, on the
contrary, understand what is being said to us.
Again, it is argued that it would be quicker, or more efficient, to spot what is odd or important in a story, if we have the norm available. The efficiency argument depends on many formal parameters. The importance argument depends on the important being the novel, as when an unexpected bomb goes off at a birthday party. However, the important may also be coextensive with the expected item, as when the present (expected) turns out to be a huge diamond (unexpected). The oddness argument may be a little like the old detection of ungrammaticality — perhaps true but, so what? Some people claim to find "John went into the supermarket and went straight up to the cash desk" both a violation of the supermarket frame and odd. Even if that were so, what would have been achieved?

A clearly applicable frame context is that of genuinely stereotyped situations such as Bornat & Brady's (1975) program to read FORTRAN cards, where each (story) card really does end with END.

One clear function of do-it frames that is rarely embraced by their proponents is that of a robot-driver: the frame is a program that would, in principle, allow a robot to actually carry out the associated activity. It is not hard to see why those who use dining or shopping frames might not want to accept this explanation of what they are doing too easily. Rieger actually does so (1975), the only odd thing being that he also considers his water-closet commonsense algorithm relevant to natural language, though he does not attempt to justify this.

The last, and best known fall-back justification of frames is the task of question-answering, which is a strong position because there can be no doubt that, in order to answer questions about subject X, we need to know something about X. This, as I hinted in an earlier chapter, is as true of questions about angels as of those about the stock market. The only residual doubt is as to just what that has to do with natural language: since the same knowledge representation could be used to answer questions using only pre-formed whole sentences as in a box-ticking questionnaire, where no question of natural language understanding need arise. The only possible mismatch that could occur would
be because the frames were at the wrong level of description, in some sense, so that if someone asked "Does Thriftimart have Musak?" a frame that "drove you round" the market wouldn't necessarily help.

Some of those who assume that question-answering must be related to natural language understanding are making what I would call the expertise fallacy: that understanding natural language consists, in general, in retrieving expert knowledge about a subject matter. When one points out that most conversations are not like that, the reply can be (indeed, was, in one case) "Well, most people don't understand what they are talking about, they are just ad hoc ing along".

That simply cannot be true: there cannot be a main sense of "understand" which 95% of the population never attain, for that is not what the word means. It means what we all do most of the time. No physicist could tell us that we had most of us been wrong all the time about what "red" meant. My reply will be seen to be no more than an old weapon from the "ordinary language philosopher's" war chest. But it must be used.

The above, then, are undoubted and possible roles for the PLH. As regards natural language understanding itself, some form of the frame hypothesis, even the PLH, may well be true, both as regards humans and efficient computer systems. But I have argued that, in addition to the dreamy and seductive dashing off of example structures (all too easy, as I showed), we also need some harder thinking on just what claims we are making and testing.
References


Bar-Hillel, Y., A demonstration of the non-feasibility of fully automatic high quality machine translation, 1958 memorandum, (most easily available as Chap. 12 of Language and Information (Reading, Mass., 1964)


Charniak, E., Organization and Inference, in TINLAP, 1975a.

Charniak, E., A partial taxonomy of knowledge about actions, in Proc. 4th International Conf. on AI, Tbilissi, 1975b.

Charniak, E., Notes for the Tutorial on Computational Semantics (Inference and Knowledge), ISSCO, 1975c.


Clark, H., Bridging, in TINLAP.


Crowell, E., Causal explanation and human action, Mind, 1975.


Feldman, J., Bad-mouthing frames, in TINLAP.


Hewitt, C. H., 'Stereotypes', in TINLAP.


Marcus, M., Diagnosis as a notion of grammar, in TINLAP.


Rieger, C., The commonsense algorithm, in TINLAP.

Riesbeck, C., Computational Understanding, Memo from ISSCO No. 4, 1974.


Schank, R., Causality and reasoning, ISSCO Memo No. 1, 1973c.

Schank, R., Understanding paragraphs, ISSCO Memo No. 6, 1974.

Schank, R., Using knowledge to understand, in TINLAP.

Sloman, A., Object-frames, Domain frames and Analogical representation, (unpublished MSS, 1975)


Wilks, Y., An Intelligent Analyser and Understander for English, Comm. ACM, 1975a

Wilks, Y., Natural Language Systems with the AI Paradigm, Stanford University, AI Lab. Memo No. 337, 1975b.

Wilks, Y., Notes for the Tutorial on Computational Semantics (Philosophy of Language), ISSCO, Castagnola, 1975c.

Wilks, Y., A Preferential Pattern-matching Semantics for Natural Language Inference, Artificial Intelligence, 1975d.
