Frames for machine translation

This second of two articles on machine translation discusses a recent concept that may help the inanimate computer "understand" the often apparently irrational utterings of human language

Dr Yorick Wilks is reader in theoretical linguistics, University of Essex

Research in artificial intelligence (AI) is aimed at simulating unconscious human abilities. For instance, humans understand language—in the sense of our ability to answer questions, make inferences, translate to another language, and carry on a sensible conversation—without conscious effort. AI workers are designing and programming systems that can do one or more of these things, in the rather limited manner of replying to a human operator at a teletype or video-screen. A number of systems have been developed in recent years based on notions rather different from those to be found in contemporary linguistics, in that they all made the representation of common sense knowledge a central feature.

More recently, the structures of such systems have become more complex, under the influences of a proposal of Marvin Minsky that the knowledge structures in use in AI —and he was writing about machine vision as well, but here we shall concentrate only on language—should be higher-order structures that he called frames.

One can see the sort of thing he was getting at by considering the statement: "John went into a supermarket and put some soap in his basket. On impulse he put a bar of chocolate in his pocket as well, but when he reached the cash desk his face went red and he said 'I didn't mean to take it'."

The question that might come up in, say, machine translation, is how we know that the "it" refers to the chocolate, and not to the soap. The two words might have different genders in some output language, and so we would have to get the decision right, and in a general and plausible manner. It is easy to see that one might need to have access, even for this apparently simple task, to some complex formalised structure expressing what normally went on in a supermarket, so that one could infer from it that putting buyable items in one's pocket was not normal behaviour. Notice that it would have to be very specific information too, because it would not be enough to know that, in a supermarket, one normally puts buyables into a container, for a pocket is certainly a container. On so general a description of the activity of shopping the "abnormal" act would slip through unnoticed.

The notion of frames

It is just such highly complex and specific knowledge structures that Minsky argued should be called frames, which, in some formalised version, would be essential to any computerised language understanding system.

Let us begin with the standard quotation from Minsky that best captures the general notion of "frame": "A frame is a data-structure for representing a stereotype situation, like a certain kind of living room, or going to *a* children's birthday party. Attached to each frame are several kinds of information. Some of this is information about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.

"We can think of a frame as a network of nodes and relations. The top levels of a frame are fixed and represent things that are always true about the supposed situation. The lower levels have many terminals . . . 'slots' that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet..."

Under the influence of Minsky's proposal, Eugene Charniak produced a frame for shopping in a supermarket (to deal with examples like that about soap and chocolate), while Roger Schank produced similar structures but called them scripts. Schank defines a script as "a predetermined causal chain of conceptualisations that describe a normal sequence of things in a familiar situation", by which he means some account, capable of simple formalisation, of the normal order of events when visiting a restaurant. He sketches a restaurant script 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 scenes 3 eating and 4 exiting. For the reader to get the general idea, we need not go into the precise definitions of the associated *primitive actions:* entities like PTRANS on the left-hand side—this one indicating physical movement—that Schank uses in his underlying semantic conceptualisations of sentences in the computer. Schank's students have written a program which will take a paragraph-length restaurant story and produce a longer story with the "missing parts" filled in from the script above; and will do this in a number of output languages, thus producing a rather new definition of machine translation.

The question that is being asked at the moment is what exactly frames are for in language-understanding systems; what hypothesis their use implicitly appeals to; and whether the benefit they confer could be obtained by other simpler means. There is no doubt they express the dynamic order of events that is part of the meaning of certain concepts, in some intuitive sense.

Moreover, the frame is potentially a powerful device for defining_topic context, a problem that has plagued all formal work with language since the earliest machine translation. So, for example, if we see the sentence "John ordered an omelette", we know that it is the "ordering food" sense rather than the "order people about" sense (and these are expressed by different words in French and German, for example, so for machine translation the right sense would have to be found). If we are processing a particular text with the aid of the "restaurant script" this problem will have been settled for us because the Schankian MTRANS (in the last line of scene 2) will be tied only to the appropriate sense of "order".

This point may be clearer if we think of a language understanding system encountering a word it did not know: suppose it encountered "John ordered scampi", although "scampi" was not in its dictionary. Suppose the system had no restaurant script, but just representations of the senses of "order", including the standard one in which ordering was normally done by humans and of physical objects. These normal objects and agents we can call the preferences of the action, because they are not absolute—we can all understand children's stories with sentences like "The dog ordered a bone in the doggy shop"—but they do enable

important semantic choices to be made. In "John ordered the numbers", for example, we can reasonably say that we select the mathematical sense of "order" *because numbers fit the preferred object for that sense,* though not the preferred physical object of the sense of "order" appropriate to "ordering things commercially".

Now we can see the payoff from the restaurant script: if we are analysing our sentences with it then we know that even the unknown "scampi" is almost certainly a food, just because that is *the preferred object of the sense of the action* tied into the script at that point. If we had only the general sense of "order" we could infer only that a physical object was ordered.

Frames or scripts, therefore, will certainly help in determining topic or overall context, provided that we can reliably decide in advance what is the appropriate frame with which to analyse a given input. This assumes reliable cues (the word "restaurant" for example) which will not always be present ("They stopped off to eat at a little place he knew"), and a way of deciding which of these enormous information structures to use when several have been cued by a single sentence ("On the way home from the cinema, they stopped off at the supermarket before dropping into Luigi's restaurant"). Later, problems arise as to when to stop following a script and get rid of it in favour of another.

Claims by frame users

The real issue, though, is not technical but concerns what claims are being made by frame users. They are, I think, making a plot line hypothesis: "Humans, or computer understanding systems, can only understand a particular story by seeing how far it follows, or diverges from (as did the chocolate and soap story), the stereotypical story of that type." As Charniak puts it: "The primary mechanism in understanding a line of a story is to see it as instantiating one or more frame statements."

The trouble is that the claim is not obviously true, as we can see by making up an imaginary frame about a more remote cultural activity. I have jotted down the following for a male puberty rite in Charniak's (1975) notation—which is more or less self-explanatory:

Frame : 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 obtains TOOLS
(d) ELDER exhorts CROWD (on proper behaviour)
(e) (general condition)
Bad behaviour by CROWD→ Activity halts
(f) ELDER checks if CHILD properly purified
(g) (special condition)
CHILD not purified activity halted
(h) ELDER marks CHILD'S back
(i) (method suggested)
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 now try to show.

Suppose we have three "story sentences":

"Little Kimathis's mother accidentally { looked away / dropped her shoga / touched his arm }

during the puberty rite. The crowd drew back in horror."

If we wish to "understand" this story, do we need the frame above to do it? The frame covers the story with line (e) in some sense, given an adequate list defining bad behaviour accessible from the frame.

And yet it is clear that we understand the sentences perfectly well without the frame. In commonsense terms we could say that we infer from the sentences that the mother touching Kimathis during the ceremony was a bad thing. We do not need that information in order to understand.

One might argue that, in order to understand the above. a program should tie two parts of its representation together with some rule equivalent to:

human display alarm→other human has performed bad action.

A Martian lacking any earthly frame could understand the stories so long as he understood this rule and the constituent words. That is, of course, why I chose a puberty rite rather than a restaurant as a frame topic, for most of us are Martians where puberty rites are concerned. If we do understand the stories (and we do) it cannot be from our associated frame, because we do not have one. So we must understand it on the basis of knowledge organised on some simpler principles.

At present there is a tension between those who believe that frames-are necessary for language understanding, and those that think whatever is necessary can be provided by a system of cues and inference rules no more complex than the "humans show alarm" rule. So, to return to the "ordering scampi" example, provided we had a restaurant cue (which even a frame needs, as we saw) we could have a special inference rule tied to that cue that said "ordering is now normally of food". The reply from frame advocates is that these inference rules would be too numerous to be accessed but, as we saw, there are also enormous problems about access to and manipulation of frames, so that this question is not settled, either by argument or by the performance of programs.

Some frame advocates are not urging the "plot line hypothesis" (PLH) in the strong form of "you must have structure X to understand" but are claiming that it is more efficient to understand text from the topmost level down in that way.

However, an efficiency PLH cannot be assessed in the absence of frame application procedures. Moreover, and this is an important point, this efficiency PLH almost certainly rests on some statistical assumption about the degree to which texts do in fact follow the frame norms: the PLH would clearly be more plausible if, say, 90 per cent of texts about X contained explicitly all and only the knowledge in the frame about X, than if, say, only 5 per cent of texts about X did so.

Another possibility that I have put forward is the use for understanding of more static frames, such as one about the functioning and use of say, a car. These would be used to deal with the frequent occasions in text where the writer ignores the normal "semantic preferences" of words. This phenomenon is often called metaphor, as in "My car drinks petrol" which we understand perfectly well, even though any semantic description of "drink" would declare it to be normally an action done by animate things. Yet this "boundary breaking" sentence is perfectly comprehensible and of a type one might expect to find two or three times in any randomly chosen newspaper paragraph.

The suggestion is that the "car frame" might contain some formalised expression of such facts as "humans inject liquid (petrol) into a car" and "the car's engine uses liquid". Given a sophisticated matching process for semantic representations we might be able to match the representation for "My car drinks petrol" to one or both of these frame statements, and infer that, in understanding the sentence, we should project the (frame) sense of "use" onto that of "drink" in the sentence, thus changing the underlying representation of it to something like "My car uses petrol". This representation would now be closer to a literal meaning that the computer system could manipulate. An application of frames along these lines is being programmed, and might conceivably be brought to bear upon metaphor, a long intractable problem in the world of machine translations. □