SPEECH ACTS AND MULTIPLE ENVIRONMENTS

Yorick Wilks
Department of Language
& Linguistics
University of Essex
Colchester, England

Janusz Bien
Informatics Institute
University of Warsaw
Warsaw, Poland

This brief paper outlines joint work that implies a view of the explication of Speech Acts different from those current in AI. It is exploratory at present, but we intend that it should lead to programs as soon as possible. We propose a system that engages in dialogue with a human user so as to discuss others and in doing so, interprets the appropriate communicative force of the user's utterances. The three key features of the explication of communicative force presented are:

1. multiple environments: the ability to interpret an expression or a description relative to some particular knowledge, e.g. relative to the beliefs of one person about another. It is a generalization of the computer science approach to expression evaluation, using the notion of an environment to fix the interpretation of names occurring in the expression.
2. dual knowledge representation: an "experiential" representation is used to memorize sentences or other episodes (exemplified by pseudo-tests, or PTs), while a frame-type representation is used to memorize knowledge needed more generally for reasoning.
3. a least effort principle of understanding: considered to be a language universal, which requires (a) that the system maintain as much representation as possible in the frame-like forms, rather than in PTs, and (b) that the system maintain a highly redundant representation in the PTs, achieved by a process we call percolation of belief.

1. INTRODUCTION

Speech Acts (SAs) originated as a notion in philosophy [1] , [9] much preoccupied with the logical properties of sentences like "I promise to pay you five pounds" and, later, with access to the intentions in a speaker's head. Linguists worked with these notions subsequently but, like AI workers, they have added little that is distinctive to the original perception that utterances have a distinctive communicative force (as a threat, a promise, a warning, etc.), in addition to literal content. We believe that an AI approach to this area that is to contribute distinctively will have to make use of general principles concerning the environmental embedding of beliefs, etc., as well as a least effort maxim for their manipulation. The first of these, advocated by Bien in 1975 [2] , has been used by Cohen [5] and others of the Toronto group, but the distinctive feature we discuss here concerns the maintenance and change of such environments under a general requirement of cognitive efficiency.

2. TIRESIAS

We propose a system able to engage in a limited conversation with a human user: let us call It TIRESIAS for a reason that will appear. The topic of the conversation would be the relationships of a group of friends (see [7]) known in varying degrees to both TIRESIAS and the user. We expect sample dialogues such as:

$1 USER: Frank is coming tomorrow, I think.
$2 TIRESIAS: Perhaps I should leave.
U Why?
T Coming from you, $1 is a threat.
U Does Frank hate you?
$3 T I don't know-but you think he does, and that is what is important right now.

The important feature we require here is TIRESIAS' ability to evaluate its descriptions of persons differently in different environments Thus, at $1, TIRESIAS detects a threat (not in itself necessarily interesting) on the basis of evaluating its representation of itself, inside that of Frank, inside that of the user, which we might write (given that [P] is the data-structure for P inside Q, that is, Qs model of P):

{{{TIRESIAS} FRANK} USER} TIRESIAS

Note that all the data-bases are in TIRESIAS, so an outer TIRESIAS bracket exists round all evaluations, though we may sometimes omit it.

whereas at $3, when questioned, TIRESIAS is seeking to evaluate simple Frank's view of him, or {{TIRESIAS} FRANK} TIRESIAS

where he discovers that he has no specific information on what Frank thinks of him. The generation of $2, and TIRESIAS understanding of $1 as a threat, may, let us suppose, depend on a specific belief that the user believes that Frank dislikes TIRESIAS (but as $3 shows, not a belief by TIRESIAS that Frank dislikes him). Let us follow this through by examining inferences plausibly generated by TIRESIAS' construction of the two nestings of environments shown above and comparing their results.

The crucial difference here is between the evaluation of {FRANK} in {USER} and in {TIRESIAS} . Another assumption is important to what follows: the nesting of environments {FRANK} USER that the example requires will not in general exist already in the system (as it does in [5] et al). as a partition of the space of beliefs: we find it implausible and comptationally overweight to assume that all such partitions are already marked in the system. {FRANK} USER will here be constructed by "pushing down" {FRANK} into {USER} . To put it simply, TIRESIAS will construct the user's view of Frank by inference from his views of the two participants: he will assume that user believes about Frank what he does, unless explicit information about Frank likes the user, and that Frank believes that he (Frank) hates the user, then {{FRANK} USER} TIRESIAS will contain Frank dislikes user, since the belief of the user, that Frank likes the user, is overridden by explicit counter information. Thus, beliefs are transferred by counting in what it is reasonable for TIRESIAS to believe others believe, rather than counting out: that is to say, believing about another's beliefs only what we have explicit evidence for. To see how such structures are obtained, we must consider briefly the form of knowledge representations in the system.

3. FRAMES AND PSEUDO-TEXTS

The notion of frame is now well known [8] , and here we assume in addition a variant called a pseudo-text [1] which is an information structure that is the semantic representation of a text, obtained by a semantics-driven parser such as [10] . A PT is an interconnected set of proposition representations, called templates, each of which consists of three principal nodes or terminals, that represent in right-left order, the actor, action, and object of an event. The terminals are, in fact, trees of semantic primitives, but here we shall place the English names of those trees at the terminals. In such a notation, we may write the template for "Frank is coming tomorrow", said by the user, as follows in the PT in TIRESIAS for {USER}

[* asserts □]
[Frank come □
tomorrow]

where □ is a dummy node, and * is a pointer to the name of the PT in which we are ({USER} in this case). The same sentence would be parsed into a corresponding representation in {FRANK} with * now replacing "Frank" rather than "user". It should be emphasized that the PTs are a surface semantic representation of such sentences, and are not frames in the sense of a structure with an a priori determined list of slot and filler names to facilitate inferencing; they are essentially episodic text representations, but it is those we shall need, in addition to some frame information, to deal with the example to hand.

It will be evident that, in the description of input we have given, there is no general representation of Frank: information about him will be clustered in {FRANK} , but will also appear, as we saw, in {USER} and possibly many other PTs.

We have postulated in various publications [2] [10] the operation of a general principle of preference in language: one that asserts that the effectiveness of natural languages depends on them being constructed so that, in general, correct interpretations are those found with least effort: one could put this by saying that a system should prefer those representation that it is easiest to construct and maintain.

Now consider again the contrasting evaluations of {{{TIRESIAS} FRANK} USER} TIRESIAS and {{TIRESIAS} FRANK} TIRESIAS

If we further assume that at the privileged shallow level there is always a frame structure corresponding to the episodic PT, which contains standard slot names such as DISLIKE, then we may expect:

[Φ DISLIKE *]

in {TIRESIAS} given that the system knows no one who dislikes it, although in {USER} may be a template (corresponding to a belief of the user)

[Frank dislikes Tiresias]

Then, on "pushing down" {TIRESIAS} into {FRANK} to get {{TIRESIAS} FRANK} TIRESIAS}
in the way we sketched in the last section, no inconsistency of content is noted and no beliefs are reset. But on pushing the value of this object down into USER to get the deepest nesting shown earlier, we encounter

[Frank dislikes Tiresias]

as well as the two connected templates for $1 shown earlier in this section, and both of these items will become part of the constructed "systems view of USERs view of Frank".

4. BELIEF PERCOLATION AFTER PUSH-DOWN

Two effects may now be expected:

i) the inference from Frank dislikes Tiresias in {USER} ; together with plausible inference rules about dislike, fear, and leaving places, will generate the reply $2 even though no such consequence was derivable in {{TIRESIAS} FRANK} TIRESIAS} alone as we see at $3.

ii) on a least effort principle, the key template belief Frank dislikes Tiresias that will have been copied from {USER} into {TIRESIAS} FRANK} [*dislikes Tiresias] and into {TIRESIAS} TIRESIAS} as [Frank dislikes*] during the construction of the desired viewpoint, will now remain present when the push-down is, as it were, pulled apart: it remains in the PT copy. In this way, the beliefs of the user about Frank (about Tiresias) percolate to {TIRESIAS} FRANK} , or rather to the relevant part of {FRANK} (the privileged shallow level), and similarly to {TIRESIAS} TIRESIAS}.

This resulting, highly redundant, representation may seem a paradoxical result of a least effort principle but we maintain that this spread of beliefs from PT to PT, via push-down evaluation, leads to a structure requiring less effort when the same or similar input is encountered in the future. This is because a push-down of environments as we have described it may be considered to require considerable resources: so, for example, {FRANK} should be checked, before being evaluated in {USER} : to see if that push-down had been done lately, and, if it had, it *need* not be done again, since the relevant beliefs, Frank dislikes Tiresias in this case, would already have percolated into it. One may notice that such percolations, if they exist, deny the whole basis of intensional logic (the denial of p believes x -> x), but seem consistent with common sense, and more technically with the sleeper effects [6] , the inference of beliefs from unreliable sources.

It should not be emphasized that we are not trespassing into possible worlds: for they require the assignment of all variables. In the recursive pushing down of PTs we envisage only relevant beliefs will be reset. It is certainly true, however, that these suggestions will impinge on the area on intensional logic (see [2] and, of course, Tiresias in the Greek myth was the seer who knew that Jocasta was the mother of Oedipus (who knew that Jocasta was the wife of Oedipus, and even that Jocasta was Jocasta, but NOT that key fact known to Tiresias). But there is no space to explore these connexions, although we do believe that even the simple considerations set out here could lead to a richer procedural explication of SAs than those current in AI.

REFERENCES

(1) J.L. Austin, "Performative utterances"(1961) Philosophical Papers. Oxford.

(2) J.S. Bien, "Towards a multiple environment model of natural language" Proc. IJCA1 4,(1975) 237-240.

(3) J.S. Bien, "Computational explication of intensionality" in A.J.C.L. Microfiche. No. 97, (1976).

(4) H. Clark &. C. Marshall, "Definite Reference and Mutual Knowledge", Sloan Workshop, Pennsylvania, May 1978.

(5) P. Cohen, "Planning Speech Acts" Ph.D. Thesis, Dept. Computer Science, Univ. Toronto (1978).

(6) C. Cruder et al. "The absolute sleeper effect" Personality and Social Psychology, 36, 1978, 1061-1073.

(7) P. Hayes & M. Rosner, "Ully". Proc. AI3 B Conf., Edinburft 1976, 46-49.

(8) M. Minsky, "Matter, Mind and Models" in Bobrow (ed.) Semantic Information Processing, MIT, 1968, 203-43.

(9) J. Searle, Speech Acts, Oxford, 1969.

(10) Y. Wilks, "A preferential pattern-seeking semantics, Artificial Intelligence, 6, 1975, 53-74.

(11) Y. Wilks, "Making preferences more active", Artificial Intelligence, 11 , 1978, 197-223.