Intentions in Communication

Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (editors)
(SRI International, University of Illinois, and SRI International)

Reviewed by
Allan Ramsay
University College Dublin

Consider the following hypothetical exchange between A and B:

> *A:* Do you know the time?
> *B:* It's OK, the boxing club have cancelled their booking.

It is fairly easy to imagine a situation in which this exchange makes perfect sense. The participants might, for instance, be playing badminton in a hall that is usually used by a boxing club after the end of the badminton session, and A wants to know whether there is time for another game before the boxing club arrive. It is also fairly easy to tell an informal story about how the conversational exchange above would serve this purpose. The usual story is that A thinks she would need to know what the time is in order to see whether she and her partner can go on playing. She doesn't know the time, but she knows that if she can persuade B to tell her then she will know it. She also knows that B won't be able to tell her unless she (B) knows it herself. If she demands to be told, and B in fact *doesn't* know what it is, then one or the other of them might be embarrassed. So A delicately asks whether B has this information, rather than asking for it outright.

At this point B is supposed to recognize that A is unlikely to have any goals that could be directly furthered by finding out whether B knows the time. Since A has no goals that will be directly furthered by B's response to this question, and since people don't usually say things unless they expect some response that will help them get on in the world, B starts to wonder what A is up to. She realizes that the question that A actually produced was only a preliminary move, and that it was uttered in order to facilitate asking the follow-up:

> *A:* What time is it?

But B has an inquiring turn of mind, and she wonders why A might want to know what time it is. A reasonable guess is that she wants to do something that is constrained by the time, and she wants to know whether she will in fact be able to do it. What might she want to do? It's very possible that she will want to go on playing badminton. Now B does not in fact know the time, but she does know that the usual constraint on when they have to stop playing no longer holds, so she tells A about this fact, rather than saying anything at all about the time. The conversation worked perfectly, despite the fact that when it finished A did not know the time, and indeed did not even know whether B knew it.

How you turn this informal story into a computational system that can take part in conversations is a fascinating, and unsolved, problem. Artificial intelligence and computational linguistics work in this area has very largely been based around attempts to encode Searle's (1969) notion of *speech acts* in the STRIPS notation for describing actions, and then to use standard AI planning techniques (Fikes and Nilsson 1971; Sacerdoti 1977) to manipulate such actions. The best-known work in this area is probably by Cohen and Perrault (1979) and Allen and Perrault (1980), though there were a number of other people exploring much the same ideas at about the same time. The volume under review here consists of a collection of papers presented at a 1987 workshop on "Intentions and Plans in Communication and Discourse." The contributors to the workshop are all eminent workers in the field, and the collection justifies their eminence. These papers present good descriptions of the basic framework, argue for interesting variations and extensions, and probe a number of quite subtle foundational issues. It is fair to say that they represent the state of the art in the area (or at any rate the state of the art in 1987, and unfortunately you can't ask for anything much more recent to be available in the form of a properly edited hardback collection).

What more can we say about them? There are fourteen substantive papers here, together with an introductory chapter and commentaries on most of the main papers (with some of the commentaries themselves presenting new ideas rather than simply discussing the relevant papers). They include contributions from philosophers (Bratman, Jones, Vanderveken, Thomason, Searle), AI people (Cohen and Levesque (two papers), Perrault, Pollack, Kautz, Litman, and Allen), computational linguists (Grosz and Sidner), and "ordinary" linguists (Pierrehumbert and Hirschberg, Clark and Wilkes-Gibbes). The striking thing about this list, which is derived from the editors' own classification of the contributors in the overview chapter, is the way linguists are outnumbered by philosophers and computer scientists. One of the good points about the collection is that it contains contributions from people attacking the problem from a wide range of viewpoints, but I have to say I was surprised at the relative numbers. In a discussion of an essentially linguistic matter we find three papers out of fourteen written by people who would primarily call themselves linguists (Grosz and Sidner, Pierrehumbert and Hirschberg, Clark and Wilkes-Gibbes). Of course the others are all thoroughly at home with the relevant linguistic literature, but by their own descriptions of themselves they are philosophers or computer scientists.

Most of the contributors, then, are philosophers and computer scientists. As philosophers and computer scientists, they quite reasonably seem to be primarily interested in the gap between the literal meaning of an utterance and its intended and/or actual effect on a hearer. This is fair enough, especially in view of the declared focus of the book, but it does presuppose that we have a very good idea of what the literal meaning is. Sadock's discussion of the papers by Vanderveken and by Cohen and Levesque, for instance, contains the following diagram:

$$\text{FORM} \xleftrightarrow[\text{grammar}]{} \text{CONTENT} \xleftrightarrow[\text{pragmatics}]{} \text{EFFECT}$$

The papers in the current volume are almost entirely directed to the right-hand side of this diagram—to the relation between CONTENT and EFFECT. The problem is that you cannot work on this part of the problem without assuming some view of the left-hand side of the diagram. It is perfectly sensible to say that you think there are a number of reasonable-looking theories of how to get from FORM to CONTENT, that theory X looks as good as any other to you, and that you are therefore going to assume that your starting point is the kind of thing that theory X would lead to. You must, however, at least say this much about the relation between FORM and CONTENT.

Unfortunately very few of the papers in the current volume say anything at all about the left-hand side of the diagram. It does not matter that they say nothing about the way that syntactic and lexical facts about the utterance encode propositional content, but it does matter that they say very little about what they expect propositional content to be like. In particular, there is virtually no discussion of the widespread view in semantics that propositional content should be seen more as a relation between situations than as a function from situations to truth values. Barwise and Perry (1983), Heim (1983), Kamp (1984), Groenendijk and Stokhof (1987), among others, have all argued for theories where the function of literal meaning is to change the discourse situation by adding constraints. The details of what a discourse situation is like, how it should be described, and what sorts of constraints can be added may vary from theory to theory, but the general framework does seem to be common. The view of meaning as a relation between situations has led to revealing analyses of a number of linguistic phenomena—anaphora resolution (Kamp 1984), clefts and presuppositions (Delin 1989), habitual propositions and generic NPs (Schubert and Pelletier 1989)— and it seems very likely that this kind of approach to semantics might provide a rich background for the kind of phenomena under discussion in the current collection. As far as I can tell, however, most of the authors in the current collection assume that literal meaning should be described in terms of the truth conditions of propositions expressed in first-order logic, or at best in some modal epistemic logic. Two of the papers refer to Barwise and Perry, but only to say that these authors have some contribution to make on the matter without going into any detail. One (Thomason's) mentions Kamp and Heim in passing. The rest leave it to the reader to guess what the source of their semantic theory is, though in every case it looks as though the basic representation language is either predicate calculus or modal epistemic logic.

It would therefore have been interesting to have had more contributions that tried to do more work on the left-hand side of the diagram—that tried to cram more information into the literal meaning. Decisions about whether to use anaphora or not, about whether to use clefts or simple declaratives, about whether to extrapose complements in order to affect the discourse structure all seem to obey fairly well-defined rules that would qualify them to count as semantic rather than pragmatic choices (Halliday 1985). The more you can do when going from FORM to CONTENT, the less you will have to do to get from CONTENT to EFFECT. The only person to have taken this approach is Vanderveken, and his contribution, though interesting, would have benefitted from closer analysis of existing relational theories of meaning.

The choice of modal logic as a framework for discussing knowledge and belief also seems likely to hinder progress in this area. Much of the discussion in this book is concerned with reasoning about what the participants in a dialogue might know and believe. It would be very odd if this were not so, since the effects of linguistic acts are very largely changes in the knowledge and belief of the participants, so reasoning about what it might make sense to say will be very largely reasoning about what the participants know and believe and what you would like them to know and believe. As such, a good treatment of knowledge and belief is essential.

Modal treatments of knowledge and belief are convenient. If you regard knowledge and belief as modalities, you can borrow possible worlds semantics for them; and if you use possible worlds semantics you can do your inference using a perfectly ordinary theorem prover for predicate calculus, with the possible worlds encoded as an extra argument and the accessibility relationship treated either by special purpose routines (by so-called "procedural attachment") or simply within the ordinary framework of predicate calculus. Hintikka's (1962) observation that modal logic and epistemic logic had something in common has proved very fruitful, as has Moore's

(1980, 1984) realization that the relations between possible worlds could be described in predicate calculus and hence that you did not need to change your existing theorem provers very much to deal with modalities. Nonetheless, the kinds of task involved in reasoning about the epistemic effects of natural language utterances seem to require rather greater subtlety than is available within a modal framework.

In particular, modal treatments of knowledge and belief require that the set of propositions that any individual knows or believes must be consistent and closed under logical consequence, that individuals must be unable to distinguish between logically equivalent propositions, and that the proper treatment of quantifiers should be in terms of rigid designators. These are all unavoidable consequences of the decision to think in terms of modalities and possible worlds. You can't escape by issuing disclaimers. Epistemic logic as modal logic leads inexorably to possible worlds semantics, and possible worlds semantics leads inexorably to consistency, closure, indistinguishability, and rigid designators.

Real belief sets satisfy none of these properties. People can easily believe contradictory things (often as many as six before breakfast), they often fail to recognize the consequences of their beliefs, they frequently believe one proposition without believing another that is logically equivalent. Satisfactory accounts of the differences between the two readings of *John wants to buy a dog* do not appeal to the notion that in one case but not the other John has a rigid designator for the dog that he wishes to buy. And so on and so on. The only contributor who seems to be aware of this as a fundamental problem, rather than a technical inconvenience, is Vanderveken, whose notion of strict implication is subtler than the usual material implication but is still prone to many of the problems of relevance logic (see for instance Epstein's (1990) account of relatedness logic, which considers these issues in detail).

It would therefore have been good to have had more papers directly addressing the problems of intensionality (the other sort, not intentionality). A good deal of recent work in formal semantics, and in particular in the foundations of formal semantics, has been concerned with investigating various highly intensional logics to see whether they provide an adequate background (Bealer 1982; Chierchia et al. 1989; Turner 1990; Ramsay 1990). Such logics acknowledge that propositions necessarily have truth conditions, and that these are very likely to be of interest to anyone who is trying to interpret the reasons why someone might produce a specific utterance. In much the same way, people necessarily have genders that are often of interest to other people who are deciding how to interact with them. We would not, however, wish to equate a person with their gender. There is much more to an individual than whether they are male or female, and anyone who fails to recognize this will have a seriously impoverished social life. Similarly, there is much more to a proposition than a description of the circumstances under which it will be true, and anyone who fails to recognize this will have a seriously impoverished view of semantics.

These remarks should not be taken too negatively. The collection does show what some of the leading researchers in the area thought in 1987, and what they thought in 1987 is at least as interesting as what most of us are thinking in 1991. I just feel that it may be time to reconsider some of the foundational assumptions that underlie this work. In particular the assumptions that literal meaning is basically about the truth conditions of propositions and that modal logic provides an appropriate framework for epistemic and doxastic reasoning seem likely to act as brakes on further progress. The current volume shows that you can do a lot even on the basis of these assumptions, and shouldn't rush off to reconstruct the foundations of speech act theory until you have fully assimilated the achievements of the people whose papers are collected here. Once you do appreciate what can be done on this basis, however, it may be time to

see whether richer theories of natural language semantics and subtler representations of knowledge and belief may help you get even further.

References

Allen, James F. and Perrault, C. Raymond (1980). "Analyzing intention in utterances." *Artificial Intelligence*, **15**, 148–178.

Barwise, Jon and Perry, John (1983). *Situations and attitudes*. Cambridge, MA: The MIT Press.

Bealer G. (1982). *Quality and concept*. Oxford: Clarendon Press.

Chierchia, Gennaro; Partee, Barbara H.; and Turner, Raymond (1989). *Properties, types and meaning, volume I: Foundational issues* and *volume II: Semantic issues*. Dordrecht: Kluwer.

Cohen, Philip R. and Perrault, C. Raymond (1979). "Elements of a plan-based theory of speech acts." *Cognitive Science*, **7**(2), 171–190.

Delin, Judith (1989). *Cleft constructions in discourse*. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Epstein, R.L. (1990). *The semantic foundations of logic, volume 1: Propositional logics*. Dordrecht: Nijhoff International Philosophy Series.

Fikes, Richard E. and Nilsson, Nils J. (1971). "STRIPS: A new approach to the application of theorem proving to problem solving." *Artificial Intelligence*, **3**(4), 251–288.

Groenendijk, J. and Stokhof, M. (1987). "Dynamic predicate logic: Towards a compositional and non-representational discourse theory." Technical report, University of Amsterdam.

Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Arnold.

Heim, Irene (1983). "File change semantics and the familiarity theory of definiteness." In: Bäuerle, Rainer et al. (eds.) *Meaning, use and the interpretation of language*. Berlin: Walter de Gruyter, 164–189.

Hintikka, Jaakko (1962). *Knowledge and belief: An introduction to the two notions*. Ithaca: Cornell University Press.

Kamp, Hans (1984). "A theory of truth and semantic representation." In: Groenendijk, J.; Janssen, J.; and Stokhof, M. (eds.) *Formal methods in the study of language*. Dordrecht: Foris. 277–322.

Moore, Robert C. (1980). "Reasoning about knowledge and action." Technical report 191, Artificial Intelligence Center, SRI International.

Moore, Robert C. (1984). "A formal theory of knowledge and action." In: Hobbs, Jerry R. and Moore, Robert C. (eds.) *Formal theories of the commonsense world*. Norwood, NJ: Ablex. 319–358.

Ramsay, Allan M. (1990). *The logical structure of English: Computing semantic content*. London: Pitman.

Sacerdoti, Earl D. (1977). *A structure for plans and behavior*. Amsterdam: Elsevier North-Holland.

Schubert, Lenhart and Pelletier, F. Jeffry (1989). "Generically speaking, or using discourse representation theory to interpret generics." In Chierchia et al. 1989, 193–268.

Searle, John R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.

Turner, Raymond (1990). *Truth and modality for knowledge representation*. London: Pitman.

Allan Ramsay is Professor of Artificial Intelligence at University College Dublin. He holds a Ph.D. in AI from the University of Sussex. His recent publications include a text on the applications of non-standard logic in AI (*Formal Methods in Artificial Intelligence*, Cambridge University Press, 1988) and an attempt at a formal semantics for English (*The Logical Structure of English*, Pitman, 1990). Ramsay's address is: Department of Computer Science, University College Dublin, Dublin 4, Ireland; e-mail: amramsay@ccvax.ucd.ie