SRI International

METHODOLOGICAL NOTES ON A COMPUTATIONAL MODEL OF REFERRING

Technical Note 434

April 20, 1988

By: Amichai Kronfeld

Artificial Intelligence Center
Computer and Information Sciences Division

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

The research reported in this paper was supported, in part, by the National Science Foundation Grant DCR-8407238. The views and conclusions expressed in this paper are those of the author and should not be interpreted as a representative of the views of the National Science Foundation or the United States Government.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Methodological Notes on a Computational Model of Referring*

Amichai Kronfeld

April 20, 1988

Contents

*The research reported in this paper was supported, in part, by the National Science Foundation Grant DCR-8407238. The views and conclusions expressed in this paper are those of the author and should not be interpreted as representative of the views of the National Science Foundation or the United States Government. I am grateful to Doug Appelt, David Israel, and Martha Pollack for their helpful comments.

1 Introduction

When two people talk to each other, they had better both know what they are talking about. This, I hasten to add, is not a condescending directive enjoining nonexperts to keep quiet. It is rather a general rationality constraint on our use of language. We frequently mention *things:* we may promise to return a book, request that a window be closed, remark that a certain tool is useful, and so on. But, as rational agents, we can expect such speech acts to be successful only if our addressee knows what book, window, or tool is being discussed. Thus the question arises: how do we let our audience know what we are talking about? How, in other words, do speaker and hearer form an agreement as to which entities are the subject of the conversation?

The question seems almost trivial: our hearer is expected to know what we are talking about because it is assumed that he understands the language we use and, moreover, we have already *told* him what is being discussed. If, say, we are talking about my mother, then I have presumably indicated that fact already by using an appropriate noun phrase, such as "my mother." Since my addressee understands English, he comprehends that the person I am talking about is my mother. Could anything be simpler?

Nonetheless, the ease with which a native speaker can *refer* — i.e., indicate what entity is being discussed — is deceptive. Like other mechanisms of language use, referring is easy to do but extremely difficult to explain or simulate (not to mention the problem of *really* making a computer refer — whatever that is supposed to mean). The case of "my mother" is indeed quite simple because the hearer knows that each and every person in the world has one and only one mother. Thus, the question, "Which of your mothers are you talking about?" never arises. But it is rather surprising that the act of referring is hardly ever that straightforward. "Did you find Maya's shirt?" my wife asks me, and I know she is talking about the shirt my daughter left in the playground. But my daughter is not the only girl named "Maya" and, besides, my daughter has more than one shirt. My wife's referring act is therefore successful, even though she never *tells* me explicitly what shirt she is talking about. In more complex cases, the success of the referring act is even more puzzling: "You know," I tell my wife while watching a couple at a party, "Jane's husband seems to be quite a romantic guy." "He is not her husband, you fool!" my wife answers. She, obviously, knows exactly who I am talking about, and I know that she does despite the fact that what I *told* her has to do with a different person entirely.

These two simple examples are neither unique nor exceptional. As a matter of fact, we hardly ever *tell* our hearers explicitly what we are talking about, although we expect them to figure it out. Indeed, in general, we are

very good at providing hearers with just enough information — no more and no less than necessary — to enable them to understand the subject of the conversation. How do we do that? And how can we teach a computer to do it?

The goal of my current research is to outline an answer to this question. Such an answer will ultimately take the form of a computational — i.e., algorithmic — model of referring. However, implementation is not my main concern, and the reader should not expect a blueprint for the construction of a specific system. Rather, my intention is to specify the general principles that ought to be incorporated in any particular implementation. These principles are derived from three methodological tenets The first is that a theory of referring should explain how noun phrases usage is related to objects in the world, not how one use of a noun phrase is related to another. The second is that referring is a speech act. The third principle is that a theoretical account of referring must be firmly anchored in a well-defined, well-defended philosophical framework. In this paper I describe what exactly is meant by these three methodological tenets.

2 Internal and external perspectives

The act of referring is performed through the use and interpretation of noun phrases in a conversation. But we should be careful to distinguish between two perspectives — one *internal*, the other *external* with respect to the discourse. From the internal perspective, our main interest is the relation of coreference among symbols. From the external perspective, on the other hand, what interests us is the relation between symbols and the objects they represent (cf. [4, 44-45]). Consider the following exchange between Representative Louis Stokes and Assistant Attorney General Charles J. Cooper during the Iran-Contra hearings:

1 **Stokes:** *And lastly, when you and the Attorney General interviewed Colonel North, he was not under oath at that time, was he?*

Cooper: *No, he was not.*[1]

Although neither North nor the Attorney General, nor Cooper himself, was under oath when they met, it is clear that, when Cooper says "No, he was not," he means North. How do we know that? How is the connection between the expressions "Colonel North" and "he" established? These are the

[1] *The New York Times*, June 26, 1987, p. 4.

typical questions that are asked from the internal perspective. The questions that are asked from the external perspective, on the other hand, are different. How is the connection between the expression "Colonel North" and the *person* North established? What does it take for a hearer to recognize who the Attorney General is? When Stokes says "you," whom does he mean and how do we know what he means? This is what matters to us from the external perspective. Note that it is entirely possible that a native speaker of English would succeed in matching the expression "Colonel North" with the right occurrence of "he" without understanding at all who Colonel North is. His success must be explained from the internal perspective, his failure from the external one.

It is tempting to identify the internal/external dichotomy with the distinction between linguistic knowledge and world knowledge. This would be a mistake, however, although these types of knowledge may be characterized respectively as "internal" and "external." For example, the linguistic knowledge that proper names are capitalized in English is important to understanding that "North" stands for the person Oliver North. At the same time, the nonlinguistic ("external") knowledge to the affect that only when a man is *answering* questions does it matter whether or not he is under oath plays an important role in interpreting the pronoun "he" in Cooper's response: since no one else was answering questions in the situation described by Stokes, the query as to whether someone had been under oath must pertain to North. Thus, linguistic knowledge is relevant for explaining the connection between an expression and an object (the external perspective), while world knowledge is often used to disambiguate the anaphoric link of pronouns (the internal perspective). The crucial difference between the two perspectives is this: from the external perspective, the criterion of success for the hearer is correct identification of the object being discussed. From the internal perspective, the criterion of success is the right matching among symbols.[2]

In the field of artificial intelligence (AI) there is a tendency to blur the distinction between these perspectives — partly, I think, because the external perspective is not well understood or appreciated. In all the natural-language systems with which I am familiar, reference to objects is handled under what may be called the *standard-name assumption.* According to this assumption, *all* objects in the domain have standard names that are known

[2] As Martha Pollack has pointed out to me, a hearer can make the right match between two noun phrases without assuming that both designate the same thing. For example, in "I returned to the house and the door was open" a hearer must be aware that the door is the house door. Note that the inference can be drawn internally, that is, a hearer can deduce the connection without knowing which house is being discussed.

to all participants in the discourse. In such systems, the act of referring is successful when (and only when) the machine associates the right standard name with the noun phrase. For example, if a user types "The screwdriver is broken," referring to the object whose standard name is, say, the constant S_1, the referring act succeeds if and only if the machine associates the constant S_1 with the expression "The screwdriver." Given this approach to referring, it is easy to ignore the external perspective entirely. Since standard names are simply labels, we tend to take the relation between the standard name and its bearer for granted. All that is left for us to do is to show how one symbol (the noun phrase) is associated with another (the standard name). This places us firmly where our view is from the internal perspective.

But the external perspective is indispensable. First, most of the objects we talk about do not have standard names; hence labeling is hardly the right model for explaining how referring is done in natural language.[3] Therefore, if we are ever to explain how natural language is capable of representing reality for us, we must give up the standard-name assumption and pay closer attention not only to the way symbols are associated with one another, but also to the way they correspond to the objects they stand for. Furthermore, if we ignore the external perspective, we lose the basic rationale for the act of referring itself. Consider the following examples:

2 (a) *The farmer down the road owns a donkey named Buridan. He feeds it.*

(b) *The average farmer owns a donkey. He feeds it.*

(c) *If John owns a donkey, he feeds it.*

(d) *If a farmer owns a donkey, he feeds it.*

In example 2(a), in contrast with examples 2(b), the speaker has in mind a particular farmer and a particular donkey. But from the internal perspective, this fact is of limited interest, since both "The farmer down the road," and "The average farmer" have equal potential for initiating anaphoric chains. Similarly, in example 2(c), there is a particular owner that the hearer is expected to identify. No such identification is required for the interpretation of example 2(d). Still, from the internal perspective all three — "John," "a farmer," and "a donkey" are treated equally: they are assigned *discourse entities* [9,23], which are basically "conceptual coathooks" on which a hearer "hangs" subsequent noun phrases in the anaphoric chain. Whether the

[3] Of course, we can always generate new labels whenever a new object is introduced, but we are still left with the problem of explaining how the object is identified in the first place.

discourse entity corresponds to a real object or not is immaterial as far as the internal perspective is concerned.[4]

But the question of whether the noun phrase corresponds to a particular object that the hearer is expected to identify is of prime importance for a natural-language system. Consider a speaker who is attempting to achieve something by means of language. Suppose, for example, that Luke Skywalker of Star Wars instructs his trusted robot C3PO to look for Han Solo's spaceship. It makes a great deal of difference to Luke whether it is understood that he has a particular spaceship in mind and furthermore, whether the robot will be able to identify it. If the robot simply associates the phrase "Han Solo's spaceship" with the correct standard name and then switches itself off, Luke has not succeeded in his speech act. Any system that combines linguistic and nonlinguistic actions, and that is capable of cooperative behavior, must be able to talk about objects. It must distinguishes when a noun phrase has a referent in the real world from when it does not, when a particular type of knowledge of the referent is required from when it is not, when knowledge of the referent is presupposed from when it should be actively sought. Without the external perspective, we cannot even ask these questions.

Quite bluntly, my own view is that *only* the external perspective is relevant to a theory of referring if such a theory is supposed to explain how expressions represent things, and how such representations can be used by a speaker for a specific purpose. This, of course, does not mean that the internal perspective is not important. On the contrary, the internal perspective is essential to our understanding of natural language. Moreover, although research from the internal perspective is largely independent of its external counterpart, the opposite is not true, since a theory of referring needs the constraints that a theory of anaphora resolution provides. Thus, if an internal theorist asks me why he should adopt the external perspective, my answer would be that he does not have to. There is enough to be done from the internal perspective for generations to come. If, on the other hand, I am asked why we need the external perspective at all, my answer would be that, without it, both semantics — in the sense of relating language to the world — and pragmatics — in the sense of correlating language with the purpose of its use — are impossible.

Throughout this section, I have talked about particular entities that a speaker may have in mind and a hearer is supposed to identify. Needless to say, there are many kinds of such entities. For example, *institutions* (the Supreme Court, the Presidency), *abstract entities* (the number 47, the the-

[4]The term "conceptual coathook" is attributed to William Woods.

ory of relativity), *events* (the shooting of J. F. Kennedy, World War II), and so on. We can talk about all such entities, and a comprehensive theory of referring should undoubtedly explain our ability to do so. We should, however, concentrate on the familiar class of physical objects that, so to speak, are simply "out there" in the physical world for everyone to see: plants, cars, shirts, persons, houses, animals, etc. From now on, whenever I talk about objects, it is such physical objects that I have in mind. The problems associated with referring to other types of entities will be deliberately disregarded in the following discussion.

There are a number of related reasons for restricting ourselves in this way. From a theoretical point of view — as Strawson [22] has argued — the category of physical objects is *basic* in the sense that, without it, identification of particular entities in other categories would be impossible (Strawson uses the term "material bodies"). From a practical point of view, a computer system that, among other things, is capable of referring is most likely to operate in a context in which the perception and manipulation of physical objects are of prime importance. But the main reason is essentially methodological. Physical objects of the type we are considering are more permanent than events and, unlike abstract and institutional entities, they can be perceived. Consequently, they can be recognized, identified, and reidentified in the most obvious and immediate manner. Focusing our attention on objects possessing such features simplifies the discussion significantly: we begin with a firm intuitive grasp of the category of things we are talking about, individuation is less of a problem (we all know what counts as one shirt or one person..), and identification and reidentification are easier. When we shift to other kinds of entities, such as the Presidency, World War II, or the integer 3, it is much harder to understand the relation between a noun phrase and an intangible object because the latter is much more difficult to grasp. All in all, physical objects seem to provide a good point of departure. If a theory of referring cannot handle physical objects, it stands little chance of coping with anything else.

3 Referring as planning

The act of referring is done typically by means of noun phrases in conversation. However, not all noun phrases are intended to be used in this manner, not even those that have the form of a definite description. For example, in the sentence "The whale is a mammal" (uttered, say, in a biology class), the speaker is making a general statement in which no referring relation is presupposed between the noun phrase "the whale" and any individual whale.

Let us reserve the term *referring expressions* for those instances of noun phrase usage that are intended to indicate that a *particular* object is being talked about. Note that one and the same noun phrase may sometimes function as a referring expression and at other times not. While discussing the whale at the Monterey aquarium, for example, I may comment "...but it costs a lot to feed the whale." Here, the noun phrase "the whale" is clearly used as a referring expression, in contrast to the above example.[5]

Thus, whether or not a particular noun phrase is a referring expression depends on the way it is intended to be interpreted. A theory of referring, therefore, is not a theory of language but of language *use*. In general, theories of language use (that is to say, *pragmatic* theories) specify and explain the ability of humans to use language for some purpose. Consequently, an account of referring should specify and explain human competence in using referring expressions to achieve particular goals. Now, pragmatic theories have concentrated on two complementary aspects of language use. The first, which is at the heart of Grice's theory of meaning, is this: when we use language, we typically achieve some of our goals by making our audience recognize our intentions to achieve them. For example, I can succeed in congratulating you simply by making you recognize my intention to do so. Once you have recognized my intention, you are thereby congratulated and nothing else is necessary. This is a unique feature of communication, as Grice [8] was the first to notice. The second aspect of language use, which is a central element in Searle's speech acts theory is this: we make our audience recognize our intentions by following mutually known rules that determine what the utterance of a particular expression *counts as*. For example, underlying the recognition of an intention to pay one's debt is a rule that is mutually known by both speaker and hearer; this rule specifies that the utterance of "I hereby promise to pay my debt" *counts as* placing the speaker under the obligation of paying his debt [20].[6]

These two general principles of language use determine the structure of pragmatic theories. For any communication act, such theories should state precisely what relevant speaker's goals are involved, and on what basis a speaker expects and intends these goals to be recognized (cf. [7]). Moreover, since the relation between language use and a speaker's goals is what needs to be explained, it is natural, within the context of computational linguistics, to consider language use as a *planning* problem [1,2,5,7]. What underlies

[5]It may be argued that the noun phrase "the whale" in "The whale is a mammal" is a referring expression, referring, as it were, to the *species*. Such noun phrases can still be contrasted with others that do not refer at all — for example, "He left me in *the lurch*."

[6]This is the rule that defines the institution of promising — see ibid., 60. For a discussion of such rules (*constitutive rules*, as Searle calls them), see ibid., 33–42.

the *generation* of an utterance is a plan (constructed by the speaker) to achieve certain goals through available means (linguistic or otherwise). The *understanding* of the utterance involves the hearer's recognition of the goal, as well as of the plan itself (or perhaps just a part of it). By regarding language use as a special case of planning, we are provided with a large array of computational tools that have been developed within the field of AI in recent years. Moreover, since planning is a special form of rational behavior, the justification of rules for language use can be grounded upon a general theory of rationality [6,10,11].

Such a computational approach to language use should govern the referring model we are after. A plan-based account of referring is an integral part of a plan-based theory of speech acts. At a certain point in the planning of a speech act, it may become obvious that, as a precondition for further steps in the plan, the speaker must make the hearer identify a particular object as being relevant to the conversation. To achieve this goal an act of referring then becomes necessary. Thus, a computational model of referring must show how the successful use of a referring expression in a given context is due to the solving of a planning problem — given also a goal, various rationality assumptions, and relevant linguistic institutions.

Note that, as is the case in other plan-based accounts of communication acts, the effects of referring are intended to be primarily on the hearer's model of the world, which naturally includes a representation of the speaker's model. In general, if my intention is that you recognize my goal, the typical way to satisfy my intention is to let you know what I think, hoping that I shall thereby alter your model of my mental state. The same reasoning applies to referring: by effecting changes in the hearer's model of the speaker model, the speaker may be successful in his referring act. That is, he may succeed in making the hearer recognize which particular object is now relevant to the conversation. Of course, the speaker's model, in turn, includes a representation of the hearer's model; hence, by making the hearer recognize the object in question, the speaker's model itself changes. Thus, a model of referring — like other plan-based accounts — should describe how after successful referring, the *hearer's* model of the speaker's mental state and the *speaker's* model of the hearer's mental state are both changed.

In sum, what I have been saying so far is this: a pragmatic theory of referring is one that specifies and explains human competence in using *referring expressions* to achieve certain goals. Since the relation between referring expressions and a speaker's goals is what must be explained, it is natural to consider referring as planned action. This, in turn, requires showing how the use of referring expressions is systematically related to changes in both the hearer's and speaker's mental states.

The view that the referring act is a planned effort to achieve certain goals through linguistic means simply follows from the fact that referring is a *speech act*, since all speech acts are attempts to achieve goals through linguistic means. Following Searle, I distinguish between two kinds of speech acts: *propositional* and *illocutionary*. There are two kinds of propositional acts, according to Searle, namely, referring and predicating, both of which are generally performed as parts of larger, complete speech acts. The latter constitute the illocutionary acts: promising, stating, requesting, congratulating, commanding, questioning, asserting, thanking, warning, advising and the like. Now, although referring is a speech act, it is different from illocutionary acts in the following four important ways.

Literal goals. In performing one and the same speech act, a speaker may have many distinct goals. For example, by uttering "The house is on fire!" a speaker may intend to inform the hearer that the house is on fire, scare the hearer half to death, and/or make the hearer leave. Only the first goal, however, is what I call a *literal* one. Literal goals are the goals of Gricean communication intentions, i.e., they are intended to be achieved through recognition of the intention to achieve them.[7] Thanks to Austin, Grice, Searle and others, we have a fairly clear notion of what the literal goals of *illocutionary* acts are. For example, the literal goal of a promise is to let the hearer know that the speaker places himself under an obligation to do something, while the literal goal of a request is to let the hearer know that the speaker desires something. But it is not clear at all what the literal goal of referring is. Needless to say, without a clear notion of a literal goal, the task of treating referring as a planned speech act cannot even get off the ground.

Conditions of satisfaction. Illocutionary acts have *illocutionary force* and *propositional content*. An assertion that it is raining and a question whether it is raining share the same propositional content but have different illocutionary forces. A promise to come home early and a promise to pay one's debt have the same illocutionary force, but differ in propositional content. Part of the illocutionary force is the illocutionary *point* [21], which specifies the point or purpose of the (type of) act. The illocutionary point together with the propositional content determine what Searle [19] calls the act's *conditions of satisfaction*: a request that a door be opened is satisfied

[7]The term "literal goal" is taken from Kasher [13], where literal *purposes* are introduced. Our use of the two terms is virtually the same, except that Kasher wishes to explain what literal purposes are in a way that is independent of Gricean intentions.

if and only if the hearer indeed opens it; by the same token an assertion that the door is closed is satisfied (true) if and only if the door is indeed closed.

A referring act, however, has neither illocutionary force, nor propositional content, and although it obviously has a purpose, it is not clear what its conditions of satisfaction are. As with literal goals, we must have a clear notion of what it takes for a referring act to be satisfied if we want to view referring as a planned speech act.

Note that specifying the conditions of satisfaction of a referring act is not the same as specifying its literal goal. The literal goal of a speech act and its conditions of satisfaction are usually distinct: if I tell you that I want the door closed and you understand me, the literal goal of my request has been achieved. But it is still up to you whether or not you will satisfy my request.

Compositionality. Speech acts can be combined with one another, creating new and more complex speech acts. For example, a question can be seen as a request to inform. It is also possible to either request a promise or promise to request. As Cohen and Perrault [7] have shown, the appropriate planning of composite speech acts is a powerful adequacy test for a planning system that can generate speech acts. Now, since referring is not a complete speech act (that is, it is not an illocutionary act), one cannot refer to promise or refer to request (although one can refer to a *particular* promise or request). However, one can certainly request or promise a hearer that referring will be done, or refer and, at the same time, make an indirect request ("The envelope, please?"). Combining such speech acts is no less an adequacy test for a planning system than is the composition of *request* and *inform*. Viewed from this standpoint of speech act compositionality, referring is not different, in principle, from illocutionary acts.

But propositional acts (in particular, referring) *are* different from illocutionary acts. Seen from a certain perspective, propositional acts are related to illocutionary ones as the structure and meaning of noun and verb phrases are related to the structure and meaning of sentences. Referring and predicating are the building blocks out of which illocutionary acts are constructed; in pragmatics, as in syntax and semantics, it must be shown how the whole is a function of its parts. One way of stating the problem is in terms of *pragmatic presuppositions*. The pragmatic presuppositions of a speech act can roughly be described as the class of propositions that is characteristically associated with the felicitous performance of that speech act. The truth of these propositions is mutually believed to be taken for granted by the participants [12]. Now, it is difficult to see how such a class of pragmatic presuppositions is generated unless the pragmatic pre-

suppositions associated with illocutionary acts are largely a function of the pragmatic presuppositions associated with *parts* of the illocutionary acts. For example, a pragmatic presupposition of the command "Show me the letter!" is that it is mutually believed that a certain letter exists and that both speaker and hearer know which one it is. This pragmatic presupposition is generated, in turn, through other presuppositions associated with the referring act: for example, that it is mutually believed that the use of the definite article in this case signals, say, an anaphoric link with a referring expression mentioned earlier.

In a planning system generating illocutionary acts, there would be operators whose executions would correspond to the performance of a particular speech act. Let us say that **REFER** is the operator corresponding to the referring act. In such a system, the pragmatic presuppositions generated by the act of referring will be represented by the mutually known effects of the execution of **REFER**. One way of capturing the relation between referring and complete speech acts, then, would be to show how the mutually known effects of the referring act contribute to the mutually known effects of illocutionary acts.

Syntax and semantics. In [direct] illocutionary acts, we have a fairly precise correlation between syntax and semantics, on the one hand, and illocutionary point on the other. Assertions and commands, for example, have their syntactic counterparts in indicative and imperative sentences, while performative verbs represent those illocutionary acts that are being performed. But whereas a serious utterance of an imperative sentence is almost always taken as a directive type of speech act, the serious utterance of a noun phrase – even a definite noun phrase – is not necessarily an act of referring, as we have already seen. Similarly, one can promise, say, to pay one's debt by stating "I hereby promise to pay my debt," but merely uttering "I hereby refer to a friend of mine" is hardly satisfactory. Hence, the semantic and syntactic clues that enable the hearer to recognize an illocutionary act do not help much as far as referring is concerned.

Thus, we have four problems with respect to referring acts that seem harder to resolve than their counterparts in a theory of illocutionary acts:

1. What is the literal goal of a referring act?
2. What are its conditions of satisfaction?
3. How does referring contribute to the success of illocutionary acts?
4. When is a use of a noun phrase a referring use?

Answers to the first and second questions are given in [15]. To a limited degree, Appelt and Kronfeld [3] address the third question: we show how changes that take place in a hearer's mental state as a result of a referring act contribute to other changes resulting from a broader speech act (in particular, a request). The fourth question has two parts. First, there is the problem of specifying the algorithm by means of which a hearer recognizes a noun phrase as a referring expression. This is an extremely difficult question and I shall not be discussing it in this paper. Second, there is the problem of specifying the mental state that underlies a speaker's use of a noun phrase as a referring expression. In essence the solution to this problem amounts to a definition of referring. Such a definition can be found in [3].

4 Philosophical foundation

Let us return to the external perspective. From its standpoint we have the object that the speaker is thinking of, we have the referring expression used by him, and we ask how a hearer makes the connection. Once the question is formulated in these terms, however, it is easily apparent that it masks a more general problem. Never mind how the hearer recognizes the connection between a referring expression and an object. How is this connection established in the first place? What does it mean to say that the speaker has a particular object in mind? Does it mean that he is able to identify that object when he sees it? Does it mean that he knows something that is true of that particular object and no other?

Such questions lead to what may be called the philosophical problem of *reference*, which can roughly be phrased as follows: "How can thoughts (and sentences that articulate them) be *about* objects?" The problem seems simple enough, but, as was the case with the *referring* problem (i.e., how do we let our audience know what we are talking about?), the simplicity is deceptive. In affect a solution to the problem of reference elucidates the general mechanism that enables the mind (and, derivativly, language) to represent the world for us. It is not surprising, therefore, that the problem of reference has occupied a central position in the philosophical debate that has been going on since the very beginning of this century. The question is whether anyone interested in a computational model of referring should get involved in the problem of reference, philosophical baggage and all.

What are the options? It should be obvious that the referring problem and the problem of reference are not mutually independent. We could avoid philosophy altogether and start from scratch. This has been done in AI, on occasion — with lamentable results, however. On the other hand, philosoph-

ical debates drag on forever. Major philosophical problems may go in and out of fashion, but they are never really "solved." If a computational model is what we are after, it would be hopeless for us to wait for the philosophical discussion of reference to reach a consensus.

The way out of this impasse is to recognize that the philosophy of language and mind can offer *research programs*. The term was introduced by Lakatos [17], as a rational synthesis of Kuhn's notion of a scientific paradigm and Popper's principle of falsification [16,18]. It essentially denotes a general scientific framework that offers a methodological foundation for investigating certain scientific problems. When I say that philosophy can provide research programs for the study of language, I mean something very much like what Lakatos had in mind: a theoretical framework that, when stated explicitly, is invaluable as a source of general principles. Thus, rather than ignoring the philosophical debate or attempting to maintain "neutrality," we should scan the philosophical landscape for the most promising general approach to the problem at hand. Once we find it, we should isolate its central theses, systematically identify the main objections to it, and evaluate the possibility of modifying the program as to overcome these objections. If the program still looks promising, we should stay with it, using it as a general guide for computational research. I present such a philosophical framework in [14].

References

[1] James F. Allen. *Recognizing Intention in Dialogue.* PhD thesis, University of Toronto, 1978.

[2] Douglas E. Appelt. *Planning English Sentences.* Cambridge University Press, Cambridge, England, 1985.

[3] Douglas E. Appelt and Amichai Kronfeld. A computational model of referring. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 640–647, Milan, Italy, 1987.

[4] Noam Chomsky. *Knowledge of Language.* Preager, 1986.

[5] Philip R. Cohen. *On Knowning What to Say: Planning Speech Acts.* PhD thesis, University of Toronto, 1978.

[6] Philip R. Cohen and H. Levesque. Speech acts and rationality. In *Proceedings of the 23rd Annual Meeting*, pages 49–59, Association for Computational Linguistics, 1985.

[7] Philip R. Cohen and C. Raymond Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:117–212, 1979.

[8] H. Paul Grice. Meaning. *Philosophical Review*, LXVI(3):377–388, 1957.

[9] Hans Kamp. A theory of truth and semantic representation. In Groenendijk et. al., editor, *Truth, Interpretation, and Information*, Foris, Dordrecht, Netherlands, 1984.

[10] Asa Kasher. Conversational maxims and rationality. In A. Kasher, editor, *Language in Focus*, pages 197–216, D. Reidel Publising Co., Dordrecht-Holand, 1976.

[11] Asa Kasher. Gricean inference revisited. *Philosophica*, 29(1):25–44, 1982.

[12] Asa Kasher. Philosphy and discourse analysis. In *Handbook of Discourse Analysis, Vol. 1*, chapter 9, pages 231–248, Academic Press, Inc., 1985.

[13] Asa Kasher. What is a theory of use. *Journal of Pragmaics*, 1:105–120, 1977.

[14] Amichai Kronfeld. *The descriptive approach to reference: why it is difficult to live with, and why we have to.* Technical Note, Artificial Intelligence Center, SRI International, 1988.

[15] Amichai Kronfeld. *The literal goal and discourse purpose of referring.* Technical Note, Artificial Intelligence Center, SRI International, 1988.

[16] Thomas Kuhn. *The Structure of Scientific Revolutions.* Chicago, 1962.

[17] Imre Lakatos. Falsification and the methodology of research programmes. In I. Lakatos and A. Musgrave, editors, *Criticism and the Growth of Knowledge*, Cambridge University Press, 1970.

[18] Karl R. Popper. *The Logic of Scientific Discovery.* Hutchinson and CO (publishers) LTD, 1959.

[19] John R. Searle. *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press, Cambridge, England, 1983.

[20] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge, England, 1969.

[21] John R. Searle. A taxonomy of illocutionary acts. In *Expression and Meaning: Studies in the Theory of Speech Acts*, pages 1–29, Cambridge University Press, Cambridge, England, 1979.

[22] Peter F. Strawson. *Individuals.* Methuen, 1959.

[23] Bonnie L. Webber. So what can we talk about now? In M. Brady and R. Berwick, editors, *Computational Models of Discourse*, pages 331–371, MIT Press, Cambridge, Massachusetts, 1983.