

SRI International



THE LITERAL GOAL AND DISCOURSE PURPOSE OF REFERRING

Technical Note 439

May 16, 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.

The Literal Goal and Discourse Purpose of Referring*

Amichai Kronfeld

May 16, 1988

Contents

1	Introduction	2
2	The literal goal of referring	4
3	The discourse purpose of referring	8

*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.

1 Introduction

Suppose I ask a guest of mine to take out the garbage. In making this request, I may have a whole range of intentions and goals that have motivated my utterance. For example, I may have intended to make my visitor feel at home, to avoid doing something I hate, to get the house cleaned, and so on. However, not all of these objectives are relevant to our discussion. What we are after are *communication* intentions and goals — that is, goals that are intended to be *recognized* or, more precisely, to be achieved (at least in part) through their recognition. For example, deception is not a communication goal because it is not (and could not be) intended to be recognized. A speaker may very well intend the hearer to recognize the speaker's attempt to eliminate boredom, but the elimination of boredom is still not a communication goal because its mere recognition would hardly contribute to making things more interesting. In contrast, my goal of obtaining some information from you is indeed a communication goal because the most efficient plan to achieve it would include letting you know what it is. Here not only do I intend that you recognize my need for this information, but I also intend that your recognition play a role in inducing you to provide me with it. Such goals, along with the intentions to satisfy them, are the communication goals and intentions that are of interest to us.

This view of communication intentions originates with Grice's analysis of the concept of meaning [7]. But much research in computational linguistics, though obviously influenced by Grice, has nevertheless stressed the role of intention and goal recognition in discourse, quite independently of a theory of meaning. Allen's dissertation [1] and his subsequent work with Perrault [3,2], for example, emphasize the importance of goal recognition for inferring the speaker's plans. So do Grosz and Sidner [10], as well as Sidner herself in her own work [22,21]. These authors show how the recognition of what the speaker is "up to" contributes to coherence and comprehensibility of the discourse (see also [18]), and is essential for the hearer's generation of an appropriate response. However, not all goals that are intended to be recognized are alike.

Sometimes the recognition of a goal suffices for its satisfaction. For example, if my objective is to congratulate you, I can succeed if (and in this particular case, only if) you recognize my intention to do so. Once you have recognized my intention, you are thereby congratulated. I call such goals, whereby recognition is sufficient for success, the *literal goal* of the speech act. In addition, there are *discourse purposes*, which are the goals underlying both the choice to engage in discourse in the first place (rather than in a nonlinguistic activity) and the choice of a particular propositional

content to be expressed [10]. In the case of congratulating, the literal goal and the discourse purpose are one and the same — but this is the exception rather than the rule. For example, in the case of the guest who is asked to take out the garbage, the literal goal of the request is essentially to make the guest *realize* what he is asked to do. The discourse purpose is to make him actually *do* it. Note that, unlike the case of literal goals, recognition of the discourse purpose, albeit vital for the success of the speech act, is not enough for the objective to be achieved. The guest may very well recognize that my purpose is to induce him to take out the garbage, but, alas, this in itself does not guarantee his cooperation. So the recognition of literal goals is sufficient for its success (sometimes it is also necessary), while the recognition of discourse purposes, as distinct from literal goals, is neither sufficient (it does not guarantee cooperation) nor necessary (my guest may take the garbage out without realizing that he is complying with my wishes). Rather, given certain assumptions about the disposition of discourse participants, it is *rational* to expect that hearer's recognition of a discourse purpose will enhance the speaker's chances of achieving his goal.

The discourse purpose should characterize the *point* of the institution of referring and, consequently, determine under what conditions an act of referring can be called a success. In this sense, the discourse purpose of referring is analogous to Searle's illocutionary point. However, melding Grosz and Sidner's notion of discourse purpose with Searle's notion of the point or purpose of a speech act masks a subtle difference between the two concepts. In Grosz and Sidner's theory, discourse purposes are *perlocutionary* [4]: they are defined in terms of effects on the *hearer's* beliefs and actions. Searle, on the other hand, insists that "the terminology of 'point' or 'purpose' is not meant to imply, nor is it based on the view, that every illocutionary act has a definitionally associated perlocutionary intent" [20, 3]. But the difference between discourse purposes and illocutionary points should not concern us too much. Grosz and Sidner's focus on perlocutionary effects is a natural consequence of their computational approach and, is quite typical of the research in their field. Searle's eschewal of any general interpretation of illocutionary points as perlocutionary effects is consistent with his rejection of Grice's theory of meaning [19, sec. 2.6]. Searle, of course, does not deny that some speech acts do have perlocutionary intentions associated with them (for example, requests), and I take referring to be one of them. If we regard it intuitively, the point of referring is to make the hearer identify an object — a perlocutionary effect *par excellence*.¹

Literal goal and discourse purpose are obviously not independent of each

¹It is not surprising that Cohen [5] views referring as a request for identification.

other: my expectation that my guest will actually carry out the garbage depends partially on my expectation that he will understand what I have said. This is indeed the basis for our use of language as a means for attaining some of our objectives, and is the foundation for a plan-based theory of speech acts. For example, the distinctions between surface speech acts and illocutionary acts [2], and between utterance-level and discourse-level intentions [10], are, in my opinion, particular computational realizations of the distinction between literal goals and discourse purposes. Since I accept the plan-based approach and consider referring to be a speech act, I must specify the literal goals and discourse purposes that typically motivate the use of referring expressions. Needless to say, the literal goal and the discourse purpose must be characterized in a way that is useful for the referring model. That is, they must be spelled out in a vocabulary that makes computational sense.

2 The literal goal of referring

Let us begin with Grice's theory of communication intentions. The motivation for Grice's philosophical project is an attempt to analyze semantic concepts in terms of propositional attitudes. In a nutshell, Grice begins by distinguishing between natural and nonnatural meaning (as well as between natural and nonnatural signs). Roughly defined, natural meaning is the meaning of *symptoms*: smoke is a natural sign of fire, spots on the skin are a natural sign of measles, and when we say that smoke in the house means that there is fire somewhere, or that spots on the patient's skin mean that he has measles, the sense of "mean" here is a *natural* one.

In contrast, when we say that sentences have meaning and that agents can mean something by uttering them, the concept of "meaning" here is a *nonnatural* one: the utterance "the patient has measles" means that the patient has measles, but the utterance, unlike the spots, is surely not a symptom of the disease.

The conceptual dependency in Grice's theory between what utterances mean (what Grice calls *timeless meaning of an utterance type*) and what speakers mean when they use these utterances (*utterer's occasion meaning*) is roughly as follows. The statement

.1 *Expression E means P*

is to be understood in terms of the statement

.2 *In uttering expression E, speaker S means that P,*

where the analysis of (.2) is to be founded on the analysis of the statement

.3 *S*, in performing act *A*, means something.

Now, the keystone of Grice's analysis of meaning *something* — in fact, the foundation of his entire approach — consists of three intentions that are supposed to be both necessary and sufficient for a speaker to mean anything [7]:

Int₁: *S* intends to produce an effect in hearer *H*.

Int₂: *S* intends *H* to recognize his intention Int₁.

Int₃: *S* intends that intention Int₁ be satisfied by means of *H*'s recognition of intention Int₂.

In other words, if *S* is to mean anything at all, he must intend to produce an effect in *H* *by means of H's recognition of this intention*. The profound insight contained in Grice's paper is the discovery of this defining feature of communication. It is the observation that communication necessarily involves intentions whose very recognition suffices to satisfy them.

As a theory of meaning, Grice's original account was burdened by a number of difficulties; confronted by a series of counterexamples, he was compelled to modify his three intentions substantially [6]. However, my interest here is not in a theory of meaning. Whether Grice's account is, or could be made to be, a correct analysis of the concept of meaning is debatable, but quite apart from that, his original insight about the nature of communication intentions and goals remains valid. My point of departure, therefore, is Grice's concluding remarks in his 1969 essay:

I see some grounds for hoping that, by paying serious attention to the relation between nonnatural and natural meaning, one might be able not only to reach a simplified account of utterer's occasion-meaning, but also to show that *any human institution, the function of which is to provide artificial substitutes for natural signs, must embody, as its key-concept, a concept possessing approximately the features which I ascribe to the concept of utterer's occasion-meaning*. (Ibid., 177. Italics mine.)

Now, whether or not referring is a special case of utterer's occasion meaning, clearly the prime function of the linguistic institution of referring is to provide an artificial substitute for natural signs. The whole point of referring is that one does not need to carry around samples of the things one wants to talk about. So, if Grice is right, referring intentions must include communication intentions with a structure corresponding rather closely to

the Gricean picture. Actually, it can be argued that the institution of referring is more essentially a *communication* institution than its relatives, the illocutionary acts. Communication requires at least two agents, i.e., our ubiquitous speaker and hearer; in this respect Grice's theory of meaning has been criticized for requiring the presence of an audience for a person to mean anything. Surely, it is argued, a person could correctly be said to have meant something without addressing anyone in particular (for example, when making entries in a diary or scribbling notes while trying to prove a mathematical theorem). Under such circumstances, an agent may mean something without intending to produce an effect in anyone through recognition of his intention. How damaging this argument is to Grice's theory is unclear. In any case, such an assessment is not crucial to our purpose because, unlike asserting, promising, congratulating and the like, referring *does* require an audience. I can congratulate myself for being such a nice fellow, promise myself to go on a diet next week, tell myself that the situation in Israel is terrible, and so on. But there seems to be little point in *referring*, unless I do it for someone else. Of course, when I mutter to myself "The situation in Israel is terrible" I am, in a sense, referring to the situation in Israel, but this type of "referring" is substantially different from performing the same act for a hearer. To see why, consider what the intuitive point of referring is: to make the hearer identify what is being talked about by supplying him with adequate information. This is what I do when I tell *you* that the situation in Israel is explosive. But none of that is required when I talk to myself, since I already know what I am thinking about. Thus, it seems that referring is essentially a communication act in a way that illocutionary acts are not and, if this is so, one would expect a key referring intention to correspond approximately to Grice's original analysis. What we need to find, therefore, is the kind of referring intention whose fulfillment requires no more than its recognition. This intention would provide the *literal goal* of the referring act.

Once we adopt the descriptive approach to reference, such an intention is not hard to find. Consider what an intuitive account of referring looks like from a descriptive standpoint. A speaker has a mental representation denoting an object; by using a noun phrase that is intended to be interpreted as a linguistic representation of the object, the speaker intends to invoke in the hearer a mental representation denoting that very same object. But it should be noted that such an intention has precisely the Gricean quality we are looking for. Once the hearer recognizes the intention that he have a mental representation denoting the same object that the speaker has in mind, the hearer *does* have such a representation. For example, if I recognize your intention to convey to me a representation of whatever object you

are talking about, then I *do* have a representation of that object in the presentation mode: *the object the speaker is talking about*. In other words, *the mere recognition of the intention to represent an object suffices to produce a presentation mode of that object*. A central referring intention, therefore, is the intention to invoke a representation of a particular object in the hearer by means of his recognition of this intention. The goal of satisfying this intention is the literal goal of referring.

We can also describe the literal goal of referring differently, namely, in terms of features of noun phrases. As Grice points out

Characteristically, an utterer intends an audience to recognize ...some "crucial" feature *F* [of the utterance], and to think of *F* ... as correlated in a certain way with some response which the utterer intends the audience to produce. (Ibid., 163)

If the property of being a referring expression is taken to be a feature of some noun phrases, and if this feature is correlated with the invocation of a mental representation in the hearer's mind of whatever object the speaker is talking about, then we can say that the literal goal of referring is to have the hearer recognize that an utterance of a noun phrase is to be interpreted as a referring expression. Of course, noun phrases that are used as referring expressions have other features relevant to the act of referring. For example, when used as a referring expression, the noun phrase "the gray whale" has a feature that is conventionally correlated with the act of invoking in the hearer a representation of a gray whale. Thus, the literal goal of referring is not merely to invoke the representation *the particular object the speaker wishes to talk about*, but to invoke a representation of that object with properties that correspond to certain features of the noun phrase.

In [13], I defined an *individuating set* as an exhaustive list of presentation modes, all believed by the agent to denote the same object. Let us distinguish between two kinds of individuating sets: the *quasi-permanent* type and the *local* type. Quasi-permanent individuating sets are those more or less permanent representations of objects that we carry around in our mental data base. They can be formed and modified independently of any discourse situation and, in general, can be viewed as the data structures that allow us to *think* of things. *Local* individuating sets, on the other hand, are those that are constructed, modified, merged with other individuating sets, or discarded during an actual discourse. They are a subset of the class of discourse entities [12,23], figure prominently in participants's representations of the discourse focusing-structure [9,8,10]; and, by and large, reflect which objects are being discussed. The distinction between quasi-permanent and local individuating sets is not necessarily sharp, and there

are obviously many links between the two types: a local individuating set may turn into a quasi-permanent one when the conversation ends, and access to a quasi-permanent set may be required while a local one is being processed. Nevertheless, the distinction is useful very much in the same sense as the distinction between long-term and short-term memories.

Given the concept of a local individuating set as a data structure representing a particular object during discourse, we can restate the literal goal of referring as follows. If we assume a speaker's intention to refer to an object by using a noun phrase, the literal goal of his referring act is to make the hearer generate a local individuating set that determines that object by virtue of the hearer's recognition that the noun phrase is to be interpreted as a referring expression. In computational terms, this means that when it is recognized that the noun phrase is to be interpreted as a referring expression, the hearer generates a local individuating set containing a single presentation mode, namely,

- .4 *The one and only x such that x is the ... that the speaker wants to say something about*

The ellipsis in (.4) is to be filled by the descriptive content of the referring expression. For example, if the noun phrase is "the gray whale," the corresponding element in the hearer's newly created local individuating set is

- .5 *The one and only x such that x is the gray whale that the speaker wants to say something about.*

Naturally, some noun phrases do not have any descriptive content conventionally associated with them (e.g., "this"). In such cases, the generated presentation mode is simply the *object* the speaker wants to say something about. Thus, it is important to note that the descriptive content of referring expressions is not at all indispensable for the initial individuation of the referent. The literal goal of referring can be achieved even if the descriptive content denotes nothing, is applied incorrectly, or is lacking altogether.

3 The discourse purpose of referring

Success in achieving the literal goal of referring is hardly enough. "And then we'll go out, Piglet, and sing my song to Eeyore." says Pooh. "Which song, Pooh?" asks Piglet. "The one we're going to sing to Eeyore" explains Pooh [17, p. 3]. Pooh has certainly succeeded in achieving his literal goal, but Piglet is still puzzled. He is yet to *identify* the song Pooh is talking about.

Referent identification is the discourse purpose of referring. But the concept of identification itself is so vague and ambiguous that it is of little help to a computational approach. Let us begin, therefore, with two important distinctions.

First, the speaker's sense of "identify" is quite different from that of the hearer. A speaker is said to identify an object *for* or *to* a hearer, while the hearer is said to identify that object if the referring act is successful. Searle [19], paying scant attention to the hearer's sense, bases his account of reference as a speech act on the speaker's sense of referent identification. He is, after all, interested first and foremost in showing how language use is a structured, rule-governed behavior, and he wants to show what it takes for a speech act to be performed *feliculously*, i.e., according to the rules. He is less interested in what makes a speech act a *success*, namely, what allows it to achieve its intended effects. As regards a computational model within the plan-based approach, on the other hand, effects on the hearer's mental state are really what counts. If no intended effects are formulated, there is little point in planning. Thus, in what follows I shall concentrate on the hearer's sense, taking the speaker's sense simply to be his plan to achieve the hearer's identification of the referent. Of course, Searle's notion of language as a rule-governed behavior still holds: conformity to the institutional rules of referring would constitute an important step in this plan.

Second, identification, as the discourse purpose of referring, should be carefully distinguished from identification in the sense of *knowing who* someone is. The former may be considered a *pragmatic* notion of identification. The latter is an *epistemic* one. Although the pragmatic notion of identification is also connected with knowledge (in the sense of knowing who the speaker is talking about), the two are quite distinct. To illustrate the difference, consider the example in which the detective, referring to Smith's murderer, says "The man whose fingerprints these are, whoever he is, is insane." Neither speaker nor hearer in this case knows who Smith's murderer is, and they cannot "identify" him if identification is interpreted epistemically. From a pragmatic point of view, however, there is a clear dichotomy: if the hearer makes the connection between "the man whose fingerprints these are" and "Smith's murderer," he has identified the person the speaker is talking about. Otherwise he has not.

The pragmatic and the epistemic senses of identification are obviously closely related conceptually, and the distinction between them is not as sharp as my example might suggest. Nevertheless, the distinction is important. As I have argued elsewhere [16], part of the argument against a descriptive theory of referring is that the notion of identification, on which such a descriptive account is founded, cannot be applied uniformly to both "ref-

erential" and "attributive" uses of definite descriptions. In the referential use, it is argued, there is an object to be identified. In the attributive, on the other hand, there is none. But the sense of "identification" in this argument is clearly epistemic. It pertains to the question of whether speaker and hearer know, in some manner, *who* the person the speaker is talking about is (e.g., who murdered Smith). However, the sense of "identification" that a descriptive theorist regards as being common to *all* successful uses of referring expressions (be they "referential" or "attributive") should be a pragmatic one. It should indicate when the hearer's grasp of what the speaker is talking about is adequate, *relative to the speaker's goals*. In planning terms: if, in the course of the speaker's attempt to achieve his objectives, it becomes clear that, as a precondition for further steps in the plan, he must achieve certain changes in the hearer's state of mind, and if the speaker believes that, by a successful act of referring, he can effect such changes, then these changes define the pragmatic notion of referent identification, i.e., the discourse purpose of referring. In other words, the hearer's sense of the pragmatic notion of referent identification is to be defined in terms of the appropriate changes in hearer's beliefs regarding a particular object that the speaker has in mind. These changes are "appropriate" relative to the goals of the speaker. Whether the hearer should ultimately come to know *who* the referent is does not necessarily have to be part of these goals (although sometimes it may be).

Not having to deal with the complicated concept of *knowing who* is certainly a relief, but we are hardly in the clear as yet. The pragmatic notion of identification seems as elusive as the epistemic one. Consider the following examples:

- .6 (a) *Look at this fellow move! He must be doing at least a five-minute mile!*
 (b) *Do you remember the little playwright who used to hang out with us in the sixties? He has just won the Pulitzer prize.*
 (c) *Tell me what other plays were written by Shakespeare.*
- .7 (a) *Your friend has just won \$10,000.*
 (b) *My friend has just won \$10,000.*
 (c) *A friend of mine has just won \$10,000.*

When we try to formulate the conditions under which a hearer can be said to identify the referents correctly in these examples, we find that in each

case there is a different standard. In Sentence .6 (a), the hearer is asked to identify the runner visually, but visual identification of Shakespeare is clearly not required in Sentence .6 (c), although the hearer is still expected to identify the referent in some other way. In Sentence .6 (b), locating the playwright in one's field of vision is also not called for, but the requirements for correct identification are surely different from Shakespeare's case. In the latter situation, the hearer needs to associate the name with a store of shared cultural knowledge. In the case of the talented playwright, a connection between the description and, perhaps, a memory image is essential.

Not only do methods of identification differ from utterance to utterance, but there are also differences in expected *degrees* of identification (analogous, perhaps, to variations in degrees of illocutionary force). In Sentence .7 (a), the speaker would usually expect the hearer to know (or inquire) which friend the speaker is talking about. In uttering .7 (c), on the other hand, the speaker usually signals that the hearer need not know which friend the speaker means. All that the hearer is expected to do is to note the existence of a particular individual, a friend of the speaker, who is lucky enough to have won such a large sum of money. The identification requirements in .7 (b) may be similar to either .7 (a) or .7 (c), depending on contextual factors and on the purpose of the conversation.

Obviously, there is not just one "correct" method of referent identification. Still, we can look for a way to express the differences as variations in applying a single mechanism. If the reason for referent identification is the need to establish mutual agreement as to which object is being talked about, then a necessary precondition for successful referring is that the hearer understand the ground rules for reaching such an agreement. In general, such ground rules follow from the propositional content of the illocutionary act itself, general knowledge about the discourse (in particular, the speaker's goals), and principles of rational behavior. For example, according to any theory of action, if the speaker asks the hearer to pick up an object, it would have to follow that the hearer can comply with the request only if he can perceive the object he is about to manipulate. This fact implies a rather specific mode of identification: locating the object in one's immediate vicinity. Different ways to identify the referent are similarly derived in other circumstances. However, if we accept the descriptive approach [16], we are committed to expressing or formalizing the rules for pragmatic identification *in terms of presentation modes*. In other words, the assumption is that a necessary and sufficient condition for a hearer to identify who or what the speaker is talking about is that the hearer come to think of the referent in terms of one or more appropriate presentation modes. For now, let us not trouble ourselves as to what an "appropriate" presentation mode is. The

important point is that pragmatic identification is regarded as *entirely* a matter of having a representation whose descriptive content denotes the referent and which is appropriate in a sense to be explained. Of course, this does not mean that the appropriate presentation mode required for pragmatic identification can always be verbalized as a definite description. When I want you to locate an object in your vicinity, I want you to see it (or touch it, etc.). Now, perception is as good a source of presentation modes as any, but obviously, I do not necessarily expect you to translate what you see into words.

How can we begin to represent such a view in our model? If the hearer is cooperative, his pragmatic identification of the referent commences once the literal goal has been achieved. At this point, the hearer thinks of the referent in terms of a single presentation mode that is guaranteed to denote whatever the speaker has in mind. What happens next depends on circumstances, but it would be wrong to assume that there is always a unique representation upon which alone the hearer's identification succeeds or fails. For example, suppose that during a discussion of German theater I mention Bertolt Brecht to you. I believe that you have an individuating set of presentation modes associated with the name "Bertolt Brecht" that allows you to understand who it is I am talking about. But, although I hope that a significant segment of your individuating set would be rather similar to mine, there is usually no *particular* presentation mode that must be there, nor, of course, do I expect your individuating set to be identical to the one I have. What I expect, rather, is that your individuating set satisfy certain *constraints*: it should include, for example, a sufficient number of individuating facts concerning important plays that Brecht wrote, but the list does not have to correspond exactly to my list (nor does it have to be exhaustive).

Thus, instead of representing pragmatic identification as a relation between a hearer and a presentation mode, it is better to represent it as a relation between a hearer and a pair of formal entities: an *individuating set* and a set of *identification constraints*. If the literal goal of referring is to make the hearer generate an individuating set that contains a presentation mode of the referent, then the discourse purpose of referring is first, to make the hearer understand what identification constraints are operative, and second, to make him apply these constraints to the newly generated individuating set. Possible identification constraints may include the following:

- A requirement that the relevant individuating set should contain a *perceptual* presentation mode, to be acquired now or later.
- A requirement that the hearer should be able to merge the newly generated local individuating set with a pre-existing, quasi-permanent one

(additional identification constraints may be required so as to ensure that the latter set will be adequate for current purposes — see below).

- A requirement that the hearer be able to connect the new local individuating set with an existing *local* one. In particular, this includes such cases as “An old girlfriend of mine got married yesterday. *The lucky man* met her only three weeks ago.”²
- An indication that identification is not expected *yet*, but that the hearer should “stay tuned”: “Although *the lucky bastard* does not deserve it, John got a raise last week.”
- A requirement that the relevant individuating set contain one or more presentation modes that are *privileged* with respect to the goals of the speaker. Suppose Art asks Ben: “Do you think *Ronald Regean* will send the marines to invade Nicaragua?” Ben may have a very rich individuating set for Regean. He may know, say, each and every detail of Regean’s Hollywood career. But if this individuating set does not contain a presentation mode such as *current president of the United States* or *commander in chief*, pragmatic identification has not been accomplished.
- The null identification constraint, under which success of the literal goal is already sufficient for pragmatic identification. “*An old friend of mine* once told me that ...” is an example. Since the speaker merely indicates the existence of an old friend whose actual identity is irrelevant, no further identification is necessary once the literal goal has been achieved.

This view of pragmatic identification in terms of individuating sets and constraints upon them is but a tentative suggestion at this stage. From a theoretical standpoint, such a view depends on an aforementioned central descriptive thesis — namely, that to a large extent individuating representation is both necessary and sufficient for a belief to be *de re*. A defense of this thesis can be found in [15]. From a computational standpoint, this approach to pragmatic identification awaits careful specification and formalization of identification constraints and their derivation and satisfaction by a hearer — not an easy task by any means. But an important advantage of this view

²I intentionally refrain from using the term “anaphora” here. Although anaphora resolution is essential for pragmatic identification, it belongs to the *internal* perspective (see [14]) and is a much more general phenomenon. The two should be kept apart, at least conceptually. The same goes for cataphora [11], i.e., “backward anaphora,” as shown in the next example of an identification constraint.

can already be pointed out: it enables us to eliminate the *standard-name assumption*.

All natural-language systems I am aware of assume that every object in the domain has a standard name that is known to everyone. In these systems, referent identification is accomplished when (and only when) the system associates the referring expression that is being processed with the right standard name. This is undoubtedly quite useful if the system is expected to handle only a small number of objects in a very limited way; as a general principle, however, the standard-name assumption is obviously too strong. It is unreasonable to assume that every object that can be talked about has a universally known standard name. A doctor and a nurse, for example, can discuss a patient's left ear without bothering to give it a standard name. Similarly, we all know that the number 2^{64} has a standard name, but few of us know what it is. This does not prevent us from referring to this number — as, in fact, I just did.

Moreover, within the framework of the logic of knowledge and belief, the standard-name assumption has a rather nasty consequence, namely, that agents can never be wrong about the identity of objects they talk about. For example, under the standard-name assumption, it would have been impossible for Oedipus not to know that his mother was his wife. Since he could identify both, he must have had a standard name for each. The two names denoted the same thing in all possible worlds, including, of course, all possible worlds compatible with what Oedipus knew. Hence, Oedipus could not fail to know that his wife and his mother were one and the same. But surely an agent can be confused about the identity of an object while being perfectly capable of identifying it pragmatically. Representing pragmatic identification in terms of individuating sets and identification constraints solves this problem and, in general, eliminates the need for the standard-name assumption (although it does not *exclude* standard names). The question, of course, is whether the interpretation of pragmatic identification in terms of identification constraints can really be made to work computationally. This remains to be seen.

To sum up, here are the goals that motivate an act of referring: (1) the literal goal is to activate in the hearer's mind, by means of the hearer's recognition of this goal, a presentation mode denoting the referent; (2) the discourse purpose of referring is pragmatic identification. This latter goal is achieved, if the hearer is cooperative, in two steps: first, the hearer derives the appropriate identification constraints; he then attempts to identify the referent by applying those constraints to the set containing the initial presentation mode.

References

- [1] J. F. Allen. *Recognizing Intention in Dialogue*. PhD thesis, University of Toronto, 1978.
- [2] J. F. Allen and C. R. Perrault. Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143-178, 1980.
- [3] J. F. Allen and C. R. Perrault. Participating in dialogues: understanding via plan deduction. In *Proceedings*, Canadian Society for Computational Studies of Intelligence, 1978.
- [4] J. L. Austin. *How to Do Things with Words*. Oxford Univ. Press, London, 1962.
- [5] P. R. Cohen. Referring as requesting. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 207-211, 1984.
- [6] H. P. Grice. Utterer's meaning and intentions. *Philosophical Review*, 78:147-177, 1969.
- [7] H. P. Grice. Meaning. *Philosophical Review*, 66(3):377-388, 1957.
- [8] B. J. Grosz. Focusing and description in natural language dialogues. In A. Joshi, I. Sag, and B. Webber, editors, *Elements of Discourse Understanding*, pages 85-105, Cambridge Univ. Press, Cambridge, 1981.
- [9] B. J. Grosz. Focusing in dialogue. In *TINLAP-2*, pages 96-103, University of Illinois, Urbana-Champaign, 1978.
- [10] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175-204, July-September 1986.
- [11] M. A. K. Halliday and R. Hasan. *Cohesion in English*. Layman Press, London, 1976.
- [12] H. Kamp. A theory of truth and semantic representation. In T. M. V. Janssen J. A. G. Groenendijk and M. B. J. Stokhof, editors, *Truth, Interpretation, and Information*, Foris, Dordrecht, 1984.
- [13] A. Kronfeld. Donnellan's distinction and a computational model of reference. In *Proceedings of the 24th Annual Meeting*, pages 186-191, Association for Computational Linguistics, 1986.
- [14] A. Kronfeld. *Methodological Notes on a Computational Model of Referring*. Technical Note 434, Artificial Intelligence Center, SRI International, 1988.

- [15] A. Kronfeld. *Reference and Denotation: The Descriptive Model*. Technical Report 368, Artificial Intelligence Center, SRI International, 1985.
- [16] A. Kronfeld. *The descriptive approach to reference: why it is difficult to live with, and why we have to*. Technical Report 435, Artificial Intelligence Center, SRI International, 1988.
- [17] A. A. Milne. *The house at Pooh corner*. Methuen, London, 1965.
- [18] J. F. Allen Perrault, C. R. and P. R. Cohen. Speech acts as a basis for understanding dialogue coherence. In *TINLAP-2*, pages 125-132, University of Illinois, Urbana-Champaign, 1978.
- [19] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge Univ. Press, Cambridge, 1969.
- [20] J. R. Searle. A taxonomy of illocutionary acts. In *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge Univ. Press, Cambridge, [1975]1979.
- [21] C. L. Sidner. Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1):1-10, 1985.
- [22] C. L. Sidner. What the speaker means: the recognition of speakers' plans in discourse. *International Journal of Computers and Mathematics*, 9(1):71-82, 1983.
- [23] B. 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, Mass., 1983.