DONNELLAN'S DISTINCTION  
AND A COMPUTATIONAL MODEL OF REFERENCE

Amichai Kronfeld  
Artificial Intelligence Center  
SRI International  
333 Ravenswood Avenue  
Menlo Park, CA 94025  
kronfeld@sri-warbucks

Center for the Study of  
Language and Information  
Stanford University  
Stanford, CA 94305

INTRODUCTION

In this paper, I describe how Donnellan's distinction between referential and attributive uses of definite descriptions should be represented in a computational model of reference. After briefly discussing the significance of Donnellan's distinction, I reinterpret it as being three-tiered, relating to object representation, referring intentions, and choice of referring expression. I then present a cognitive model of referring, the components of which correspond to this analysis, and discuss the interaction that takes place among those components. Finally, the implementation of this model, now in progress, is described.

ABSTRACT

In this paper, I describe how Donnellan's distinction between referential and attributive uses of definite descriptions should be represented in a computational model of reference. After briefly discussing the significance of Donnellan's distinction, I reinterpret it as being three-tiered, relating to object representation, referring intentions, and choice of referring expression. I then present a cognitive model of referring, the components of which correspond to this analysis, and discuss the interaction that takes place among those components. Finally, the implementation of this model, now in progress, is described.
they of course interact with one another. The notion of a conversationally relevant description provides a basis for explaining how the interaction operates.

In the following sections, the three aspects are presented, their interactions discussed, and an initial attempt to achieve an implementation that takes them into account is described.

CRITERIA

How is the referential to be distinguished from the attributive? Two criteria are usually offered:

1. Even though, when used attributively, the description must denote the intended referent, in the referential use this is not necessary.

2. In the referential use, the speaker has a particular object in mind, whereas in the attributive he does not.

These criteria have been taken to be equivalent: any use of a definite description that is referential according to one criterion should also be classified as referential according to the other (and similarly for the attributive use). However, the equivalence of the two criteria is really an illusion: some uses of definite descriptions are referential according to one criterion, but attributive according to the other. For example, let us suppose that John, a police investigator, finds Smith's murdered body, and that there are clear fingerprints on the murder weapon. Now consider John's utterance: “The man whose fingerprints these are, whoever he is, is insane.” Note that John intended to speak of Smith's murderer, and he may very well have been successful in conveying his intended referent, whether or not the fingerprints indeed belonged to the murderer. Hence, according to the first criterion, the description, “The man whose fingerprints these are,” was used referentially. On the other hand, John did not have any particular person in mind. Hence, according to the second criterion, the description must have been used attributively.

Many, including Donnellan, regard the second criterion as the more significant one. But even this criterion is given two conflicting interpretations. On the one hand, “having a particular object in mind” is taken as an epistemic concept: this is not necessary. For example, let us suppose that John, a police investigator, finds Smith's murdered body, and that there are clear fingerprints on the murder weapon. Now consider John's utterance: “The man whose fingerprints these are, whoever he is, is insane.” Note that John intended to speak of Smith's murderer, and he may very well have been successful in conveying his intended referent, whether or not the fingerprints indeed belonged to the murderer. Hence, according to the first criterion, the description, “The man whose fingerprints these are,” was used referentially. On the other hand, John did not have any particular person in mind. Hence, according to the second criterion, the description must have been used attributively.

Thus, there are three aspects of Donnellan's distinction that should be carefully separated. These aspects can be represented in terms of three dichotomies:

- Using a definite description “the φ” to refer to whoever or whatever the φ may be, versus using “the φ” to refer to an object x, whether or not x is indeed the φ (the speech act distinction).

- Having knowledge of an object versus not having such knowledge (the epistemic distinction).

- Using a description as a rigid designator versus using it as a nonrigid one (the modal distinction).
who a certain person is (relative to purpose $P$), the relevant individuating set must include a privileged term determined by $P$, or that, for an agent to have knowledge of an object, the relevant individuating set must be grounded, and so on.

Since individuating sets are part of the database, this is where the epistemic distinction belongs.

**Referring Intentions**

A speaker may have two distinct types of referring intentions. First, he may select a particular term from the relevant individuating set, and intend this term to be recognized by the hearer. Second, the speaker may intend to refer to the object determined by an individuating set, without intending any particular term from the set to be part of the proposition he wants to express. Consider, for example, the following two statements:

1. **The author of Othello wrote the best play about jealousy.**

2. **Shakespeare was born in Stratford-upon-Avon.**

In making both statements, a speaker would normally be referring to Shakespeare. But note the difference in referring intentions between the two: in the first statement, the speaker selects a particular aspect of Shakespeare, namely, the fact that he is the author of *Othello*, and intends the hearer to think of Shakespeare in terms of this aspect. In the second statement, the speaker does not select any particular aspect of Shakespeare from the relevant individuating set. Indeed, he may not care at all how the hearer makes the connection between the name "Shakespeare" and the referent.

The two types of referring intentions yield two distinct types of propositions. When the speaker does not intend any particular aspect of the referent to be recognized by the hearer, the proposition expressed in this way is *singular*, that is, it does not contain any individual concept of the referent. Consequently, the referring expression chosen by the speaker (be it a proper name, a demonstrative, or even a definite description) is used as a rigid designator, which means that it picks out the same individual in all possible worlds where the referent exists. On the other hand, if a particular aspect of the referent is meant to be recognized by the hearer, then the individual concept corresponding to that aspect is part of the proposition expressed and should therefore be taken into account in evaluating the truth value of what is said. Thus, it is the speaker’s referring intentions that determine whether or not he will use a definite description as a rigid designator (the *modal* distinction). Since referring intentions are represented in the planner, this is where the *modal* distinction belongs.

Note that the two types of referring intentions can be described as intentions to place constraints on the way the hearer will be thinking of the referent. In Appelt and Kronfeld [2],

| DISTINCTION | INTERPRETATION | COMPONENT |
|-------------|----------------|-----------|
| Epistemic   | Type of individuating set | Database |
| Modal       | Type of referring intentions | Planner |
| Speech act  | Choice of definite noun phrase | Utterance generator |

Table 1: *Donnellan’s distinction, its interpretation(s), and the corresponding computational components.*

---

Kripke and Searle [12,15] explain the referential use as a case in which speaker’s reference is distinct from semantic reference. This leaves an important question unanswered: why must speaker’s reference and semantic reference coincide in the attributive use?\(^3\)

Sometimes two definite descriptions are equally useful for identifying the intended referent, yet cannot be substituted for each other in a speech act. The description employed, besides being useful for identification, has to be relevant in some other respect. Consider the utterance: “New York needs more policemen.” Instead of “New York,” one might have used “The largest city in the U.S.” or “The Big Apple,” but “The city hosting the 1986 ACL conference needs more policemen” won’t do, even though this description might be as useful in identifying New York as the others. The latter statement simply conveys an unwarranted implication.

As a generalization, we may say that there are two senses in which a definite description might be regarded as relevant. First, it has to be relevant for the purpose of letting the hearer know what the speaker is talking about.\(^4\) A description that is relevant in this sense may be called *functionally relevant.* Second, as the example above indicates, a description might exhibit a type of relevance that is not merely a referring tool.

---

\(^3\)As redefined by the *speech act* distinction.

\(^4\)Whether the hearer is also expected to identify the referent is a separate issue.
A description that is relevant in this noninstrumental sense might be called "conversationally relevant."

Every use of a definite description for the purpose of reference has to be functionally relevant. But not every such use has to be conversationally relevant. If indicating the referent is the only intended purpose, any other functionally relevant description will do just as well.

In other cases, the description is supposed to do more than just point out the intended referent to the hearer. Consider the following examples:

3 This happy man must have been drinking champagne.

4 The man who murdered Smith so brutally has to be insane.

5 The winner of this race will get $10,000.

In these examples, the speaker implicates (in Grice's sense) something that is not part of what he says. In (3), it is implicated that the man's happiness is due to his drinking. In (4), it is implicated that the main motivation for believing the murderer to be insane is that he committed such a brutal homicide. The implicature in (5) is that the only reason for giving the winner $10,000 is his victory in a particular race. In all these cases, what is implicated is some relationship between a specific characteristic of the referent mentioned in the description and whatever is said about that referent. In such cases, it does matter what description is chosen, since the relevance is both functional and conversational. No other description, even if it identifies equally well, can be as successful in conveying the intended implicature.

The conversationally relevant description may not be mentioned explicitly, but rather inferred indirectly from the context. In the fingerprint example, the speaker uses the description, The man whose fingerprints these are, but the conversationally relevant description is nevertheless Smith's murderer.

Thus, there are three general ways in which a speaker may employ a referring definite description:

1. If the discourse requires no conversationally relevant description, any functionally relevant one will do. This covers all standard examples of the referential use, in which the sole function of the definite description is to indicate an object to the hearer.

2. If a conversationally relevant description is needed, the speaker may do either of the following:
   (a) Use the description explicitly. This is what is done in standard examples of the attributive use.
   (b) Use a different, functionally relevant description. The speaker can do so, however, only if the context indicates the aspect of the referent that corresponds to the conversationally relevant description. This explains the ambiguity of the fingerprint example. As the definite description uttered is only functionally relevant, its use appears to be referential. Yet, unlike the referential case, a conversationally relevant description is implied.

In sum, when the description used is conversationally relevant, the speaker intends that the specific way he chose to do his referring should be taken into account in interpreting the speech act as a whole. Consequently, if the description fails, so does the entire speech act. On the other hand, if the description is only functionally relevant, the context may still supply enough information to identify the intended referent.

INTERACTIONS

When a speaker plans a speech act involving reference to an object, he must determine whether or not a conversationally relevant description is needed. However, the nature of the individuating set, on the one hand, and constraints on choices of referring expressions, on the other, may influence the speaker's planning in various ways. For example, if the individuating set contains only one item, say, the shortest spy, the definite description "the shortest spy" must be conversationally relevant. This is true both on formal and pragmatic grounds. From a formal standpoint, the description is conversationally relevant by default: no other functionally relevant description can be substituted because no such description is available. From a pragmatic standpoint, the description "the shortest spy" is very likely to be conversationally relevant in real discourse, simply because all we know about the referent is that he is the shortest spy. Thus, whatever we may have to say about that person is very likely to be related to the few facts contained in the description.

Even if it is clear that a conversationally relevant description is needed for the speech act to succeed, constraints on choices of referring expressions may prevent the speaker from using this description. One such constraint results from the need to identify the referent for the hearer. If the conversationally relevant description is not suited for identification, a conflict arises. For example, in "John believes Smith's murderer to be insane," the speaker may be trying simultaneously to represent the content of John's belief and to identify for the hearer whom the belief is about. Sometimes it is impossible to accomplish both goals with one and the same description.

IMPLEMENTATION

This paper is part of an extensive analysis of the referential/attributive distinction, which I use in the construction of a general model of reference [13]. My ultimate research objective is to provide a computational version of the reference model, then to incorporate it into a general plan-based account of definite and indefinite noun phrases. An experimental program that implements individuating sets has already been written. Called BERTRAND, this program interprets a small subset of English statements, and stores the information in its database, which it then uses to answer questions. Individuating sets are represented by an equivalence relation that holds among referring expressions: two referring expressions, R1 and R2, belong to the same individuating set if, according to the information interpreted so far, R1 and R2 denote the same object. In constructing individuating sets, BERTRAND uses a combination of logical and pragmatic strategies. The logical strategy exploits the fact that the relation "denote the same object" is symmetric, transitive, and closed under substitution. Thus, it
can be concluded that two referring expressions, \( R_1 \) and \( R_2 \), denote the same object (belong to the same individuating set) in one of the following ways:\^4

1. Directly, when the statement "\( R_1 \) is \( R_2 \)" (or "\( R_2 \) is \( R_1 \)"") has been asserted.

2. Recursively using transitivity — i.e., when, for a referring expression \( R_3 \), it can be shown that \( R_1 \) and \( R_3 \), as well as \( R_2 \) and \( R_3 \), belong to the same individuating set.

3. Recursively using substitution — i.e., when \( R_1 \) and \( R_2 \) are identical, except that \( R_1 \) contains a referring expression \( \text{sub}R_1 \) exactly where \( R_2 \) contains a referring expression \( \text{sub}R_2 \), and \( \text{sub}R_1 \) and \( \text{sub}R_2 \) belong to the same individuating set.

Note that, in the logical strategy, it is tacitly assumed that the relation of denoting the same object always holds between two identical tokens of referring expressions. This is obviously too strong an assumption for any realistic discourse: for example, two utterances of "The man" may very well denote two different people. On the other hand, the logical strategy fails to capture cases in which it is implied (although never actually asserted) that two distinct referring expressions denote the same thing. For example, "I met Marvin Maxwell yesterday. The man is utterly insane!"

To compensate for these weaknesses, BERTRAND uses a strategy based on Grosz’s notion of "focus stack" [8,10]. In conceptual terms (and without going into details), it works as follows: a stack of individuating sets, representing objects that are "in focus," is maintained throughout the "conversation." When a new referring expression is interpreted, it is transformed into an open sentence \( D(z) \) with a single free variable \( z \). An individuating set \( I \) is said to subsume an open sentence \( S \) if \( S \) can be derived from \( I \). The first individuating set in the focus stack to subsume \( D(z) \) represents the object denoted by the new referring expression. This solves the aforementioned problems: two occurrences of the same referring expression are considered as denoting the same object only if both are subsumed by the same individuating set in the focus stack, and two distinct referring expressions may still be considered as denoting the same object even though the logical strategy failed to show this, provided that both are subsumed by the same individuating set.

Once the concept of an individuating set has been implemented, referring intentions can be represented as intentions to activate appropriate subsets of individuating sets. For example, the intention to use a conversationally relevant description can be represented as the plan to activate a subset of an individuating set that contains the term associated with the description. This is the topic of a current joint research effort with D. Appelt [2] to investigate the interaction that takes place between individuating sets and referring intentions.

\^4What belongs to an individuating set, of course, is not a referring expression but the logical structure associated with it. For the sake of simplicity, however, I do not make this distinction here.

\^5For example, "The man from the city by the bay" is transformed into \( \text{Man}(z)\&\text{From}(z,X) \)

where \( X_1 \) is an "internal symbol" associated with \( \text{City}(y)\&\text{By}(y,X) \), and \( X_1 \) is associated with \( \text{Bay}(z) \).

\[ \text{concept activation actions} \]

The next stage in the development of BERTRAND — the implementation of referring intentions — will be based on this research. In the final stage, individuating sets and referring intentions will be used to generate actual referring expressions.

**ACKNOWLEDGMENTS**

This research was supported by the National Science Foundation under Grant DCR-8407238. I am very grateful to Doug Appelt and Barbara Grosz for detailed comments on earlier drafts, as well as to members of the Discourse, Intention and Action seminar at the Center for the Study of Language and Information for stimulating discussions of related issues.

**REFERENCES**

[1] Douglas E. Appelt. Some pragmatic issues in the planning of definite and indefinite noun phrases. In Proceedings of the 23rd Annual Meeting, Association for Computational Linguistics, 1985.

[2] Douglas E. Appelt and Amichai Kronfeld. Toward a model of referring and referent identification. Forthcoming. Submitted to the AAAI convention, Philadelphia, August 1986.

[3] Jon Barwise and John Perry. Situations and Attitudes. The Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1985.

[4] Philip R. Cohen. Referring as requesting. In Proceedings of the Tenth International Conference on Computational Linguistics, pages 207–211, 1984.

[5] Philip R. Cohen and Hector Levesque. Speech acts and the recognition of shared plans. In Proceedings of the Third Biennial Conference, Canadian Society for Computational Studies of Intelligence, 1980.

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

[7] Kieth S. Donnellan. Reference and definite description. Philosophical Review, 75:281–304, 1966.

[8] Barbara 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 University Press, Cambridge, England, 1980.

[9] Barbara J. Grosz, A. Joshi, and S. Weinstein. Providing a unified account of definite noun phrases in discourse. In Proceedings of the Twenty-first Annual Meeting, pages 44–50, Association for Computational Linguistics, 1983.

[10] Barbara J. Grosz and Candace L. Sidner. Discourse structure and the proper treatment of interruptions. In Proceedings of the Ninth International Joint Conference on Artificial Intelligence, pages 832–839, 1985.
[11] David Kaplan. Dthat. In Peter Cole, editor, *Syntax and Semantics, Volume 9*, Academic Press, New York, New York, 1978.

[12] Saul Kripke. Speaker reference and semantic reference. In French et al., editor, *Contemporary Perspectives in the Philosophy of Language*, University of Minnesota Press, Minneapolis, Minnesota, 1977.

[13] Amichai Kronfeld. *Reference and Denotation: The Descriptive Model*. Technical Note 368, SRI International Artificial Intelligence Center, 1985.

[14] Amichai Kronfeld. *The Referential Attributive Distinction and the Conceptual-Descriptive Theory of Reference*. PhD thesis, University of California, Berkeley, 1981.

[15] John Searle. Referential and attributive. In *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge, England, 1979.

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

[17] Peter F. Strawson. On referring. In J.F Rosenberg and C. Travis, editors, *Reading in the Philosophy of Language*, Prentice Hall, Englewood, New Jersey, 1971.