1. Overview

Linguistic theories typically assign various linguistic phenomena to one of the categories, syntactic, semantic, or pragmatic, as if the phenomena in each category were relatively independent of those in the others. However, various phenomena in discourse do not seem to yield comfortably to any account that is strictly a syntactic or semantic or pragmatic one. This paper focuses on particular phenomena of this sort—the use of various referring expressions such as definite noun phrases and pronouns—and examines their interaction with mechanisms used to maintain discourse coherence.

Even a casual survey of the literature on definite descriptions and referring expressions reveals not only defects in the individual accounts provided by theorists (from several different disciplines), but also deep confusions about the roles that syntactic, semantic, and pragmatic factors play in accounting for these phenomena. The research we have undertaken is an attempt to sort out some of these confusions and to create the basis for a theoretical framework that can account for a variety of discourse phenomena in which all three factors of language use interact. The major premise on which our research depends is that the concepts necessary for an adequate understanding of the phenomena in question are not exclusively either syntactic or semantic or pragmatic.

The next section of this paper defines two levels of discourse coherence and describes their roles in accounting for the use of singular definite noun phrases. To illustrate the integration of factors in explaining the uses of referring expressions, their use on one of these levels, i.e., the local one, is discussed in Sections 3 and 4. This account requires introducing the notion of the centers of a sentence in a discourse, a notion that cannot be defined in terms of factors that are exclusively syntactic or semantic or pragmatic. In Section 5, the interactions of the two levels with these factors and their effects on the uses of referring expressions in discourse are discussed.

2. The Effects of Different Levels of Discourse Coherence

A discourse comprises utterances that combine into subconstituents of the discourse, namely, units of discourse that are typically larger than a single sentence, but smaller than the complete discourse. However, the constituent structure is not determined solely by the linear sequence of utterances. It is common for two contiguous utterances to be members of different subconstituents of the discourse (as with breaks between phrases in the syntactic analysis of a sentence); likewise, it is common for two utterances that are not contiguous to be members of the same subconstituent.

An individual subconstituent of a discourse exhibits both internal coherence and coherence with the other subconstituents. That is, discourses have been shown to have two levels of coherence. *Global coherence* refers to the ways in which the larger segments of discourse relate to one another. It depends on such things as the function of a discourse, its subject matter, and rhetorical schema [Grosz, 1977, 1981; Reichman, 1981]. *Local coherence* refers to the ways in which individual sentences bind together to form larger discourse segments. It depends on such things as the syntactic structure of an utterance, ellipsis, and the use of pronominal referring expressions [Sidner, 1981].

The two levels of discourse coherence correspond to two levels of focusing—global focusing and centering. Participants are said to be globally focused on a set of entities relevant to the overall discourse. These entities may either have been explicitly introduced into the discourse or be sufficiently closely related to such entities to be considered implicitly in focus [Grosz, 1981]. In contrast, centering refers to a more local focusing process, one relates to identifying the single entity that an individual utterance most centrally concerns [Sidner, 1979; Joshi and Weinstein, 1981].

---

1 This research was supported in part by the National Science Foundation under Grant MCS-8115105 to SRI International, and Grant MCS81-07290 to the University of Pennsylvania.
The two levels of focusing/coherence have different effects on the processing of pronominal and nonpronominal definite noun phrases. Global coherence and focusing are major factors in the generation and interpretation of nonpronominal definite referring expressions. Local coherence and centering have greater effect on the processing of pronominal expressions. In Section 5 we shall describe the rules governing the use of these kinds of expressions and shall explain why additional processing by the hearer (needed for drawing additional inferences) is involved when pronominal expressions are used to refer to globally focused entities or nonpronominal expressions are used to refer to centered entities.

Many approaches to language interpretation have ignored these differences, depending instead on powerful inference mechanisms to identify the referents of referring expressions. Although such approaches may suffice, especially for well-formed texts, they are insufficient in general. In particular, such approaches will not work for generation. Here the relationships among focusing, coherence, and referring expressions are essential and must be explicitly provided for. Theories—and systems based on them—will generate unacceptable uses of referring expressions if they do not take these relationships into account.

3. Centering and Anaphora

In our theory, the centers of a sentence in a discourse serve to integrate that sentence into the discourse. Each sentence, S, has a single backward-looking center, Cb(S), and a set of forward-looking centers, C(S). Cb(S) serves to link S to the preceding discourse, while C(S) provides a set of entities to which the succeeding discourse may be linked. To avoid confusion, the phrase *the center* will be used to refer only to Cb(S).

To clarify the notion of center, we will consider a number of discourses illustrating the various factors that are combined in its definition (abstractly) and in its identification in a discourse. In Section 5 we define center more precisely, show how it relates to Sidner's [1981] immediate focus and potential focal, and discuss how the linkages established by the centers of a sentence help to determine the degree of intelligibility of a discourse. We begin by showing that the center cannot be defined in syntactic terms alone. The interaction of semantics and centering is more complex and is discussed in Section 4.

The following examples, drawn from Reinhart [1982], illustrate the point that the notion of center is not syntactically definable, i.e., the syntax of a sentence S does not determine which of its NPs realizes Cb(S). (The reasons for the use of this terminology are discussed in Section 4.)

(1a) Who did Max see yesterday?
(1b) Max saw Rosa.
(2a) Did anyone see Rosa yesterday?
(2b) Max saw Rosa.

Although (1b) and (2b) are identical, Cb(1b) is Max and Cb(2b) is Rosa. This can be seen in part by noticing that *He saw Rosa* seems more natural than (1b) and *Max saw her* than (2b) (a fact consistent with the centering rule introduced in Section 5). The subject NP is the center in one context, the object NP in the other.

Even when the NP used to realize Cb(S) can be syntactically determined, the Cb(S) itself is not yet fully determined, for Cb(S) is typically not a linguistic entity (i.e., it is not a particular linguistic expression). Rosa, not *Rosa* is the Cb(2b). Consider the discourse:

(3a) How is Rosa?
(3b) Did anyone see her yesterday?
(3c) Max saw her.

Here, Cb(3c) is Rosa, but clearly would not be in other contexts where the expression *her* still realized the backward-looking center of *Max saw her.* This is seen most simply by considering the discourse that would result if *How is Joan?* replaced (3a). In the discourse that resulted, Joan, not Rosa, would be the center of (3c).

4. Centering and Realization

The interactions of semantic and pragmatic factors with centering and their effects on referring expressions are more complex than the preceding discussion suggests. In the examples given above, the NPs that realize Cb(S) also denote it, but this is not always the case: we used the term *realize* in the above discussion advisedly. In this section, we consider two kinds of examples in which the choice of which choice of and interaction among different kinds of interpretations of definite noun phrases are affected by the local discourse context (i.e., centering). Second, the role of pragmatic factors in some problematic cases of referential uses of definite descriptions [Donnellan 1966] is discussed.

4.1. Realization and Value-Free and Value-Loaded Interpretations

The distinction between realization and semantic denotation is necessary to treat the interaction between value-free and value-loaded interpretations [Barwise and Perry, 1982] of definite descriptions, as they occur in extended discourse. Consider, for example, the following sequence:

---

3. They differ in other respects also. Reichman [1981] and Grosz [1981] discuss some of these.

4. Initial attempts to incorporate focusing mechanisms in generation systems are described in [Appelt, 1981 and McKeown, 1982].
If an intermediate sentence allows both interpretations but prefers one in a given context, then either is possible in the third sentence. A use with preference for a value-loaded interpretation followed by a use indicating the value-free interpretation is illustrated in the sequence:

John thinks that the telephone is a toy.  
He plays with it every day.  
He doesn't realize that it is an invention that changed the world.

The preference for a value-free interpretation that is followed by a value-loaded one is easiest to see in a dialogue situation:

s1: The vice president of the United States is also president of the Senate.  
s2: I thought he played some important role in the House.  
s3: He did, but that was before he was VP.

4.2. Realization and Referential Use

From these examples, it might appear that the concepts of value-free and value-loaded interpretation are identical to Donnellan's [1966] attributive and referential uses of noun phrases. However, there is an important difference between these two distinctions. The importance to our theory is that the referential use of definite noun phrases introduces the need to take pragmatic factors (in particular speaker intention) into account, not just semantic factors.

Donnellan [1966] describes the referential and attributive uses of definite descriptions in the following way:

"A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion states something about whoever or whatever is the so-and-so. A speaker wishes to assert something about whatever or whoever fits that description; but in the referential use the definite description is merely one tool for doing a certain job—calling attention to a person or thing—and in general any other device for doing the same job, another description or a name, would do as well. In the attributive use, the attribute of being the so-and-so is all important, while it is not in the referential use."

The distinction Donnellan suggests can be formulated in terms of the different propositions a sentence $S$ containing a definite description $D$ may be used to express on different occasions of use. When $D$ is used referentially, it contributes its denotation to the proposition expressed by.
S; when it is used attributively, it contributes to the proposition expressed by S a semantic interpretation related to the descriptive content of D. The identity of this semantic interpretation is not something about which Donnellan is explicit. Distinct formal treatments of the semantics of definite descriptions in natural language would construe the appropriate interpretation differently. In semantic treatments based on possible worlds, the appropriate interpretation would be a (partial) function from possible worlds to objects; in the situation semantics expounded by Barwise and Perry, the appropriate interpretation is a (partial) function from resource situations to objects.

As just described, the referential-attributive distinction appears to be exactly the distinction that Barwise and Perry formulate in terms of the value-loaded and value-free interpretations of definite noun phrases. But this gloss omits an essential aspect of the referential-attributive distinction as elaborated by Donnellan. In Donnellan's view, a speaker may use a description referentially to refer to an object distinct from the semantic denotation of the description, and, moreover, to refer to an object even when the description has no semantic denotation.

In one sense, this phenomenon arises within the framework of Barwise and Perry's treatment of descriptions. If we understand the semantic denotation of a description to be the unique object that satisfies the content of the description, if there is one, then Barwise and Perry would allow that there are referential uses of a description D that contribute objects other than the semantic denotation of D to the propositions expressed by uses of sentences in which D occurs. But this is only because Barwise and Perry allow that a description may be evaluated at a resource situation other than the complete situation in order to arrive at its denotation on a given occasion of use. Still, the denotation of the description relative to a given resource situation is the unique object in the situation that satisfies the description relative to that situation.

The referential uses of descriptions that Donnellan gives examples of do not seem to arise by evaluation of descriptions at alternative resource situations, but rather through the *referential intentions* of the speaker in his use of the description. This aspect of referential use is a pragmatic rather than a semantic phenomenon and is best analyzed in terms of the distinction between semantic reference and speaker's reference elaborated in Kripke [1977].

Consider the following discourses drawn from Kripke [1977]:

(6a) Her husband is kind to her.
(6b) No, he isn't. The man you're referring to isn't her husband.
(7a) Her husband is kind to her.
(7b) He is kind to her but he isn't her husband.

With (6a) and (7a), Kripke has in mind a case like the one discussed in Donnellan [1966], in which a speaker uses a description to refer to something other than the semantic referent of that description, i.e., the unique thing that satisfies the description (if there is one). Kripke analyzes this case as an instance of the general phenomenon of a clash of intentions in language use. In the case at hand, the speaker has a general intention to use the description to refer to its semantic referent; his specific intention, distinct from his general semantic intention, is to use it to refer to a particular individual. He incorrectly believes that these two intentions coincide and this gives rise to a use of the referring expression *her husband* in which the speaker's reference and the semantic reference are distinct.6 (The speaker's referent is presumably the woman's lover).

From our point of view, the importance of the case resides in its showing that Cf(S) may include more than one entity, that is realized by a single NP in S. In this case, *her husband* contributes both the husband and the lover to Cf(6a) and Cf(7a). This can be seen by observing that both discourses seem equally appropriate and that the backward-looking centers of (6b) and (7b) are the husband and the lover, respectively, realized by their anaphoric elements. Hence, the forward-looking centers of a sentence may be related not semantically but pragmatically to the NPs that realize them.

Hence, the importance of the referential/attributive distinction from our point of view is that it leads to cases in which the centers of a sentence may be pragmatically rather than semantically related to the noun phrases that realize them.

5. Center Movement and Center Realization--Constraints

In the foregoing sections we have discussed a number of examples to illustrate two essential points. First, the noun phrase that realizes the backward-looking center of an utterance in a discourse cannot be determined from the syntax of the utterance alone. Second, the relation N realizes c between noun phrases N and centers c is neither solely a semantic nor solely a pragmatic relation. This discussion has proceeded at a rather intuitive level, without explicit elaboration of the framework we regard as appropriate for dealing with centering and its role in explaining discourse phenomena. Before going on to describe constraints on the realization relation that

6There are, of course, several alternative explanations: e.g., the speaker may believe that the description is more likely than an accurate one to be interpreted correctly by the hearer. Ferreting out exactly what the case is in a given situation requires accounts of mutual belief and the like. A discussion of these issues is beyond the scope of this paper.
explain certain phenomena in discourse, we should be somewhat more explicit about the notions of center and realization.

We have said that each utterance S in a discourse has associated with it a backward-looking center, \( C_S(S) \), and a set of forward-looking centers, \( C(S) \). What manner of objects are these centers? They are the sort of objects that can serve as the semantic interpretations of singular noun phrases.\(^7\) That is, either they are objects in the world (e.g., planets, people, numbers) or they are functions from possible worlds (situations, etc.) to objects in the world that can be used to interpret definite descriptions. That is, whatever serves to interpret a definite noun phrase can be a center.

For the sake of concreteness in many of the examples in the preceding discussion, we have relied on the situation semantics of Barwise and Perry. The theory we are developing does not depend on this particular semantical treatment of definite noun phrases, but it does require several of the distinctions that treatment provides. In particular, our theory requires a semantical treatment that accommodates the distinction between interpretations of definite noun phrases that contribute their content to the propositions expressed by sentences in which they occur and interpretations that contribute only their denotation—in other words, the distinction between value-free and value-loaded interpretations. As noted, a distinction of this sort can be effected within the framework of possible-worlds* approaches to the semantics of natural language. In addition, we see the need for interpretations of definite noun phrases to be dependent on their discourse context. Once again, this is a feature of interpretations that is accommodated in the relational approach to semantics advocated by Barwise and Perry, but it might be accommodated within other approaches as well.\(^8\)

Given that \( C_S(S) \), the center of sentence S in a discourse, is the interpretation of a definite noun phrase, how does it become related to S? In a typical example, S will contain a full definite noun phrase or pronoun that realizes the center. The realization relation is neither semantic nor pragmatic. For example, \( N \) realizes \( c \) may hold in cases where \( N \) is a definite description and \( c \) is its denotation, its value-free interpretation, or an object related to it by a "speaker's reference.* More importantly, when \( N \) is a pronoun, the principles that govern which \( c \) are such that \( N \) realizes \( c \) derive from neither semantics nor pragmatics exclusively. They are principles that must be elicited from the study of discourse itself. A tentative formulation of some such principles is given below.

Though it is typical that, when \( c \) is a center of \( S \), \( S \) contains an \( N \) such that \( N \) realizes \( c \), it is by no means necessary. In particular, for sentences containing noun phrases that express functional relations (e.g., *the door,* *the owner*) whose arguments are not exhibited explicitly (e.g., a house is the current center, but so far neither its door nor its owner has been mentioned),\(^9\) it is sometimes the case that such an argument can be the backward-looking center of the sentence. We are currently studying such cases and expect to integrate that study into our theory of discourse phenomena.

The basic rule that constrains the realization of the backward-looking center of an utterance is a constraint on the speaker, namely:

If the \( C_S \) of the current utterance is the same as the \( C_S \) of the previous utterance, a pronoun should be used.

There are two things to note about this rule. First, it does not preclude using pronouns for other entities as long as one is used for the center. Second, it is not a hard rule, but rather a principle, like a Gricean maxim, that can be violated. However, such violations lead at best to conditions in which the hearer is forced to draw additional inferences.

As a simple example, consider the following sequence, assuming at the outset that John is the center of the discourse:

\( (8a) \) He called up Mike yesterday. (he=John)  
\( (8b) \) He was annoyed by John's call.

\( (8b) \) is unacceptable, unless it is possible to consider the introduction of a second person named *John.* However, intervening sentences that provide for a shift in center from John to Mike (e.g., *He was studying for his driver's test*) suffice to make (8b) completely acceptable.

Sidner's discourse focus corresponds roughly to \( C_S(S) \), while her potential focus correspond approximately to \( C(S) \). However, she also introduces an actor focus to handle multiple pronouns in a single utterance. The basic centering rule not only allows us to handle the same examples more simply, but also appears to avoid one of the complications in Sidner's account. Example D4 from Sidner [1981] illustrates this problem:

\( (9-1) \) I haven't seen Jeff for several days.  
\( (9-2) \) Carl thinks he's studying for his exams.  
\( (9-3) \) But I think he went to the Cape with Linda.

On Sidner's account, Carl is the actor focus after (9-2) and Jeff is the discourse focus (\( C(S) \)). Because the actor focus is preferred as the referent of pronominal expressions, Carl is the leading candidate for the entity referred to by he in (9-3). It is difficult to rule this case out without invoking fairly special rules. On our account, Jeff is \( C_S(9-2) \) and there is no problem. The addition of actor focus was made to handle multiple pronouns—for example, if (9-3) were replaced by

He thinks he studies too much.

The center rule allows such uses, without introducing a

\(^7\) In a fuller treatment of our theory we will consider centers that are realized by constituents in other syntactic categories.

\(^8\) Israel [1983] discusses some of these issues and compares several properties of situation semantics with Montague semantics.

\(^9\) Grosz [1977] refers to this as "implicit focusing": other examples are presented in Joshi and Weinstein [1981].
second kind of focus (or center), by permitting entities other than $Cb(S)$ to be pronominalized as long as $Cb(S)$ is\textsuperscript{10}.

Two aspects of centering affect the kinds of inferences a hearer must draw in interpreting a definite description. First, the shifting of center from one entity to another requires recognition of this change. Most often such changes are affected by the use of full definite noun phrases, but in some instances a pronoun may be used. For example, Grosz [1977] presents several examples of pronouns being used to refer to objects mentioned many utterances back. Second, the hearer must process (interpret) the particular linguistic expression that realizes the center.

Most previous attempts to account for the interaction of different kinds of referring expressions with centering and focusing (or "topic") have conflated these two. For example, Joshi and Weinstein [1981] present a preliminary report on their research regarding the connection between the computational complexity of the inferences required to process a discourse and the coherence of that discourse as assessed by measures that invoke the centering phenomenon. However, several of the examples combine changes of expression and shifts in centering.

Violations of the basic centering rule require the hearer to draw two different kinds of inferences. The kind required depends on whether a full definite noun phrase is used to express the center or whether a pronoun is used for a noncentered entity. We will consider each case separately.

Several different functions may be served by the use of a full definite noun phrase to realize the currently centered entity. For instance, the full noun phrase may include some new and unshared information about the entity. In such cases, additional inferences arise from the need to determine that the center has not shifted and that the properties expressed hold for the centered entity. For example, in the following sequences

(10) I took my dog to the vet the other day. 
\textit{The mangy old beast}....

(11) I’m reading \textit{The French Lieutenant’s Woman}. The book, which is \textit{Fowles best}...

the full definite noun phrases that are in boldface do more than merely refer.

When the current center is not pronominalized (it may not be present in the sentence), the use of a pronoun to express an entity other than the current center, is strongly constrained. The particular cases that have been identified involve instances in which attention is being shifted back to a previously centered entity (e.g., Grosz, 1977; Reichman, 1975) or to one element of a set that is currently centered. In such cases, additional inferences required to determine that the pronoun does not refer to the current center, as well as to identify the context back to which attention is shifting. These shifts, though indicated by linguistic expressions typically used for centering (pronouns), correspond to a shift in global focus.

6. Summary

The main purpose of the paper was to sort out the confusion about the roles of syntactic, semantic, and pragmatic factors in the interpretation and generation of definite noun phrases in discourse. Specific mechanisms that account for the interactions among these factors were presented. Discourses were shown to be coherent at two different levels, i.e., with referring expressions used to identify entities that are centered locally and those focused upon more globally. The differences between references at the global and local levels were discussed, and the interaction of the syntactic role of a given noun phrase and its semantic interpretation with centering was described.

References


\textsuperscript{10}Obviously, if $Cb(S)$ is not expressed in the next sentence then this issue does not arise.


