

The Representation and Processing of Coreference in Discourse

PETER C. GORDON AND RANDALL HENDRICK

University of North Carolina at Chapel Hill

A model is presented that addresses both the distribution and comprehension of different forms of referring expressions in language. This model is expressed in a formalism (Kamp & Reyle, 1993) that uses interpretive rules to map syntactic representations onto representations of discourse. Basic interpretive rules are developed for names, pronouns, definite descriptions, and quantified descriptions. These rules are triggered by syntactic input and interact dynamically with representations of discourse to establish reference and coreference. This interaction determines the ease with which coreference can be established for different linguistic forms given the existing discourse context. The performance of the model approximates that observed in studies of intuitive judgments of grammaticality and studies using online measures of language comprehension. The model uses the same basic interpretive mechanisms for coreference within and between sentences, thereby linking the domain traditionally studied by generative linguists to domains that have been of concern primarily to psychologists and computational linguists.

The phenomenon of coreference—where two linguistic expressions refer to the same thing—is a central topic in attempts to understand the meaning and structure of language. This is true both for disciplines that attempt to model knowledge of language and for disciplines that attempt to model language processing. Despite their often divergent goals, the disciplines of generative linguistics, psycholinguistics, computational linguistics and formal semantics have all expended considerable energy on aspects of coreference, particularly on the distinction between coreference with full referring expressions, such as names or descriptions, and with reduced referring expressions, such as pronouns. We believe that progress in these different disciplines has provided important ideas that converge in the

Direct all correspondence to: Peter C. Gordon, Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270; E-Mail: pcg@email.unc.edu; Randall Hendrick, Department of Linguistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3155; E-Mail: hendrick@email.unc.edu.

domain of local coreference, that is coreference within sentences and short strings of sentences that are semantically coherent. In this paper we attempt to integrate that progress in such a way as to provide a unified account that addresses some of the central concerns of each of these disciplines.

This account is based on the following principles: (1) The primary function of pronouns (and other reduced expressions) is to refer to things that have already been mentioned in a discourse and which are mentally represented in a discourse model; therefore pronouns are the natural vehicle for coreference. In contrast, the primary function of names (and of other full expressions) is to introduce entities into the discourse model; therefore coreference with repeated names requires additional mental processes. (2) The syntactic and sequential structure of language strongly influence mental representation in the discourse model which in turn influences the interpretation of referring expressions in linguistic input. (3) The discourse model is constructed incrementally with each utterance directly adding conditions that further specify the meaning embodied in the model; exceptions to this direct incremental construction occur in well-marked cases where linguistic forms indicate that a phrase or clause serves to modify the meanings that follow it rather than the meanings that precede it.

The paper is organized as follows: First, we briefly review how reference and coreference are approached by different disciplines within cognitive science, with particular attention on how the work contributes to the present effort. Next, we sketch out some basics of the approach taken towards coreference in Kamp and Reyle's (1993) Discourse Representation Theory; our work employs the mechanisms developed in that approach to formal semantics. The main body of our work then shows how the three principles outlined above can explain a wide range of basic phenomena in coreference. The final section of the paper places the work in a broader context by considering how the plausibility of events can shape the interpretation of referential expressions and by evaluating the symbol processing framework adopted in this paper in relation to other formalisms used to model cognitive processes.

Approaches to Reference and Coreference

Generative Linguistics. In the early days of generative grammar, the occurrence of a pronoun, as shown for example in (1), would have been created by a transformation applying to a deep structure containing two instances of the proper name, as shown in (2).

- (1) *Bill Clinton* contends *he* will win.
- (2) Bill Clinton contends Bill Clinton will win.

While considerable effort was expended on the characterization of this pronominalization transformation (Langacker, 1969; Lees & Klima, 1963; Ross, 1967), a series of steps fundamentally changed the focus of work on coreference (Bach, 1970; Jackendoff, 1972; Lasnik, 1976; Newmeyer, 1986 for a review). Pronominalization transformations were abandoned in favor of free insertion of names and pronouns, with indices to represent intended reference. Thus (1) might have the form in (3) or (4) depending on whether the

pronoun referred to Bill Clinton or to some other individual, so that coreference in (1) is not the result of any grammatical principle, but is “accidental” in the sense that linguistic theory has nothing to say about this process beyond the slogan “index freely” (Lasnik, 1976; Chomsky, 1981; and others). An important assumption of this approach is that proper names and pronouns accomplish their referential work in the same way, in that their reference is encoded by the same type of indices.

- (3) Bill Clinton_i contends he_i will win.
 (4) Bill Clinton_i contends he_j will win.

From this perspective, linguistic theory is limited to explaining what prevents coreference, as for example between *Bill Clinton* and *him* in (5). A fundamental goal of this approach has been to search for privileged structural relations that force disjoint reference between a pronoun and a possible antecedent. Principle B of the Binding Theory (Chomsky, 1981; 1986) is a prominent formulation of such a restriction. It requires that pronominals (e.g., *he*, *she*, *it*) not appear in the same local domain with a c-commanding antecedent.¹ This locality condition accounts for the contrast between (3) where coreference is permitted and (5) where it is blocked; in (3) the pronoun is not in the same local domain as its antecedent, whereas in (5) it is. Principle B also predicts that a pronoun and its antecedent can appear in the same local domain so long as the antecedent fails to c-command the pronoun. For this reason, there is a contrast between (5) and (7) where in both examples the coreferential elements are in the same local domain but differ in whether the pronoun has a c-commanding antecedent.²

- (5) *Bill Clinton_i respects him_i.
 (6) Bill Clinton_i respects him_j.
 (7) Many supporters of Bill Clinton_i respect him_i

A similar approach is adopted for coreference involving names and other unreduced referring expressions such as descriptions. Principle C of the Binding Theory attempts to explain why coreference is prevented in sentences like (8). It says that names are prohibited from having c-commanding antecedents. The contrast between sentences like (8) and (9), taken from Reinhart (1983), is often cited to illustrate this principle. Coreference is excluded in (8) because *she* c-commands the name *Zelda*; (9) admits coreference as possible (although not necessary) because *her* does not c-command *Zelda*.

- (8) *[_{IP} She_i [_{VP} adores [_{NP} Zelda's_i teachers]]]
 (9) [_{IP} [_{NP} Those that [_{VP} know her_i]] [_{VP} adore Zelda_i]]

Thus, both Principles B and C provide information on disjoint reference, not on coreference. Both principles are also built upon Reinhart's (1976, 1983) notion of c-command, an exclusively hierarchical relation between the coreferential expressions. Before Reinhart's work, it was generally believed that in addition to hierarchical requirements, there

was a left-to-right restriction that required an antecedent to precede a coreferential pronoun in some circumstances. Reinhart's isolation of c-command broke with this traditional belief by making the restrictions on pronouns subject only to hierarchical arrangement, a position that has won widespread adherence.

Gordon and Hendrick (1997) conducted a systematic investigation of what competent speakers of English judge to be acceptable patterns of coreference involving names and pronouns. This is the task that linguists usually perform informally on themselves and their colleagues, but Gordon and Hendrick (1997; 1998a; 1998b) brought more formal methodology to the task by controlling the presentation of stimulus materials, testing subjects who were unaware of the hypotheses, and performing statistical analyses on the data. The results showed that naive subjects have consistent intuitions of grammaticality that agree with some of the principles of contemporary binding theory but not others. In particular, subjects have strong intuitions that reflexives and pronouns are in complementary distribution within a clause; they accept coreference in sentences such as (10) but reject it in sentences such as (5). This pattern supports contemporary Binding Theory.

(10) Bill Clinton_i respects himself_i.

Beyond that, subjects' judgments of coreference are systematically influenced by syntactic structure in ways that are not consistent with syntactic theory.

In contrast to expectations based on the Binding Theory, Gordon and Hendrick (1997) found the following pattern of judgments. There is a strong left-to-right effect on the relation between form of referring expression and the acceptability of coreference. Coreference is highly acceptable in sentences where a name precedes a pronoun [a Name-Pronoun sequence such as (11) or (12)].

(11) Lisa_i visited her_i brother at college.

(12) Lisa's_i brother visited her_i at college.

Coreference is considerably less acceptable in sentences containing repeated names [Name-Name sequences such as (13) and (14)].

(13) Lisa_i visited Lisa_i brother at college.

(14) Lisa's_i brother visited Lisa_i at college.

Coreference is least acceptable in sentences where a pronoun precedes a name [Pronoun-Name sequences such as (15) and (16)].

(15) She_i visited Lisa_i brother at college.

(16) Her_i brother visited Lisa_i at college.

Further, a c-command relationship between the two referring expressions reduces the acceptability of coreference in Name-Name sequences; it enhances acceptability in Name-

Pronoun sequences and it generally has little effect on Pronoun-Name sequences. The effect of c-command on disjoint reference is greater when the antecedent is within the subject of the main clause, while the effect of c-command on coreference is greater when the antecedent is not within the subject. Gordon and Hendrick (1997) formulated a general construct of syntactic prominence to account for the interaction of c-command and subject status. Finally, coreference in a pronoun-name sequence is widely accepted when the pronoun is in a preposed adjunct phrase as shown in (17).

(17) If he_i wins the lottery John_i will be happy.

The finding that syntactic prominence of the antecedent induced disjoint reference in name-name sequences and positive coreference in name-pronoun sequences is analogous to reading-time results to be discussed below, which were obtained by Gordon (Gordon et al., 1993; Gordon & Chan, 1995; Gordon & Scarce, 1995) for intersentential coreference. Thus, the work of Gordon and Hendrick (1997) combines with Gordon's earlier work on discourse comprehension to provide an empirical basis for a theoretical synthesis of the comprehension of intrasentential and intersentential anaphora.

Psycholinguistics. While generative linguists have come to focus on factors promoting disjoint reference between two expressions, psycholinguists have maintained an interest in factors leading to the referential interpretation of an expression. These factors are of two sorts: semantic and structural.

It is generally agreed that a pronoun, or other referential expression, will ultimately be interpreted in the manner that makes the greatest communicative sense in the particular context in which it is used. A number of studies have provided empirical evidence that knowledge of events influences the interpretation of pronouns (Gernsbacher, 1989; Gordon & Scarce, 1995; and others), but these studies have not attempted to provide a principled analysis of how event knowledge constrains pronoun interpretation. Doing so would require a comprehensive theory of general knowledge. Some work has examined how inference mechanisms in specific domains of knowledge could help to resolve pronominal reference. One such instance involves the implicit causality of verbs (Grober, Beardsley, & Caramazza, 1978) where it has been shown that for some verbs the grammatical subject is seen as causing the action described by the verb, while for other verbs causality is attributed to the object. When two utterances (clauses or sentences) are joined by a causal connective (e.g., "because"), then a pronoun at the beginning of the second utterance will be interpreted as coreferential with whichever argument of the first verb is perceived as causal (Ehrlich, 1980). For example, in (18) the pronoun is judged coreferential with Bill because Bill occupies the role of stimulus for the verb "admire".

(18) John admires Bill because he is reliable.

Recent studies have attempted to determine precisely the time course of implicit causality effects, though the results of this line of inquiry are unclear. McDonald and MacWhin-

ney (1995) have claimed that implicit causality has an immediate influence on pronoun interpretation while Garnham, Traxler, Oakhill, and Gernsbacher (1996) have argued that its influence emerges later when the description of an event is fully integrated. However, the conflicting results that have been obtained to date are reminiscent of findings in the domain of parsing where methodological complexities have made describing the architecture of parsing very difficult (e.g., the contrasting results of Ferreira & Clifton, 1986 and Trueswell, Tanenhaus, & Garnsey, 1994). The present work takes the view that structural factors initially drive referential interpretation and that the plausibility of events influences interpretation of referential expressions only after the description of an event is integrated (Garnham et al., 1996). We will return to consideration of this type of semantic effect in the Final Discussion.

When semantic forces are neutral, structural factors have clear effects on the interpretation of pronouns, though there is some disagreement over the details of their operation. At least four competing proposals can be identified. The first is parallel function, where a pronoun is interpreted as coreferential with an antecedent that has the same grammatical role (Grober, Beardsley, & Caramazza, 1978; Sheldon, 1974). The second is subject-assignment, where a pronoun is preferentially interpreted as coreferential with the subject of the preceding clause (Crawley, Stevenson, & Kleinman, 1990; Fredericksen, 1981). The third is grammatical matching, where a pronoun preferentially corefers with the antecedent that shares the most grammatical features (Smyth, 1994). The fourth is structural prominence, where a pronoun is interpreted as coreferential with an antecedent that is maximally prominent due to a combination of syntactic embeddedness and linear position (Gordon, et al, 1993; Gordon & Scearce, 1995). In many cases, these different proposals make identical predictions. For example, pronouns tend to occur most frequently in subject position and tend to corefer with earlier subjects (Fredericksen, 1981), a fact that is consistent with all four accounts. The predictions of these proposals diverge primarily in cases involving the interpretations of postverbal pronouns as shown in (19) where the pronoun *him* could be coreferential with either John or Derek.

(19)³ John encouraged Derek and Carol asked him how it was going.

Unfortunately, it is difficult to control the effect of semantic factors on the interpretation of postverbal pronouns because of the rich semantic information conveyed by the verb and the preceding discourse. Further, studies of pronoun interpretation in such cases have tended to use offline measures of interpretation so that it is difficult to determine whether the results are due to immediate processing or later integrative processing (Crawley et al., 1990; Smyth, 1994).

As an alternative to studying how potentially ambiguous pronouns are interpreted, some researchers have investigated how different forms of referring expressions contribute to the construction of a discourse model by comparing comprehension of pronominal coreference and repeated-name coreference in cases where the coreferential interpretation is unambiguous (Hudson-d'Zmura, 1988; Gordon et al., 1993; Gordon & Chan, 1995; Gordon & Scearce, 1995). Such studies have demonstrated a *repeated-name penalty* (Gordon et al.,

1993), where the time to read a sentence containing a repeated name is slower than the time to read that sentence with a pronoun, and have shown that the repeated-name penalty is greater when the antecedent expression that introduces the referent is structurally prominent (Hudson-d'Zmura, 1988; Gordon et al., 1993; Gordon & Chan, 1995). This effect would be observed in a comparison of reading times for sentences such as (20c) and (20c'). Corroborative findings have come from cross-modal semantic priming tasks where pronominal coreference facilitated processing of semantic associates more readily than did repeated-noun coreference (Cloitre & Bever, 1988). Such findings have supported the view that pronouns are directly interpreted as referring to previously introduced discourse entities, while coreference with repeated names requires more elaborate interpretive processes (Cloitre & Bever, 1988; Gordon et al., 1993; Garrod, Freudenthal, & Boyle, 1994). This view is in stark contrast to one emerging from probe-word tasks, a method in which subjects are asked to judge as quickly as possible whether a word has occurred in the preceding text (sentence or discourse). Studies using that technique have shown that repeated-name coreference speeds responses to probes consisting of the repeated name and slows responses to probes consisting of a name from the text that was not repeated; pronominal coreference does not immediately have such effects (Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989; MacDonald & MacWhinney, 1990). In fact, one series of studies showed that pronouns have no effect on responses to probe words leading to the conclusion that pronouns are not interpreted at all by the automatic processes responsible for basic understanding in reading (Greene, McKoon, & Ratcliff, 1992). Both Gordon et al. (1993) and Garrod et al. (1994) have argued that results from reading-time studies and semantic-priming studies are more informative about coreferential processing than are probe-word studies, because the probe-word task could be tapping processes that are not actually involved in language comprehension. Further, the idea that pronouns are interpreted more directly than repeated names is consistent with the notion from the Binding Theory (Chomsky, 1981) that there are more restrictions on coreference with names than on coreference with pronouns, and with ideas that have been developed in computational linguistics about how pronouns and names contribute to discourse coherence.

Computational Linguistics. Unlike the artificial languages that have been designed for machines, natural languages exhibit significant amounts of nondeterminism. That is to say, situations arise in the processing of natural language where the next step to be undertaken is not uniquely determined. Ambiguities of various sorts best exemplify this problem. For example, in parsing it often must be determined whether a given word (e.g., "permit") is being used as a verb or a noun (Allen, 1987; Steedman, 1996); a similar issue is encountered in determining which sense of a homonymous word (e.g., bank) is being used. It is from this interest in ambiguity that researchers in computational linguistics have long considered the interpretation of reference and coreference as critical to the processing of natural language (Winograd, 1972). The principal focus has been on how the context of an utterance contributes to the "resolution" (or determination) of a pronoun's reference. A central question that has been confronted concerns the relation between inference in coreference and in coherence more generally. Charniak (1972) and Hobbs (1979) offered impor-

tant attempts to resolve pronominal reference by nonexplicit inferences. According to Hobbs (1979), coreference is in effect a byproduct of general inference mechanisms that are used to make a text coherent. Shank (1973) and others developed theories of special domains of knowledge that could assist this inferencing mechanism (cf. Partridge, 1991). Others elaborated how the achievement of specific tasks (such as questions) assists in making appropriate inferences (Appelt, 1985). These attempts share to varying degrees the belief that pronoun resolution can be accomplished by unstructured semantic inferences.

A contrasting view is that language processing must take advantage of the contextual structure of language, particularly with regard to reference, in order to constrain processes of inference and make them computationally tractable (Grosz, 1977). This idea has been developed in centering theory (Grosz et al. 1983; 1986; 1995), which provides an analysis of the structure of discourse segments that has had substantial impact on the development of natural language processing systems (Walker, Joshi, & Prince, 1997). According to this approach, the local coherence of discourse arises from the manner and extent to which successive utterances make reference to the same entities. In particular, each utterance in a locally-coherent discourse realizes a single entity, the backward-looking center (**Cb**), that provides a crucial link to the preceding utterance. This entity must be realized by a pronoun for the link to be established. This notion is illustrated in (20a-c), which marks the **Cb**s in a sample discourse as well as showing partial markings of the forward-looking centers (a construct to be discussed shortly).

- (20) a. Susan gave Betsy a pet hamster.
Cf = {Susan, Betsy, hamster₁}
- b. She reminded her that such hamsters were quite shy.
Cb = {Susan} **Cf** = {Susan, Betsy, hamsters}
- c. She asked Betsy whether she liked the gift.
Cb = {Susan} **Cf** = {Susan, Betsy, hamster₁}

The importance of realizing the **Cb** as a pronoun can be seen by contrasting 10c with 62c' which uses a name rather than a pronoun.

- (20) c' Susan asked Betsy whether she liked the gift.
Cb = {Susan} **Cf** = {Susan, Betsy, hamster₁}

The originators of centering theory offered the judgment that this sort of substitution makes the sentence incoherent in the context of the discourse.

In addition to having a backward-looking center, centering theory proposes that each utterance in a discourse has a set of forward-looking centers (**Cf**) that are ordered with respect to their prominence in the discourse (Grosz et al. 1983; 1986, 1995). Ranking in the **Cf** has two important consequences: it affects the likelihood that an entity will be the **Cb** of the subsequent utterance and it affects the interpretation of pronouns.

The focus in Centering Theory on the different constraints that affect the coreferential use of names and pronouns provides a potential connection to syntactic work on the Binding Theory (Chomsky, 1981; 1986) which also relates the form of a referring expression to coreference. This connection is strengthened by the results of psycholinguistic research (Gordon et al., 1993; Gordon & Chan, 1995; Hudson-D'Zmura, 1988) on Centering which has identified the critical role that syntactic characteristics of an antecedent phrase play in the subsequent interpretation of referential expressions. This general focus on the role of syntactic factors in coreference is supported by findings obtained by Lappin and Leass (1994) on the success of algorithms for pronoun resolution. Their approach derives from syntactic structures a measure of discourse salience that in turn is used in resolving the reference of a pronominal expression. A heterogeneous class of intrasentential syntactic factors contributes to this measure, such as sentence recency, being a subject, being a direct object, and not being contained within another noun phrase. Performance of this algorithm in resolving pronouns in natural language corpora is high, and is not improved by the use of supplemental semantic modeling based on the approach of Asher and Wada (1988). The approach taken by Lappin and Leass (1994) differs from Centering Theory in that it limits its scope to pronoun resolution and does not attempt to model coherence. However, the demonstration that their algorithm works very successfully using syntactic aspects of antecedent expressions is consistent with the findings of psycholinguistic research on Centering (Gordon et al., 1993; Gordon & Chan, 1995).

Formal Semantics. Formal semantics borrows from the study of mathematical logic a characterization of the meaning of a statement in terms of the conditions under which it would be true or false (Gamut, 1991). In this way, it provides a characterization of meaning as a relation between language and the world, and thus meaning is more than a mere translation of a linguistic statement into some other, uninterpreted language. The semantic value of a statement in these terms is composed from the semantic value of the syntactic parts of the statement. The semantic value of the parts of a sentence are conceived of as entities in the universe or sets of such entities. For example, a sentence like *Bill is snoring* would be classified as true or as false. This decision in turn depends on what entity the name Bill points to and whether that entity is in the set of entities that are snoring. If the entity designated by Bill is a member of the set of snorers then the sentence Bill is snoring is true; otherwise it is false. This basic intuition of how to approach meaning was shown to be viable over an important fragment of English by Montague (1974). While there is a school of semanticists (including Frege and Montague) that rejects any psychological import to this approach to meaning, other semanticists see this general approach as very useful for considering how meaning could be represented mentally (Johnson-Laird, 1983; Partee, 1995).

The truth-conditional view of meaning has been challenged on the grounds that some phenomena cannot be characterized in truth-conditional terms. Speech acts (e.g., promises) and other pragmatic aspects of language are important examples and have attracted considerable attention (see Levinson (1983) for an overview). Nevertheless formal semanticists contend that a full theory of meaning must at the least be able to explicate the use of language to express true or false statements about the world, and that an appropriate theory of

pragmatics would complement the theory of meaning developed from a truth-conditional perspective rather than displace it. Yet even when attention is restricted to statements that can be given a straightforward definition in truth-conditional terms, the semantics advocated by Montague has difficulty characterizing the meaning of a statement when it is placed in the context of preceding and following statements. Significantly Montague's semantics has difficulty capturing some basic features of our understanding of temporal sequence and reference as can be seen by considering (21)-(22).

- (21) a. Jane is walking in the park and whistles.
 b. Jane is walking in the park. She whistles.
- (22) a. Jane owns a Porsche. It is red.
 b. Jane doesn't own a Porsche. It is red.

From the perspective offered by Montague (see Gamut, 1991, for a discussion) (21a) and (21b) should always mean the same thing. Yet there is an important difference. (21a) seems to assert that Jane simultaneously walks and whistles, while (21b) appears to allow an interpretation in which Jane's whistling follows her walking in time. In (22a) the pronoun "it" naturally refers to the specific automobile that Jane owns. Yet Montague has no ready explanation for why that same pronoun sounds exceedingly strange (even nonsensical) when used to refer to "a Porsche" in (22b).

Kamp and Reyle (1993) attempt to broaden model-theoretic semantics so as to capture more closely the semantic phenomena of natural language that emerge from the succession of statements in a discourse. Their model simultaneously aspires to be a formal semantic system and to model cognition. In their model, linguistic expressions refer to discourse objects within a discourse representation structure that mediates between syntactic expressions and objects in the world. This discourse structure is incrementally elaborated as each new statement of a discourse is encountered. The representation of a preceding proposition in the evolving discourse structure can have significant consequences for the interpretation of a subsequent proposition. Thus, meaning emerges as something more than the composition of the semantic values of syntactic elements in a simple proposition as was conceived by Montague. Kamp and Reyle analyze in detail the behavior of pronouns with respect to negation and quantification, as well as the complexities of situating events described by propositions in a temporal relation.

Kamp and Reyle adopt an incremental approach to the construction of a discourse model so that their system correctly maps syntactic representations onto the discourse representations that embody the meaning of a series of statements. Their standard for success in this project is the one typically employed by logicians, a characterization of the true inferences that follow from a statement. They show that they achieve this standard by demonstrating that the interpretations given by their approach can be translated into a first-order predicate calculus that is logically consistent. Beyond the domain of formal semantics, Kamp and Reyle's use of incremental construction of the discourse model makes their approach to meaning attractive to those (e.g., psycholinguists and computational linguists)

who are attempting to characterize processes that can operate on linguistic input as it arrives in order to determine its meaning.

A Formalism for Modeling the Interpretation of Reference and Coreference

We build our explanation of coreference within the general formalism of discourse offered by Kamp and Reyle (1993). As discussed above, this formalism has several properties that we wish to exploit. It describes how a discourse model is incrementally constructed through the use of processes that interpret linguistic input in the context of the existing discourse; at the same time the formalism can be given an explicit truth-conditional interpretation. Though we use the Kamp and Reyle formalism (1993), we depart from their approach by allowing certain aspects of a linguistic utterance to have a larger impact on the construction of a mental model during comprehension than they do. In our work, the linguistic form of a noun phrase strongly influences comprehension: pronouns are interpreted preferentially as referring to entities that are already present in the discourse model while proper names and definite descriptions cause the introduction of new entities into the discourse model. In addition, the syntactic organization of an utterance into phrases influences comprehension by contributing to an ordering of entities in a discourse model that determines the accessibility of those entities as referents for subsequent expressions. We refer to this kind of mental model as a Discourse Prominence Representation.

It is the ease of accommodating syntactic influences on prominence that cause us to find Discourse Representation Theory more congenial to our purposes than other broadly-similar theories of semantics. For example, despite its interest in how the form of referring expressions influences their interpretation in a discourse, the file-change semantics proposed in Heim (1983) is less useful to us because it takes as its input an abstract syntactic level of Logical Form (cf. May, 1985), while prominence seems to be influenced by factors represented in the surface constituent structure which forms the input in Kamp and Reyle's model. Similarly, the SNePS model of Rappaport, Shapiro, and Wiebe (1997) is not suitable to our purposes because it does not assign an important role to syntactic representations in the process of semantic interpretation.

Kamp and Reyle (1993) provide explicit rules, called *Construction Rules* (CR), that map syntactic representations onto *Discourse Representation Structures* (DRS). A construction rule is made up of a triggering condition that specifies the structural condition that leads the rule to apply, and output instructions that modify the discourse representation and replace a specified portion of the linguistic representation. Any material added by a Construction Rule is termed a *condition set*; such sets consist of discourse entities (or discourse referents) and the relations between them. The condition set together with the residual linguistic representation left as the output of the construction rule comprise a Discourse Representation Structure.

This approach is illustrated in (23) by showing the construction rule used to handle a noun phrase consisting of a proper name. When it encounters such a constituent, it introduces three condition sets: (1) it posits an entity for the discourse to talk about, the dis-

- (25) **Creation of a DRS for a simple sentence.** Time t_1 : a syntactic structure is parsed of the string *John sees Mary*. Time t_2 : the syntactic structure is provided to the Discourse Representation Structure component and an interpretation for the proper name *John* constructed by application of the construction rule CR.PN. The discourse referent x is posited. *John* is predicated of that referent, and it is substituted into the syntactic representation. Time t_3 : the construction rule CR.PN is reapplied to the structure produced at t_2 ; a second discourse referent, y , is posited of which *Mary* is predicated, and y is inserted into the syntactic structure.

In the way illustrated above multiple applications of CR.PN will ultimately derive the DRS in (28).⁴

- (26)

| | |
|--------------|-----|
| x | y |
| John (x) | |
| Mary (y) | |
| x sees y | |

This representation encodes in the information that there are two individuals x and y , called *John* and *Mary* respectively, that stand in a seeing relation. An analogous set of construction rules for definite descriptions and quantified expressions is given in the Appendix to account for data described in Gordon and Hendrick (1998a) and further analysis of pre-suppositions related to descriptions is given in Hendrick (1998).

A DRS, such as the one illustrated in (26), is meant to capture the meaning conveyed by a series of statements. It does so by providing a model that can be evaluated for truth in the world, thereby providing a characterization of meaning that is independent of language. The model specifies discourse entities and the relations between them. Explicit representation of discourse entities provides a way of keeping track of which token of a type of entity is participating in a series of statements. Predicates indicate the semantic relations between entities in a manner that can be completely independent of the specifics of the linguistic forms in which they were introduced. For example, the same predicate could be employed in representing active sentences (e.g., “*John saw Mary.*”) and passive sentences (“*Mary was seen by John.*”). The predicate relations between entities that we use are very similar in form to the “propositions” used in Kintsch’s work on discourse processing (Kintsch, 1988; Kintsch & van Dijk, 1978). A major difference between Kintsch’s work and our own is that Kintsch’s model takes as input a text that has already been coded into his propositional notation. The model then attempts to account for phenomena of memory and inference in terms of operations on those propositional representations. In contrast, our model addresses the creation of the initial semantic representation of a series of statements. This creation is the result of the operation of explicitly formulated Construction Rules on specific linguistic forms in the context of the existing DRS.

Our use of a predicate notation is consistent with work in formal semantics (Kamp & Reyle, 1993) and in cognitive psychology (Kintsch, 1988). That use does not mean that we believe that the predicate calculus is an appropriate model of human inference; a substantial body of empirical research indicates that the inferences that people make diverge from those allowed by formal logics (e.g., Johnson-Laird, 1983). Rather, we believe that the representations that are generated by our model could serve as input to inference mechanisms, such as those described by Johnson-Laird (1983), that more adequately describe human cognition.

Basic Mechanisms of Coreference

The portion of DRT outlined so far provides a mechanism for handling reference by proper names. Before turning to the mechanisms for handling coreference, we review basic findings concerning how the form of two referring expressions affects their coreferential interpretation. We believe that the following general statement captures a great deal of the variation in ease of coreference in a sequence of referring expressions: Coreference is more readily achieved in Name-Pronoun sequences than in Name-Name sequences which in turn admit coreference more readily than do Pronoun-Name sequences. This ordering, shown in (27) captures the essence of a large set of findings on different types of linguistic relations that have been obtained with different methods.

(27) Name-Pronoun > Name-Name > Pronoun-Name

This ordering of ease of coreference in the types of sequences is observed in the grammaticality judgments of naive subjects on a range of types of noun phrases (Gordon & Hendrick, 1997). It occurs when the names or pronouns are in relative clauses, possessives and conjuncts. Further, it occurs when the sentence containing the sequence is preceded by a discourse context (a question) that contains one of the names. The pattern holds both for categorical judgments of grammaticality and for ratings of grammaticality. The advantage of coreference in Name-Pronoun sequences as compared to Name-Name sequences is also demonstrated by the repeated-name penalty observed in self-paced reading tasks where sentences with repeated-names are read more slowly than matched sentences with coreferential pronouns (Gordon et al., 1993; Gordon & Chan, 1995; Gordon & Scearce, 1995). The repeated-name penalty has been observed both for intersentential coreference (Gordon et al., 1993; Gordon & Chan, 1995; Gordon & Scearce, 1995) and for intrasentential coreference (Gordon, Hendrick, Ledoux, & Yang, 1997). Thus, the generalization that different sequences of types of referring expressions show a systematic pattern in the ease with which they admit coreferential interpretation applies to a very diverse set of findings. In the remainder of this section of the paper we show how an intuitively plausible mechanism for coreference can explain this broad generalization. In subsequent sections we will consider some refinements to this generalization as well as one important exception.

The mechanisms that we propose to handle coreference are built on the premise that the difference between pronouns and proper names is not merely a superficial difference of form but also a fundamental difference of the way in which they establish reference. This

fundamental difference can be seen in the *conventional wisdom* on pronouns and proper names that can be found in traditional grammars (e.g. Jespersen (1964), in some approaches to semantics (e.g. Prince (1981) and Heim (1983)) and also in some current approaches to the comprehension of anaphora (Garrod et al., 1994). However, it is not embodied in other approaches to reference and coreference (e.g., Gernsbacher, 1989; Greene et al., 1992). This wisdom can be paraphrased as (28).

(28) Conventional Wisdom on the Achievement of Reference

Proper names introduce entities into a discourse.

Pronouns refer to entities already mentioned in a discourse.

The construction rule previously outlined in (23) gives us a way of handling proper names that is consistent with the first premise of conventional wisdom. The second premise of conventional wisdom presents pronouns as the primary way of achieving coreference.

The construction rule for pronouns shown in (29) provides a way of achieving coreference within DRT that incorporates the second component of conventional wisdom. This rule is triggered by a pronoun and then seeks to identify that pronoun with a discourse referent that already exists in the discourse model by finding a suitable antecedent as determined by the “grammatical” features of the pronoun, specifically: number, gender, animacy, and reflexivity⁵. If no suitable antecedent is found, an instruction is executed to posit a new discourse referent. This fallback instruction would handle the deictic uses where pronouns point directly to something in the world. Deictic pronouns on this view are secondary to pronouns of coreference and to be used felicitously should be accompanied by linguistic or contextual cues (such as gestures) that make other potential antecedents unsuitable. We emphasize that on this view pronouns are foremost formal devices to evoke a referent already familiar in a discourse (cf. Prince, 1981). Our version of the construction rule for pronouns differs in this respect from the one offered by Kamp and Reyle (1993). Their construction rule always posits a new discourse referent for every pronoun, and only then undertakes to identify that new referent with a previously established discourse referent by a statement of equivalence of the type $x = y$. At its heart, their account takes the deictic use of pronouns as primary and makes their anaphoric use secondary.

(29) CR.PRO

Triggering Condition:

$[_{NP} \text{ [PRO } \alpha \text{]}]$

Instruction:

- choose a suitable antecedent v , such that v exists in the DRS, and substitute v for $[_{NP} \text{ [PRO } \alpha \text{]}]$ in the triggering condition.
- If no suitable antecedent v is present, introduce into the universe of the DRS a new discourse referent u .
- Substitute u for $[_{NP} \text{ [PRO } \alpha \text{]}]$ in the triggering condition

(30) shows an example of a sentence whose interpretation requires the use of the construction rule for pronouns.⁶

(30) Jane thinks she is sick.

The application of CR.PRO to the sentence in (30) will produce the DRS in (31) where there is one discourse referent.

(31)

| |
|---|
| x Jane (x) x thinks x is sick |
|---|

This example shows that the construction rules for proper names and for pronouns are together sufficient to provide coreferential interpretation of sentences that contain Name-Pronoun sequences. Our empirical data indicate that this is the type of sequence of referring expressions that most easily allows a coreferential interpretation.

Let us now consider coreference between two names. A sentence like (32) will result in the construction rule for proper names being triggered twice. The resulting Discourse Representation Structure (33) will contain two distinct entities, both named *Jane*, where one thinks the other is sick.

(32) Jane thinks Jane is sick.

(33)

| |
|---|
| x y Jane (x) Jane (y) x thinks y is sick |
|---|

The existence in this example of two distinct entities each of which has the same name predicated on it illustrates a fundamental ambiguity of names. More than one entity may be referred to with the same name, and our model reflects this by having the construction rule for proper names treat a name as introducing a distinct new entity into the discourse model⁷. To capture coreferential interpretation involving repeated names, the structure needs another condition set, namely that $y = x$. To achieve this result, we must postulate a construction rule for equivalence (CR.EQ), which we do in (34). The triggering condition in this case is that the Discourse Representation Structure contains two discourse entities, both of which have the same name predicated of them. This construction rule will modify the discourse representation in (33) to that in (35).

(34) CR.EQ

| |
|--|
| <p>Triggering Condition: $[\gamma \dots x \dots y \dots]$ such that $\alpha(x)$ and $\alpha(y)$</p> <p>Instruction:</p> <ul style="list-style-type: none"> • Introduce the new condition $x = y$ • Remove the condition $\alpha(y)$ |
|--|

(35)

| |
|--|
| <p>x y Jane (x) Jane (y) $x = y$ x thinks y is sick</p> |
|--|

Our empirical results showed that a sentence with a name-name sequence like (32) is less acceptable than the otherwise parallel sentence in (30) that has a name-pronoun sequence. This is because the DRS for Sentence (30), shown in (31) has fewer condition sets and is constructed by fewer construction rule applications than in (35). For this reason, constructing (31) requires less effort than does constructing (35). Further, there is a phase during the construction of (35) in which there are distinct, unequated representations of different entities with the same name creating the psychological basis for a sense of disjoint reference. There is no such phase in the creation of a discourse representation for (31) because CR.PRO preferentially does not posit a distinct discourse referent but makes use of one already present in the universe of the DRS.

The final case to consider is coreference in Pronoun-Name sequences. A sentence like (36) will first trigger the construction rule for pronouns (39). It will fail to find a suitable antecedent (because none has occurred) and will therefore treat the pronoun as deictic and on that basis posit a new discourse referent. Subsequently, the construction rule for proper names will be triggered and it will posit a new discourse referent on which the name will be predicated. This will result in the discourse representation structure shown in (37).

(36) She thinks Jane is sick.

(37)

| |
|---|
| <p>x y Jane (y) x thinks y is sick</p> |
|---|

At this point, coreference between the pronoun and name is effectively blocked given the rules that we have postulated so far. In particular, the construction rule for equivalence, which can equate the entities referred to by repetition of a name, can not equate the entities referred to by the pronoun and the name because their predication involves different identifying information. The triggering conditions for the construction rule for equivalence are

simply not met. This explains why coreference is most difficult to achieve in pronoun-name sequences.

Effects of Syntactic Prominence on Coreference

Above we stated the generalization that the ease of achieving coreference in different types of sequences of referring expressions had the following order: Name-Pronoun sequences > Name-Name sequences > Pronoun-Name sequences. However, a substantial amount of evidence indicates that within the broad categories of Name-Pronoun sequences and Name-Name sequences, the syntactic prominence of the antecedent affects the ease of establishing coreference. Specifically, a syntactically prominent antecedent facilitates coreference in Name-Pronoun sequences and it inhibits coreference in Name-Name sequences. Informally, the syntactic prominence of an NP is related to its height in a syntactic tree and therefore inversely related to its depth of embeddedness.⁸ Formally, our conception of syntactic prominence is related to the *c-command relation* of Reinhart (1981). It can be defined as follows:

- (38) α is more prominent than β if α *x*-commands β , and β *y*-commands α where $x < y$.
 (39) α *n*-commands β if there is some node γ that dominates both α and β , and there are *n* many branching nodes that are dominated by γ and that dominate α .

Empirical evidence that prominence of this sort influences coreference comes from studies of judgments of the acceptability of coreference, reading time studies, and statistical analyses of corpus data showing that antecedents tend to occupy certain syntactic roles with greater probability than other roles. Gordon and Hendrick (1997) showed that having an antecedent in a syntactically prominent position increased judgments of the acceptability of coreference in Name-Pronoun sequences and decreased judgments of the acceptability of coreference in Name-Name sequences. This was shown for syntactic prominence as demonstrated by a variety of comparisons: subject head noun vs. object head noun, head noun vs. possessive, head noun vs. component of a conjunct, and head noun directly dominated by the main clause vs. head noun dominated by a subordinate clause. Reading time data also provide evidence that syntactic prominence of the antecedent favors Name-Pronoun coreference and disfavors Pronoun-Name coreference. Studies of intersentential coreference have shown that the repeated-name penalty is greater when the antecedent is the subject of the preceding sentence than when it is the object (Gordon et al., 1993; Gordon & Chan, 1995; Kennison & Gordon, 1997). Studies of both intersentential and intrasentential coreference have shown that the repeated name penalty is greater when the antecedent is the head noun than when it is embedded in a conjunct or when it is a possessive (Gordon, Ledoux, & Hendrick, 1997). The importance of syntactic prominence that emerges from studies of judgment and reading time is corroborated by results obtained on natural-language corpora using syntax-based algorithms for pronoun resolution. The algorithms used by Lappin and Leass (1994) employ a number of syntactic criteria for evaluating potential antecedents. Those criteria favor reference to entities with syntactically prominent antecedents over reference to entities with nonprominent antecedents.

Accounting for the effects of syntactic prominence on coreference requires elaboration of our model. This is done by incorporating a concept from centering theory (Grosz et al., 1983; 1995) that discourse referents constitute an ordered set (the set of forward-looking centers) and that this ordering is heavily influenced by the syntactic prominence of the most recent expression that referred to the entity. We represent this idea formally by characterizing the list of discourse referents in a DRS as an ordered set $\langle \alpha_i, \alpha_j, \dots, \alpha_n \rangle$ with α_i more accessible than α_{i+x} . A revised CR.PRO (shown in 30) preferentially selects the most accessible suitable antecedent in the set of discourse referents by working through the set of grammatically suitable discourse referents in order of accessibility in its attempt to find a suitable antecedent. The ease of referring to an entity will therefore depend on its prominence since that determines the length of the search.⁹

(40) CR.PRO (revised)

Triggering Condition

[_{NP} [_{PRO} α]]

Instruction:

- Chose an antecedent v_i , after considering every v_j $i < j$ such that v_j and v_i exist in the ordered set of discourse referents in the DRS and are suitable antecedents, and substitute v_j for [_{NP} [_{PRO} α]] in the triggering condition.
- If no suitable antecedent v_i is present, introduce into the universe of the DRS a new discourse referent u .
- Substitute u for [_{NP} [_{PRO} α]] in the triggering condition.

A companion revision to the construction rule for equivalence CR.EQ is shown in (41). As presented earlier, this rule provides a mechanism for achieving coreference in name-name sequences. It is modified so it too is now sensitive to the ordering of the discourse referents. The sense of disjoint reference is magnified when the first name in a name-name sequence is prominent because the disjoint interpretation is sustained longer in the processing of the utterance.

(41) CR.EQ (revised)

Triggering Condition:

[$\gamma \dots \gamma \dots$]

such that $\alpha(x)$ and $\alpha(y)$

Instructions:

- Identify the discourse referent x in the ordered list of discourse referents $v_1 \dots v_n$ by checking v_i after v_j where $i < j$.
- Introduce the new condition $x = y$
- Remove the condition $\alpha(y)$

Thus, the manner in which prominence influences the two rules for establishing coreference (CR.Pro and CR.EQ) provides an explanation for the opposite pattern of effects that

prominence has on coreference in Name-Pronoun and Name-Name sequences. The Appendix further generalizes this explanation to coreference involving definite descriptions and quantified expressions.

Backwards Anaphora and Discourse Segmentation

So far we have provided an explanation of the general effect of form of a referring expression on coreference and have shown how this general effect is moderated by the syntactic prominence of the antecedent expression. However, there is one narrowly constrained set of circumstances where the patterns described above do not hold. This occurs in the case of *backwards anaphora* where a pronoun precedes the name with which it corefers, thus violating the generalization that coreferential interpretation is very difficult to achieve in Pronoun-Name sequences. A major finding of Gordon and Hendrick (1997) was that naive, competent speakers of English are far less accepting of backwards anaphora than are the linguists who have formulated the Binding Theory (Chomsky, 1981; 1986; Reinhart, 1981). This finding of Gordon and Hendrick (1997) has played a critical role in allowing us to present the relatively simple formulation of the construction rule for pronouns because it need not allow for backwards anaphora. The one class of sentences where Gordon and Hendrick (1997) found that naive subjects accepted backwards anaphora was when the pronoun was in a fronted adjunct phrase as illustrated in (42).

(42) If she wins, Jane will be happy.

Adjuncts serve to semantically modify the main clause of a sentence. Accordingly, when they are fronted they modify the clause that follows them. Gordon and Hendrick (1997) found that Pronoun-Name coreference was highly acceptable when the pronoun occurred in a fronted adjunct that was either a subordinate clause or a prepositional phrase.

Results obtained by Gordon et al. (1993) show that fronted adjuncts exhibit another very interesting pattern with respect to coreference. In particular, no repeated-name penalty is observed for referring expressions in fronted adjuncts, even when they corefer with the subject of the preceding sentence as shown in passage (43). Gordon et al. (1993) interpreted this finding in the context of centering theory (Grosz et al., 1983; 1995) as indicating that an entity in a fronted adjunct could not be the backward-looking center, that is the entity that provides the crucial semantic link to the preceding utterance. Instead, their results indicated that the grammatical subject played that role. However, subsequent results by Gordon and Chan (1995) indicated that the direct object shows a repeated-name penalty when it is the first expression that can provide a link to the preceding sentence, a finding that raises the question of why an expression in a fronted adjunct could not play that role as well.

(43) Lisa wanted to try painting portraits.
 She persuaded Joe to pose for a couple of hours.
 She painted him in an abstract, geometric style.
 In **her/Lisa's** opinion, the painting captured Joe's mood exactly.
 Other observers had more difficulty interpreting the work.

The results obtained by Gordon and Hendrick (1997) showing that backwards anaphora is acceptable with fronted adjuncts suggest the following answer to this question: pronouns are not immediately interpreted in a coreferential fashion when they are contained in fronted adjuncts. This answer makes a great deal of sense in the context of our adoption of a general approach to reference (Kamp & Reyle, 1993) where syntactic structures are used to govern the incremental construction of semantic structures. Because an adjunct provides a semantic modification of the clause to which it is attached, a fronted adjunct cannot directly contribute to the meaning of the discourse that precedes it but must first contribute to the meaning of the clause that follows it. In an important sense, a fronted adjunct interrupts the incremental building of a discourse model. This semantic intuition can be modeled by partitioning the discourse universe, a device that Kamp and Reyle (1993) use to account for several aspects of semantic interpretation including conditional expressions.

Kamp and Reyle (1993) argue that the semantic contribution of conditional expressions to a discourse model differs from that of simple declarative sentences because declarative sentences contribute directly to the current discourse model by building incrementally on the condition sets of a DRS so as to further constrain the truth conditions of that model, while conditional expressions cannot do this because the antecedent clause itself makes no assertion that has a truth value. Kamp and Reyle account for such expressions by positing that a subuniverse, distinct from the current discourse universe, is created for the representation of the antecedent clause. The subuniverse is connected by an implication operator to a subuniverse for the consequent. Together, these connected subuniverses constitute a condition set on the containing universe. The appeal to subuniverses has two functions that are given a single representation in Kamp and Reyle's treatment. It models the non-incremental elaboration of the DRS. At the same time, it represents the pattern of accessibility for pronouns in such structures. An instance of universal quantification in the antecedent will be prevented from binding a pronoun in the consequent, as illustrated in (44b), because when the universally quantified expression is hidden in a subuniverse, a pronoun cannot access it in Kamp and Reyle's system.

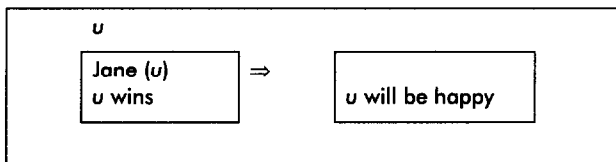
- (44) a. If Jane_i wins, she_i will be happy.
 b. * If every girl_i wins, she_i will be happy.

We suggest that it is profitable to keep these two functions conceptually distinct. To this end, let us suppose the construction rule (45) which interprets NPs that appear in conditionals. This rule establishes a new DRS universe. It also introduces a discourse referent u into that new universe in the following way: proper names have referents introduced into the containing DRS while other nominal expressions within the conditional clause have a discourse referent assigned within the subuniverse established by CR.if. The construction rule in (45) will construct the DRS in (46) for the sentence (44) on the assumption that CR.Pro is responsible for the interpretation of the pronoun.

(45) CR.if

| |
|--|
| <p>Triggering Condition: [γ ... [if IP] CP ...]</p> <p>Instructions:</p> <ul style="list-style-type: none"> • Begin a new DRS U_{k+1} • Introduce a new condition $K_n \Rightarrow K_{n+1}$ with $K_n = K_{n+1} = (\{\}, \{\})$ into U_{k+1}. • For any $[_{NP} [_{PN} \alpha]]$ within IP, introduce a new discourse referent u into K_n. • For any other $[_{NP} \alpha]$ within IP, introduce a new discourse referent u into the universe of the DRS U_{k+1}. • Introduce a new condition $\alpha(u)$ into U_{k+1} • Substitute u for $[_{NP} \alpha]$ in IP. |
|--|

(46)



From our perspective, the key to understanding backwards anaphora in fronted adjuncts derives from the semantic need to partition the adjunct phrase from the preceding discourse, a generalization of Kamp and Reyle's treatment of conditionals.¹⁰ A formal statement of this generalization is given in the construction rule for adjuncts shown in (47) which is only triggered by a configuration that includes a very clear linguistic cue (a preposition, complementizer, or other functional category) to partition the discourse universe. This generalization is warranted because neither fronted prepositional phrases nor clauses beginning with temporal or causal connectives build directly in an incremental fashion on the preceding discourse. With regard to the interpretation of pronouns, this partitioning of the discourse means that CR.Pro cannot search the existing universe for a suitable antecedent when it encounters a pronoun in the adjunct phrase. Instead, such adjuncts lead to positing discourse referents within the new discourse universe for each nominal expression encountered. When the construction rule for adjuncts encounters a pronoun in a fronted adjunct, it establishes a new discourse referent and predicates the pronoun of it. Unlike other pronouns that are directly equated with a previously existing discourse referent, the initial assignment of a discourse referent to a pronoun in a fronted adjunct must be provisional because the pronoun occurs at the beginning of a discourse segment where no discourse referent could possibly be available. The construction rule CR.EQ. Adjunct in (48) then interprets this provisional referent by equating it either with a discourse referent introduced to the containing universe in the processing of the subsequent matrix clause, or subsequently to a discourse referent in the preceding discourse universe if the two discourse universes are collapsed.

(47) CR.FRONTED.Adjunct

| |
|---|
| <p>Triggering Condition: [_{CP} [XP] CP...]</p> <p>Instructions:</p> <ul style="list-style-type: none"> • Begin a new DRS U_{k+1} • Introduce a new condition $K_n = \{\}$ into U_{k+1}. • For any [_{PN} α] within XP, introduce a new discourse referent u into the universe of the DRS U_{k+1}. • For any other [_{NP} α] within XP, introduce a new discourse referent u into K_n. • Introduce a new condition $\alpha(u)$ into U_{k+1} • Substitute u for [_{NP} α] in XP. |
|---|

(48) CR. EQ.Adjunct

| |
|---|
| <p>Triggering Condition γ: the condition set $\alpha(u)$ within K_i where α is a pronoun</p> <p>Instructions:</p> <ul style="list-style-type: none"> • Equate u with a discourse referent v that is within the universe K_{i-1} that contains K_i. |
|---|

The modified mechanisms from DRT that we propose to handle backwards anaphora in fronted adjuncts provide a straightforward explanation of the finding discussed above that no repeated-name penalty is observed in fronted adjuncts. The partitioning of the discourse universe that is occasioned by the fronted construction means that a pronoun in the fronted adjunct cannot be immediately interpreted as referring to an entity in the earlier discourse universe. Thus, interpretation of a pronoun becomes very much like interpretation of a name in the sense that there is no immediate basis for a coreferential interpretation. This means that no repeated-name penalty should occur because the pronoun provides no greater basis for linking the current utterance to the preceding utterance than does the repeated name. It is this linking function that is described by centering theory (Grosz et al., 1983; 1995) as a critical component of the coherence of a multi-utterance discourse segment. The occurrence of a fronted adjunct does not permanently block the creation of a coherent discourse segment because the main clause of the sentence can build semantically on the preceding discourse universe and this can cause the partition between the discourse universes to collapse yielding a unified discourse representation. Such a process of referential linking of the main clause of a sentence that has a fronted adjunct to the preceding discourse representation was demonstrated by Gordon et al. (1993) who found that the repeated-name penalty is observed for the subject of the main clause of a sentence with a fronted adjunct.

In summary, this section has shown that two very distinct phenomena, the judged acceptability of backwards anaphora in fronted adjuncts (Gordon & Hendrick, 1997) and the absence of a reading time advantage for pronouns over names in sentences presented in discourse context (Gordon et al., 1993) can be given a unified explanation in terms of the constraints on the incremental construction of a representation of the meaning of a discourse.

Syntax and Semantics in the Processing of Reference

In our model of the interpretation of referring expressions syntactic aspects of the linguistic input guide the construction of a universe that embodies the meaning of a discourse. Here, we address questions of how those syntactic representations are computed and how the contribution of syntactic structures to the construction of a discourse representation interacts with semantic constraints on the plausible interpretation of events.

The construction rules in our model are triggered by linguistic input that has been parsed so as to reveal the syntactic structure of noun phrases and the type of phrase marker to which those NPs are attached. While our model requires such a parse, it is neutral with respect to how the parse is performed. The required structure could be determined by parsing strategies that operate purely on syntactic information as has been advanced for the domains of attachment of clausal and prepositional phrases (Frazier, 1987). Or, the required structure could be determined using syntactic information combined with a constrained type of semantic information (Crain & Steedman, 1985). Finally, the required structure could be determined by a mechanism that simultaneously takes into account a wide variety of types of information, such as those given by the dominant usages of words (MacDonald, Seidenberg, & Pearlmuter, 1994). Consider how such different mechanisms might handle the contrast in subject NPs in (49) and (50). When applied to (49), a purely syntactic parser, like that developed in Marcus (1980), would begin an NP on encountering the determiner "The" and, looking ahead three words, would take the unambiguous verb status of "is" as an indication that the NP must be complete. The (initially) ambiguous word "square" in (49) would be classified as a (premodifying) adjective solely because this is necessary in order to make the first four words of the sentence correspond to a syntactically acceptable string in English. Alternatively, the parsing of these words could make use of additional information such as whether the dominant use of "square" is as an adjective or noun (MacDonald et al., 1994).

(49) The square box is in the center of the room.

(50) The square is in the center of the town.

As noted above, our model is neutral with respect to how the analysis of the noun phrase is achieved. What it requires is that there be a minimal analysis of the input into a noun phrase before an entity is introduced into the discourse model. For us, the crucial difference is that the syntax of (49) encodes the fact that the sentence is about a box, not a square. There is no entity in the discourse model of (49) that corresponds to a square; square is represented as a predicate of the entity box. In contrast, the syntax of (50) means that the sentence is about a square which corresponds to a distinct entity in the discourse model.

The nature of the relation between syntactic representations and semantic representations is also of concern with respect to the manner in which the plausibility of an event influences the interpretation of referential expressions. Consider the short passage below (51).¹¹

- (51) a. John sent a package to Bill.
 b. He received it two days later.

The pronoun “He” in the second sentence seems to refer to Bill. Significantly, this interpretation is not given by the construction rule for pronouns (CR.Pro shown in 30), which interprets the pronoun as referring to the most prominent entity in the discourse, in this case John. We are committed to the claim that, while this interpretation of the pronoun as referring to an entity that is not the most prominent is achieved, it can only be achieved with added cost. Gordon and Scearce (1995) found that reading times for the later part of a sentence like (51b) were significantly elevated when the meaning of its verb phrase combined with the meaning of earlier utterances so that the discourse made the greatest communicative sense when the pronoun referred to an entity that was not the most prominent one in the discourse. No such elevation in reading times was observed when a repeated name (in this case “Bill”) was used rather than a pronoun, suggesting that the elevated reading times did not result from the semantic complexity of the passage but rather from difficulty due to interpreting the pronoun. We think that this pattern of results is consistent with our formulation of the construction rule for pronouns because it suggests that determination of the ultimate interpretation of the pronoun emerges from aspects of the semantic representation (consisting of the predications in the discourse universe) after initial interpretation by the construction rule for pronouns.

To be more complete, our model does require a mechanism for reinterpreting a pronoun based on the semantic plausibility of the event being described. While the details of how this mechanism would work depend on the specifics of the event being described, we can illustrate how it might occur for (51). This passage describes what could be called a “transfer event”. If the event is analyzed in terms of thematic roles (Jackendoff, 1972 & Dowty, 1991), the verb “sent” in the first sentence assigns the thematic role of source (the origin of the event) to its subject (John) and the thematic role of theme (that which undergoes movement or change of state) to its direct object (package). Bill functions here as the goal that the theme moves toward. The verb “received” in the second sentence assigns the role of goal to its subject and also the role of theme to its direct object. Thus, the structurally preferred interpretation of “He” as referring to the subject of the first sentence (John) leads to a semantic representation where John is the source of the theme in the first sentence and the goal of that same theme in the second sentence. That configuration makes no sense with regard to the common understanding of transfer, so the pronoun is reinterpreted as referring to Bill.¹² In the case of transfer events, it appears that complete thematic assignment from both verbs is necessary in order to get a nonpreferred interpretation of the subject pronoun. To illustrate this, consider (52) which is the same as (51) through the verb of the second sentence.

- (52) John sent a package to Bill.
 He received prompt confirmation of its delivery.

Here, the pronoun “He” seems to refer to John, thereby showing that simply shifting from being the thematic source in the first sentence to being the thematic goal in the second

sentence is not a major obstacle to identity of reference. This shift is only a problem when the theme is constant across the two sentences. This supports the notion that substantial semantic information is required in order to force a nonpreferred interpretation of a pronoun in this case. Because so much semantic information is required to override the structurally favored interpretation, we argue that this effect is likely to occur subsequent to the immediate interpretations driven by the structural analysis of the input.

The idea that considerable semantic information is required to influence the interpretation of reference is also advanced by Garnham et al. (1996) based on a study of the implicit causality of verbs (Grober, Beardsley, & Caramazza, 1978). They considered sentences like (53) below, where people tend to regard the event described by the verb "punish" as being caused by the second NP (the direct object "Diane" in this case).

(53) Betty punished Diane three weeks ago because she didn't do the dishes.

When a causal connective (e.g., because) follows such a verb, readers tend to interpret an ambiguous pronoun in the subsequent clause as referring to the character that caused the action (Ehrlich, 1980). Again, this result would not be expected given our construction rule for pronouns and its dependence on syntactically-based prominence. Garnham et al. (1996) evaluated two hypotheses about how interpretations based on implicit causality are achieved. The first is a focus hypothesis which states that implicit verb causality leads to focus being placed on the causal entity (the second NP in 53) thereby making that entity the most accessible for subsequent reference. The second is an integration hypothesis that states that the implicit causality of verbs only has an effect through the integration of the meaning of the two clauses. Garnham et al. (1996) interpret the results of a series of probe-word experiments as supporting the integration hypothesis. From our perspective, this is consistent with the notion that pronoun interpretation based on the semantics of events comes relatively late in processing.

Of course, while evidence from the understanding of transfer events and implicit causality is suggestive, we would not argue that it provides a basis for concluding that semantic plausibility never influences the interpretation of pronouns at an early stage of processing. The time course of plausibility effects in referential processing remains an open empirical question. Results showing that plausibility had an effect on early stages of processing would constitute significant evidence against our model. A pattern of results that consistently showed that plausibility influences interpretation late in processing would provide support for our model.

Alternative Formalisms and Explanatory Goals

We have framed our explanation of coreference in terms of a symbol processing system. Of course, many researchers (Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989) believe that most, if not all, aspects of cognition can or should be modeled by distributed representations in connectionist networks, thereby doing without explicit symbols of this sort. Others have argued that connectionist approaches prove unilluminating (cf. Fodor & Pylyshyn (1988)) and still others (Prince & Smolensky, 1997) have argued for the

desirability of mixed approaches that draw from both kinds of architectures. Does this vigorous debate have implications for our model? We believe it is important to assess these competing modes of explanation empirically and in a non-dogmatic fashion. There are certainly aspects of our model that might usefully be expressed in a connectionist formalism so as to incorporate easily the continuous and stochastic effects naturally exhibited in that kind of model. In particular, the triggering conditions of our construction rules might be characterized as activation thresholds relative to baseline activations, so that characteristics of the input and of the current state of the system would determine the timecourse and probability of a rule being executed. Similarly, the notion of prominence that is so important to our discourse model could be implemented as differences in activation of discourse entities that are evaluated in parallel during the execution of the construction rules for pronouns and equivalence.

Nevertheless we chose not to frame our model in connectionist terms for several reasons. First, syntactic structures play a significant role in how reference and coreference are interpreted and connectionist approaches are not yet adept at capturing these kinds of relations. For example, connectionist models currently have difficulty in capturing the identifiability of the same NPs in multiple contexts (cf. Fodor & McLaughlin, 1990), but recognition of NPs is essential for the operation of our model. They also seem to have trouble capturing basic syntactic relations such as dominance and recursion (Elman, 1992). A second consideration that makes symbolic systems of more use in building our model concerns the notion of a variable. Our discourse model takes predicates to be functions that map from sets of individuals to truth values. For example, the predicate "love" can be thought of as a function with two arguments "LOVE(x, y)"; x and y are variables that can be filled by individuals to make a true or false proposition. Our construction rules posit entities that fill the variables in the predicate functions. Yet variables in this sense have proven difficult for connectionist architectures to model (Quinlan, 1991). A current attempt to address this problem is offered by Shastri and Ajjanagadde (1993) which makes use of temporal information to achieve that binding. However, the ways in which names, pronouns, definite descriptions, and indefinite descriptions are identified with the variables of a predicate in a discourse constitutes a nontrivial pattern that is addressed by our model. It is difficult to see how these different classes of variable binding can be explained in terms of a single mechanism of synchronic binding. The challenge our model poses for connectionist theorizing is to provide a principled and natural explanation for the different kinds of variable binding that seem to play an important role in language processing.

Our model addresses both people's knowledge of the acceptable distribution of coreferential expressions and the processes by which they interpret referential expressions. Thus, our model involves a close coordination between knowledge of a grammar and language processing, or between linguistic competence and performance in the terms of Chomsky's distinction (e.g., Chomsky, 1965). Linguists typically appeal to intuitive judgments of native speakers of a language as evidence for their theories of competence. Psycholinguists on the other hand typically look at some aspect of language use, such as the timecourse of reading, as a measure of language comprehension. In contrast, we employed both types of data (surveys of intuitive judgments and measures of the timecourse of language compre-

hension) as constraints on our model of language comprehension. While linguistics and psychology are separate academic modules that conduct their business independently of one another, we believe that competence and performance theories of cognitive phenomena are not independent. Bresnan and Kaplan (1982) endorse this general view in presenting the “strong competence hypothesis” which states that every competence operation or representation must be employed in some performance process. As Steedman (1996) observes, this view has never been seriously questioned. The strong competence hypothesis leads one to expect a close connection between a competence theory and a performance theory, and our model is offered in that spirit. We believe that arbitrary classifications of data into “competence data” and “psycholinguistic data” have no systematic theoretical significance.

Summary

Our model, Discourse Prominence Theory, describes the representation and processing of reference and coreference in natural language. At its core, the model consists of three principles. The first principle pertains to the form of referring expressions: The primary function of pronouns is to refer to entities that have already been mentioned in a discourse, thereby making pronouns a natural vehicle for coreference. In contrast, the primary function of names (and other unreduced referring expressions) is to introduce entities into a discourse model, thereby making repeated names a poor vehicle for coreference. The second principle pertains to how the syntactic and sequential structure of language is related to discourse models. The syntactic and sequential structure of a sentence is a major determinant of the prominence of a discourse entity. High prominence of a discourse entity affects the interpretation of subsequent referential expressions by facilitating coreference by pronouns and resisting coreference by repeated names. The third principle pertains to the incremental construction of meaning in a discourse model. For the most part, each utterance in a discourse adds conditions to the discourse model that serve to further specify the meaning of the model. An exception to this incremental construction occurs when a phrase or clause begins with an explicit linguistic marker indicating that it semantically modifies the utterance that follows it rather than the utterances that precede it. This causes a semantic partitioning of the discourse model that prevents immediate interpretation of a pronoun thereby eliminating the usual coreferential advantage of pronouns over repeated names and creating the possibility of backwards anaphora. In our model, these principles are formally expressed as construction rules that map syntactic structures onto representations of the semantics of discourse and which use the prominence of discourse entities as context for interpreting syntactic expressions.

Discourse Prominence Theory provides an integration of work on coreference from the disciplines of generative linguistics, psycholinguistics, computational linguistics, and formal semantics. It addresses key goals of those disciplines by providing an account of both the distribution and interpretation of expressions that create coreference within and between sentences through the use of an explicit set of rules that create a model with an explicit semantics.

Acknowledgements: The work in this paper was supported by NSF Grant IRI-94-04756.

APPENDIX

In this appendix, we extend the formalisms developed for proper names by providing construction rules for the interpretation of definite descriptions and for quantified expressions. These rules provide an explanation for data in Gordon and Hendrick (1998b) which show that judgments of coreference involving these two classes of expressions are very similar to those involving proper names. Accordingly, the construction rules for coreferential interpretation of these expressions are the same as those used for names. The rules differ only in two respects: Their triggering conditions differ because definite descriptions and quantified descriptions are different in form from proper names, and the semantic representation of quantified expressions differs because unlike proper names and definite descriptions, these expressions do not refer to unique entities.

Definite descriptions are handled in a fashion parallel to proper names. The construction rule responsible for their interpretation, CR.DD in (54), introduces a new discourse referent and predicates the description of that referent.

(54) CR.DD

| |
|--|
| <p>Triggering Condition: $[y \dots [NP [DET \textit{the}] N] \dots]$</p> <p>Instructions:</p> <ul style="list-style-type: none"> • Introduce a new discourse referent u into the universe of the DRS, U_k. • Introduce a new condition the $N(u)$ into the condition set of the DRS. • Substitute u for $[NP [DET \textit{the}] N]$. |
|--|

A sentence such as (55) will thus have the DRS in (56).

(55) John saw the girl.

(56)

| |
|---|
| <p style="text-align: center;"> x y John (x) the girl (y) x sees y </p> |
|---|

The construction rule responsible for universal quantifiers such as *each*, *every* and *all* is somewhat more complicated. It establishes domains within the universe of discourse that the quantified expression operates on. The rule in (57), CR.Q, constructs a DRS like (59) for a sentence such as (58), involving universal quantification. CR.Q establishes within the universe of the DRS, two subuniverses: K_1 and K_2 . This is the consequence of the first instruction in CR.Q positing $K_1 = K_2 = (\{\}, \{\})$. Such subuniverses play an important role in the justification of Discourse Representation Theory because they have an important

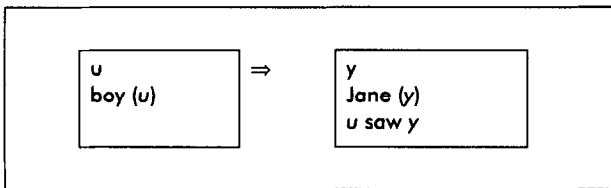
role in determining whether a quantified nominal is a suitable antecedent for a pronoun. When a quantified expression is within a subuniverse it is not available as a suitable antecedent for a pronoun in a containing universe. In the text we appealed to this aspect of Discourse Representation Theory to explain some special properties of coreference exhibited by clause initial adjuncts (e.g., fronted conditionals) and conjoined nominals. The symbol \Rightarrow can be understood as material implication in the predicate calculus.

(57) CR.Q

| |
|---|
| <p>Triggering Condition: $[\gamma \dots [NP [DET \text{ each }] N] \dots]$</p> <p>Instructions:</p> <ul style="list-style-type: none"> • Introduce a new condition $K_1 \Rightarrow K_2$ with $K_1 = K_2 = \{\}, \{\}$. • Introduce a new discourse referent u into the universe of the DRS U_{K_1}. • Introduce a new condition $N(u)$ into the condition set of K_1. • Introduce a new condition γ from γ where γ results by substituting u for $[NP [DET \text{ each }] N]$ into the condition set of K_2. • Delete γ from the list of condition statements in K. |
|---|

(58) Each boy saw Jane.

(59)

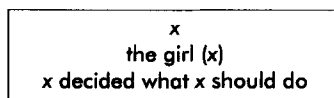


The construction rules that interpret names and definite descriptions, CR.PN and CR.DD, are comparable and relatively simple. In contrast, the construction rule responsible for universal quantification, CR.Q, is more complex in terms of the work it performs to construct its output. Not surprisingly the results of Gordon and Hendrick (1997) suggest that sentences involving coreference with universal quantified expressions are as a class somewhat less acceptable than sentences with names or definite descriptions.

This line of explanation generalizes to facts concerning definite descriptions and quantified expressions. Consider definite descriptions. Sentences such as (60) with definite description-pronoun sequences are provided with the DRS in (61) by application of CR.DD followed by CR.PRO.

(60) **The girl_i** decided what **she_i** could do.

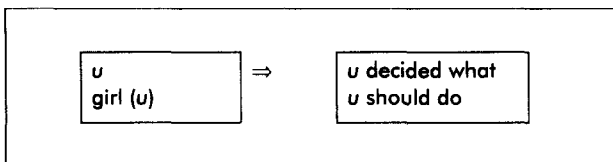
(61)



Similarly when a quantified NP-pronoun sequence is encountered, CR.Q will interpret the quantified expression followed by an application of CR.PRO triggered by the pronoun. In this way we construct a DRS like (63) for the sentence (62).

(62) Each girl_i decided what she_i could do.

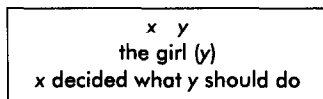
(63)



If the pronoun precedes a definite description as in (64), CR.PRO will be triggered, positioning a new discourse referent because at this moment the definite description has not yet had its discourse referent established. After the definite description is interpreted we are left with a DRS like (65) in which there are two distinct discourse referents with no relation of coreference.

(64) **She**_i decided what **the girl**_i should do.

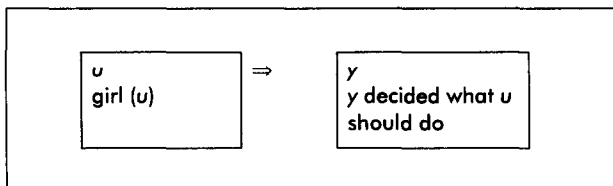
(65)



The same line of reasoning applies to structures in which a pronoun precedes a quantified expression, such as (66). The pronoun will be interpreted before a discourse referent is established by CR.Q, and as a result there will be no binding of the pronoun by the quantified expression, as can be seen in (67).

(66) **She**_i decided what **each girl**_i should do.

(67)



NOTES

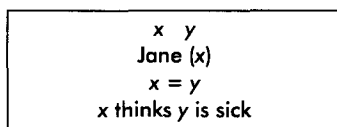
1. A constituent α is said to *c-command* another constituent β if the first branching node that dominates α dominates β as well. The local domain (or "governing category") of a pronoun is basically the smallest phrase containing the pronoun, its governor, and an accessible subject. See Harbert (1995) for a more complete discussion.

2. Principle A of Chomsky's Binding Theory puts anaphors (i.e. reflexives and reciprocals) in complementary distribution with pronominals (like he, she, it). Anaphors must have a *c*-commanding antecedent in their governing category. This requirement is satisfied in (i) but unmet in the ungrammatical (ii), where the relevant governing categories are bracketed.

- (i) [_a Bill Clinton_i respects himself_i]
 (ii) Bill Clinton_i thinks [_a she respects himself_i]

3. Adapted from Crawley, et al. (1990).
 4. We ignore here the translation of the predicate as tangential to our concerns. Moreover, as indicated above it is probable that a sub string of a sentence is parsed and assigned a DRS, although the illustration given in (15) in the text is somewhat idealized in this regard.
 5. The inclusion of reflexivity in the suitability conditions is a convenient assumption that we adopt for ease of exposition. In work in progress we attempt to explore more carefully how Principles A and B fare empirically and how they should be modeled.
 6. For Kamp and Reyle the DRS for this sentence would be (i) where two distinct discourse referents are posited initially and then latter identified.

(i)



7. The empirical investigations on which we base this claim have only examined first names. However, we believe that this ambiguity extends to more complete names. For example, the complete name "Peter Gordon" is ambiguous even in the relatively small world of psycholinguistics; experience with the addressing of mail and the compiling of author indices has shown this to be true. It is an empirically open question whether the assumption that names always introduce new entities extends to names that appear to have unique reference, for example the name of a *bona fide* celebrity such as Michael Jordan. According to Kripke's (1972) perspective on modal logic, it is essential for names to serve as "rigid designators" that refer consistently across all possible discourse universes. In our model, this function of rigid designation would be served by discourse entities rather than uninterpreted linguistic expressions. It is possible that linguistic expressions with unique reference achieve rigid designation more directly than common names because they are potentially less ambiguous. However, we think that it is more likely that comprehension is set up to treat names as introducing new entities (in part because of their usual ambiguity) and that the existence of truly unique reference is sufficiently uncommon that comprehension processes do not respond differently to names with unique reference.
8. A similar intuition is explored in Wasow (1979). We have assumed that the notion of a pronominal being bound in its governing category, Principle B, is part of a suitable antecedent in CR.Pro. Since a pronominal is only bound if it has a *c*-commanding antecedent, we will obviously still need to appeal to the notion of *c*-command. From the perspective of (38)-(39) *c*-command emerges as maximal prominence, 0-command.
9. We present this operation as an ordered search and suggest that shorter searches are easier. However, the operation could be cast in other terms. For example, one could easily envision this operation as a probabilistic one over activation levels determined by some range of preference constraints. Prominence and the matching of grammatical features that makes an antecedent suitable as a potential antecedent would be two such constraints. We see no immediate difference in the empirical consequences of these conceptualizations.
10. Kamp and Reyle address pronoun-name coreference in such sequences in a way that we find unsatisfactory. Essentially they argue that the acceptability of coreference in both name-pronoun sequences and pronoun-name sequences results from sometimes having the consequent processed before the antecedent is (completely) processed.
11. This example is modeled after constructions used by Corbett & Chang (1983), Gernsbacher (1989) and Gordon and Scarce (1995).
12. More specifically the information that John is the goal and Bill is the goal leads to the contradiction about who the package moves toward.

REFERENCES

- Allen, J. (1987). *Natural language understanding*. Menlo Park, CA: Benjamin-Cummings.
- Appelt, D. E. (1984). *Planning English sentences*. New York, NY: Cambridge University Press.
- Asher, N., & Wada, H. (1988). A computational account of syntactic, semantic, and discourse principles for anaphora resolution. *Journal of Semantics*, 6, 309-344.
- Bach, E. (1970). Pronominalization. *Linguistic Inquiry*, 1, 121-122.
- Bresnan, J., & Kaplan, R. (1982). Introduction: Grammars as mental representations of language. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. xvii-lii). Cambridge MA: MIT Press.
- Caramazza, A., Grober, E., & Garvey, C. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior*, 16, 601-609.
- Carroll, J. M., Bever, T. G., & Pollack, C. R. (1981). The non-uniqueness of linguistic intuitions. *Language*, 57, 368-383.
- Chang, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory & Cognition*, 8, 58-64.
- Charniak, E. (1972). Towards a model of children's story comprehension, Unpublished doctoral dissertation, MIT.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, the Netherlands: Foris Publishers.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- Cloitre, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes*, 3, 293-322.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, 11, 283-294.
- Crain, & Steedman (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 320-358). Cambridge, UK: Cambridge University Press.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19, 245-264.
- Dowty, D. (1991). Thematic proto roles and argument selection. *Language*, 67, 547-619
- Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, 32, 247-255.
- Elman, J. (1992). Grammatical structure and distributed representations. In Steven Davis (Ed.), *Connectionism. Theory and practice* (pp. 138-178). Oxford, UK: Oxford University Press.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Fodor, J. A., & McLaughlin, B. (1991). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. In T. Horgan & J. Tienison (Eds.), *Connectionism and the philosophy of mind* (pp. 331-354). Cambridge, MA: MIT Press.
- Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.
- Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, 4, 323-347.
- Gamut, L. T. F. (1991) *Logic, language and meaning. Vol. I: Introduction to logic; Vol II: Intensional logic and logical grammar*. Chicago: University of Chicago Press.
- Garnham, A. (1989). Integrating information in text comprehension: The interpretation of anaphoric noun phrases. In G. N. C. a. M. K. Tanenhaus (Ed.), *Linguistic Structure in Language Processing* (pp. 359-399). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, 35, 517-543.
- Garrod, S., Freudenthal, D., & Boyle, E. A. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, 33(1), 39-68.
- Gerken, L., & Bever, T. (1986). Linguistic intuitions are the result of interactions between perceptual processes and linguistic universals. *Cognitive Science*, 10, 457-476.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156.

- Gordon, P. C., & Chan, D. (1995). Pronouns, passives and discourse coherence. *Journal of Memory and Language*, 34, 216-231.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311-347.
- Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic coreference. *Cognition*, 62, 325-370.
- Gordon, P. C., & Hendrick, R. (1998a). Dimensions of grammatical coreference. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gordon, P. C., & Hendrick, R. (1998b). Intuitive knowledge of non definite NP anaphora. *Proceedings of the Chicago Linguistics Society*. Chicago, IL: University of Chicago Press.
- Gordon, P. C., Hendrick, R., Ledoux, K., & Yang, C. L. (1997). Processing of reference and the structure of language: An analysis of complex noun phrases. Manuscript submitted for publication.
- Gordon, P. C., & Scearce, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*, 23, 313-323.
- Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 266-283.
- Grober, E. H., & Beardstley, Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition*, 6, 117-133.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 175-204.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1986). Towards a computational theory of discourse interpretation. Unpublished Manuscript.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21, 203-226.
- Harbert, W. (1995). Binding Theory. In G. Webelhuth (Ed.), *Government Binding Theory and the Minimalist Program*. Blackwell.
- Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In B. Rainer, C. Schwartze, & A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language*. Berlin: de Gruyters
- Hendrick, R. (1998). Definiteness, specificity and Tongan determiners. Manuscript under review.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.
- Hudson-D'Zmura, S. B. (1988). *The structure of discourse and anaphore resolution: The discourse center and the roles of nouns and pronouns*. Unpublished doctoral dissertation, University of Rochester.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, Massachusetts: MIT Press.
- Jespersen, O. (1964). *The essentials of English grammar*. Birmingham AL: University of Alabama Press.
- Johnson-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P.N., & Garnham, A. (1980). Descriptions and discourse models. *Linguistics and Philosophy*, 3, 371-395
- Kamp, H., & Reyle, U. (1993). *From discourse to logic. Introduction to modeltheoretic semantics of natural language, formal logic, and discourse representation theory*. Dordrecht: Kluwer Academic Publishers.
- Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Keenan, E. (1974). The functional principle: Generalizing the notion of 'subject of'. *Proceedings of the Chicago Linguistics Society*, 10, 298-309.
- Kennison, S. M., & Gordon, P. C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes*, 24, 229-252.
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 85, 363-394.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language*, Dordrecht, Reidel.
- Langacker, R. (1969). "On pronominalization and the chain of command. In D. Reidbel and S. Schane (Eds.), *Modern Studies in English*. (pp. 160-186). Englewood Cliffs, NJ: Prentice Hall.

- Lappin, S. & Leass, H. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20, 535-561.
- Lasnik, H. (1976). Remarks on Coreference. *Linguistic Analysis*, 2, 1-22.
- Lees, R. B., & Klima, E.S. (1963). Rules for English Pronominalization. *Language*, 39, 17-28.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.
- MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language*, 29, 469-492.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676-703.
- Marcus, M. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.
- May, R. (1985). *Logical form*. Cambridge, MA: MIT Press.
- McCawley, J. (1968). Lexical insertion in a transformational grammar without deep structure. *Papers from the Fourth Regional Meeting of the Chicago Linguistics Society* (pp 71-80).
- McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34, 543-566.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466.
- McKoon, G., & Ratcliff, R. (1993). Discourse models, pronoun resolution and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1040-1052.
- Montague, R. (1974). *Formal philosophy. Selected papers of Richard Montague*. Edited with an introduction by R. H. Thomason. New Haven Yale University Press.
- Newmeyer, F. J. (1986). *Linguistic theory in America*. Orlando Academic Press.
- Partee, B. (1995). Lexical Semantics and Compositionality. In L. Gleitman & M. Liberman (Eds.), *Invitation to cognitive science I: Language* (pp. 311-360). Cambridge, MA: MIT Press.
- Partridge, D. (1991). *New guide to artificial intelligence*. Norwood, NJ: Ablex.
- Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, 275, 1604-1610.
- Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-256). Academic Press. New York.
- Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295-325). Philadelphia/Amsterdam: John Benjamins B. V.
- Quinlan, P. T. (1991). *Connectionism and psychology: A psychological perspective on new connectionist research*. Chicago, IL: University of Chicago Press.
- Rappaport, W. J., Shapiro, S. C., & Wiebe, J. M. (1997). Quasi-indexicals and knowledge reports. *Cognitive Science*, 21(1), 63-107.
- Reinhart, T. (1976). *The syntactic domain of anaphora*. Unpublished doctoral dissertation. MIT.
- Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
- Reinhart, T. (1981). Definite NP anaphora and c-command domains. *Linguistic Inquiry*, 12, 605-636.
- Ross, J. (1967). Constraints on variables in syntax. Unpublished doctoral dissertation. MIT.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart, J. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216-271). Cambridge, MA: MIT Press.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Shank, R. (1973). Identification of conceptualizations underlying natural language. In R. Schank & K.M. Colby (Eds.), *Computer models of thought and language* (pp. 187-247). San Francisco, CA: Freeman
- Shastri, L. & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences*, 16(3), 417-494.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272-281.
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23, 197-229.
- Steedman, M. (1996). Natural language processing. In M. Boden (Ed.), *Artificial Intelligence*, (pp 229-266). San Diego, CA: Academic Press.

- Steedman, M. (1995). Computational aspects of the theory of grammar. In L. Gleitman & M. Liberman (Eds.), *Invitation to cognitive science I: Language* (pp. 247-281). Cambridge, MA: MIT Press.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
- Walker, M. A., Joshi, A. K., & Prince, E. F. (1997). *Centering in discourse*. Oxford University Press.
- Wasow, T. (1979). *Anaphora in generative grammar*. Ghent: E. Story Scientia.
- Winograd, T. (1972). *Understanding Natural Language*. New York: Academic Press.