Title: Using Centering to Relax Gricean Informational Constraints on Discourse Anaphoric Noun Phrases[1]

Name: Rebecca J. Passonneau

Address: Bellcore

Morristown, NJ

beck@bellcore.com

Running Head: Centering and Informational Constraints

[a]This paper is a longer version of (Passonneau 1995). The author's full address is Bellcore, 445 South Street,Morristown, NJ 07960. The work reported here was partly supported under NSF grant IRI-91-13064; note that it was not supported by Bellcore. Many thanks to Robert Dale, Ehud Reiter, Megumi Kameyama and the anonymous reviewers for valuable comments on ideas presented here.

Abstract

In this paper, I explore the interaction among several factors argued to constrain discourse anaphoric expressions: the local centering context consisting of successive pairs of utterances; the informational content of a discourse anaphoric expression relative to the set of entities that need to be discriminated; the choice among different kinds of discriminating information; and the global attentional state that derives from the structure of an entire discourse. Specifications for the model are based on analysis of such factors in a corpus of spontaneous oral narratives. Then I use a reserved set of narratives for evaluating a simplified version of the generation model. The model integrates informational constraints, based on Dale's (Dale 1992) model, with centering. I show that by integrating centering constraints, the integrated model can account for two types of violation of informational constraints: informationally inadequate pronouns where the pronoun realizes the current backward-looking center, and certain full noun phrases in utterances where the backward-looking center is not overtly mentioned, and use of a pronoun would produce garden-path effects. I argue that for generation, centering constraints take precedence over informational constraints. The evaluation tests three conditions: informational constraints 1) with no focus structure imposed on the discourse, 2) with an empirically derived global focus structure, 3) and with global focus plus centering. Results improve dramatically under each next condition.

INTRODUCTION

More often than not, each new utterance in a discourse contains one or more discourse anaphoric expressions. Understanding the meaning of an utterance, and ultimately its point, generally depends on correctly interpreting these expressions. Whether a particular utterance contains a discourse anaphoric expression depends on how one defines utterance and discourse anaphoric expression. Later in the paper, I give technical definitions of both terms. For the moment, I use examples to illustrate factors that constrain the use of discourse anaphoric expressions. From these examples, I argue that while much is known about individual factors influencing a speaker's usage, less is known about how these factors interact with one another throughout a discourse. The goal of this paper is to provide specifications for a model to generate discourse anaphoric expressions, based on analysis of such factors in a corpus of spontaneous oral narratives. I use a reserved set of narratives for evaluating a generation model designed to meet these specifications.

(1) Default pronoun interpretation (centering)

a. A woman (e_j) went to the bookstore.
b. Afterwards, she (e_j) gave her aunt (e_k) a new book.
c. She (e_j)'s a true bibliophile.

(2) Semantic constraints/commonsense reasoning (abduction)

a. A woman (e_j) went to the bookstore.
b. Afterwards, she (e_j) saw her aunt (e_k).
c. She (e_k) looked pale.

(3) Lexical focus

a. A woman (e_j) went to the bookstore.
b. Afterwards, she (e_j) saw her aunt (e_k).
c. The woman (e_j) looked pale.

Figure 1: Three factors constraining discourse anaphoric reference

Fig. 1 presents three short, structurally parallel texts. In each case, the grammatical subject of the first utterance introduces a woman who is referred to again by means of a subject pronoun in the next utterance. The subscripted variable e_j represents a discourse entity that co-indexes the noun phrase that initially introduced the woman with subsequent *discourse anaphoric* pronouns and full noun phrases. I assume, following (Karttunen 1976) (Webber 1978) (Heim 1983) and (Kamp 1981), that a discourse entity is a term or variable in the discourse model that serves as a vehicle for relating information from distinct utterances to the same semantic representation. After the 'a' utterances in Fig. 1, each context contains a discourse entity (e_j) indexing a description of a woman who went into a particular bookstore. In the second utterance of each example, another discourse entity is introduced (e_k) and a situation involving both individuals is described. The context is thus updated with the new discourse entity e_k, and with new information about the existing discourse entity e_j. The 'c' examples illustrate different types of constraints on discourse anaphoric expressions.

In principle, the subject pronoun in 1c. of Fig. 1 is ambiguous between the woman who went into the bookstore and the aunt who received a book as a gift. However, there is a tendency to interpret such pronouns as indicated by the co-indexing—as coreferential with the preceding subject pronoun (e_j)—due in part to the structural parallelism of 1b. and 1c. As described later in this paper, centering theory (Joshi & Weinstein 1981) (Grosz, Joshi & Weinstein 1983) (Kameyama 1985) gives an account of the inferences for interpreting pronouns in such contexts. Example 2, also discussed later in the paper, illustrates a pronoun in a context that is structurally quite similar, but where the pronoun in 2c. is not assigned the default interpretation predicted by centering. The woman who is looked at, namely the aunt (e_k), is assumed to be the one who looks pale. Here, lexical and commonsense reasoning weight one interpretation more highly than others. Abductive reasoning (Hobbs, Stickel, Martin & Edwards 1988), as discussed later, provides a mechanism for deriving and weighting the inferences supporting this interpretation. Example 3, referred to here as *lexical focus*, illustrates that despite what is very much the same syntactic and semantic context as found in example 2, the interpretation of the 'c' utterance is "toggled" yet again. Either discourse entity could presumably be described as a woman. But because only one of them has been explicitly described as such, the subject of 3c. is interpreted to corefer with the subject of the preceding utterance (e_j).

The three contextual factors illustrated in Fig. 1 seem to have increasingly stronger effects in constraining the interpretation of discourse anaphoric expressions. Previous work has presented mechanisms integrating centering and commonsense reasoning for *understanding* discourse anaphoric noun phrases. But there has been no systematic analysis of when it would be necessary to *generate* texts like example 1 in Fig. 1. When do speakers provide less information information in their utterances than is necessary for unambiguous reference, and when do they provide just enough information?

As pointed out in (Dale & Reiter 1994), speakers sometimes provide more information than is necessary for unambiguous reference. The discourse excerpt shown in Fig. 2 depicts an unnecessarily informative discourse anaphoric noun phrase taken from a narrative monologue about a man picking pears. The utterance units

Seg	Phr.	
1	1.1	Okay,
	1.2	[pause] u-h [pause] the movie is basically about uh [pause] u-m [pause] a number of [pause] individuals,
	1.3	[pause] uh a guy (e_1) who's picking pears,
	1.4	[pause] u-m [pause] and a kid (e_2) on a bicycle.
2	2.1	Basically those are the two [pause] protagonists in this.
3	3.1	[pause] And um [pause] $\boxed{\text{the guy } (e_1) \text{ who is picking pears}}$,
	3.2	[pause] um [pause] um [pause] picks the pears and puts them in a [pause] in um [pause] these baskets that he (e_1) has.

Figure 2: Overly-informative discourse anaphoric reference

noted in the middle column are prosodic phrases, about which more will be said below when the corpus data is described. In phrases 1.3 and 1.4, the speaker describes two individuals, a guy picking pears (e_1) and a boy on a bicycle (e_2). In phrase 3.1, the speaker provides more information than is necessary for unambiguous reference. It has often been suggested that overly informative discourse anaphoric expressions occur at shifts in global discourse focus (Grosz & Sidner 1986) (Webber 1988) (Allen 1987) or narrative structure (Marslen-Wilson, Levy & Tyler 1982). The horizontal lines in Fig. 2 represent segment boundaries annotated in the corpus from which this excerpt is taken, where the boundaries derive from a study the author conducted with Diane Litman (Passonneau & Litman 1993). Note that the boxed noun phrase happens to occur at an empirically identified segment boundary, and that segments have been proposed to reflect global discourse focus (Grosz & Sidner 1986).

In this paper, I explore the interaction among the several factors argued to constrain discourse anaphoric expressions that are illustrated in Figs. 1 and 2: the local centering context consisting of successive pairs of utterances; the informational content of the anaphoric expression relative to the set of entities that need to be discriminated; the choice among different kinds of discriminating information, including lexical focus, or information inferable from successive utterances; and the global attentional state that derives from the structure of an entire discourse. The data indicates that by and large, discourse anaphoric NPs are informationally *adequate* and *economical* with respect to the global attentional state. They provide *adequate* information to unambiguously re-evoke discourse entities; the term adequacy is due to (Dale 1989). Further, they comprise the most *economical* linguistic choices. For noun phrases that violate adequacy or economy, I claim that centering theory provides an explanation. Violations of adequacy have often been discussed in the centering literature. Violations of economy have not been discussed elsewhere, but are predicted by the model I propose for integrating informational constraints with local and global attentional state.

In the section on *Background*, I review the basic assumptions about discourse structure that I adopt here.

I briefly discuss previous computational models for generating informationally adequate and economical noun phrases. In the *Corpus Analysis* section, I describe the data analyzed here and present results indicating that most of the discourse anaphoric noun phrases in the corpus are informationally adequate. In the following section, I present a relation called *ae-describe* that enforces informational adequacy and economy for all discourse anaphoric noun phrases within an utterance. In *The Need for Centering Constraints*, I present examples of violations of adequacy and economy, and discuss how centering constraints can be used to relax the informational constraints on those discourse anaphoric noun phrases that realize the current center. In the *Integrated Model*, I present the centering constraints as taking priority over the informational constraints. The results of a simplified implementation are presented in the *Evaluation* section. I evaluate three conditions: ae-describe with no global focus structure; ae-describe integrated with global focus; and ae-describe integrated with both local and global attentional state. Results improve dramatically with the addition of global focus, and again with the addition of centering. In the conclusion, I point out certain asymmetries between generation and understanding, and suggest directions for future work.

BACKGROUND

The Discourse Processing Model

(Grosz & Sidner 1986) propose a tri-partite model of discourse consisting of intentional structure, linguistic structure and attentional state. They (Grosz & Sidner 1986) make strong claims regarding the relations among these three levels, some of which have been questioned due to insufficient or conflicting evidence (e.g., (Dale 1992) (Walker 1995) (Moore & Pollack 1992).) Despite the many open questions, their model is one of the most comprehensive and familiar. Here I briefly sketch the main components of their model in order to identify the assumptions I adopt, and to set the stage for the empirical questions posed here.

The units of linguistic structure as defined in (Grosz & Sidner 1986) are segments. Fig. 3, taken from (Grosz 1977) and discussed in (Dale 1992), shows two discourse segments where one is embedded in the other. Each discourse segment has its own discourse segment purpose (DSP) in intentional structure, and a corresponding focus space (FS) in attentional state. Each focus space in attentional state represents the corresponding DSP, and the discourse entities in focus (Grosz & Sidner 1986). Attentional state represents the relative accessibility of focus spaces and the discourse entities they contain for the purpose

Seg	U	Spkr		
1	1	A	I'm going camping next weekend.	
	2		Do you have a two-person tent I could borrow?	
	3	B	Sure. I have a two person back-packing tent (e_1).	$FS_1 = \{DSP_1, \ldots e_1 \ldots\}$
2	4	A	The last trip I was on there was a huge storm.	
	5		It poured for two hours.	
	6		I had a tent (e_2), but I got soaked anyway.	$FS_2 = \{DSP_2, \ldots e_2 \ldots\}$
	7	B	What kind of a tent was it?	
	8	A	A tube tent.	
	9	B	Tube tends don't sand up well in a real storm.	
	10	A	True.	
	11	B	Where are you going on this trip?	
			...	
	15	B	Okay. I'll bring $\boxed{\text{the tent } (e_1)}$ in tomorrow.	

Figure 3: Motivation for Focus Spaces

of constraining the inferences involved in interpreting new utterances in the discourse, and in particular, discourse anaphoric noun phrases. In this excerpt, for example, two tents are mentioned, but the phrase *the tent* in utterance 15 (U_{15}) is unambiguously used to refer to one of them. The focus space associated with segment 1 in Fig 3 (FS_1) includes the discourse entity e_1, corresponding to the tent that speaker B introduces in utterance 3 (U_3). As the discourse progresses, each new utterance is incorporated into the linguistic structure and the attentional state is updated. Speaker A introduces a second tent in segment 2 pertaining to her last camping trip. With this utterance, e_2 is added to FS_2. When speaker B asks A for information about the current camping trip, segment 2 ends and segment 1 is resumed. The more recently mentioned tent (e_2) is hypothesized to be in global focus only from the time it is mentioned until DSP 2 is satisfied, at utterance 10. At utterance 11, DSP 1 and its attendant focus space is resumed and the original tent (e_1) is again in focus. Thus the discourse anaphoric noun phrase *the tent* in U_{15} is unambiguously inferred to evoke e_1.

Discourse segments are theoretical constructs in (Grosz & Sidner 1986). A contrasting view of the role of discourse segments is found in much of the discourse literature. In (Webber 1988) (Polanyi 1988) and (Hobbs 1985) for example, utterances contribute to semantic representations of discourse structure, but discourse segments are not explicitly represented in the model. Instead, when an utterance is processed, the interpretation increments what I will neutrally refer to as a node in a conceptual model of the discourse. The relations among nodes have been variously argued to comprise a tree (Webber 1991) (Polanyi 1988) (Hobbs 1985), a lattice (Moore & Pollack 1992), or any of various graph structures that can include cycles (Hearst

1994). In many of the examples presented in (Grosz & Sidner 1986), discourse segments have discrete boundaries. Recent work has exploited the potential for discourse segments to have sharp boundaries by investigating the correlations of independently verified segment boundaries with various surface cues, such as prosody (C.H.Nakatani, Hirschberg & Grosz 1995) (Grosz & Hirschberg 1992) (Litman & Passonneau 1995*b*), cue words (Morris & Hirst 1991) (Litman & Passonneau 1995*b*), and discourse anaphoric NPs (Litman & Passonneau 1995*b*). However, naturally occurring spoken language is often less easily segmented, as argued in (Passonneau & Litman To appear*b*). The transition from one segment to another may take more than one utterance. Conversely, a single utterance may contribute to distinct parts of the discourse model. The position I adopt here is that discourse segments are not theoretical primitives. I assume that one utterance could potentially contribute to distinct DSPs, and/or to distinct focus spaces, but that this occurs less often than not. I use the term segment simply to mean a sequence of utterances that are more closely related to one another semantically and pragmatically than they are to other utterances. I assume that most of the utterances within a segment contribute discourse entities to the same focus space, but that the transitions between focus spaces need not occur at a single utterance.

(Reichman 1985) claims that only discourse entities that are currently in focus are accessible for subsequent reference, and (Polanyi 1988) elaborates this claim to mean that only discourse entities in the right frontier of the discourse tree are accessible. In (Grosz & Sidner 1986) the weaker claim is made that there are different constraints on referring expressions within versus across segments. (Dale 1992) argues that no convincing evidence has been presented to substantiate either claim. Both (Grosz & Sidner 1986) and (Reichman 1985) suggest that the use of a full noun phrase where a pronoun would have been adequate correlates with segment or topic shifts. As discussed in the next section, analysis of a corpus of spoken narratives produces no evidence to support the latter claim.

In sum, I take discourse anaphoric noun phrases to evoke discourse entities that are already in the discourse model. I assume that discourse entities serve to index a semantic representation constituting the current description of the speaker's intended referent, and that these representations are updated as the discourse progresses. They derive from semantic entailments of the speaker's utterances, and from defeasible inferences (cf. (Hobbs et al. 1988)). When the discourse model is updated, the discourse entities that have been evoked are dynamically (re-)partitioned into distinct focus spaces. I assume that the focus structure partitions discourse entities into relatively more or less accessible sets, rather than into accessible versus inaccessible ones. I take the current focus space to be the most accessible, but relations among focus spaces, and the precise mechanisms for accessing previous focus spaces are open questions I do not address.

Informational Constraints

Much of the work on discourse processing cited above addresses the problem of natural language understanding of discourse anaphoric noun phrases. The problem of identifying constraints on the choice between pronouns and full noun phrases relates directly to constraints on the speaker, hence to natural language generation. Here, I briefly review relevant work on generating referring expressions.

Dale translates the Gricean (Grice 1975) injunctions to be informative and to be brief into his principles of *efficiency* and *adequacy*. These principles pertain only to the function of referring, requiring the information in a referring expression to be sufficient for unambiguous re-identification of the intended referent in the given context. Efficiency *requires that the referring expression used must not contain more information than is necessary for the task at hand.* Adequacy *requires that a referring expression should identify the intended referent unambiguously, and provide sufficient information to serve the purpose of reference.* (Dale 1989) (p. 70). To meet adequacy, (Dale 1989) requires discourse anaphoric NPs to be distinguishing descriptions. For any set of entities U, (Dale 1989) defines a distinguishing description of a discourse entity e in U to be a set of attribute-value pairs that are true of e, and of no other members of U. This enforces adequacy. He defines a minimal distinguishing description to be one where the cardinality of the attribute-value pairs cannot be reduced. This addresses efficiency.

(Reiter 1990*b*) argues that to generate maximally efficient NPs using Dale's framework is intractable. (Dale & Reiter 1994) later argue that maximal efficiency is psychologically implausible and empirically incorrect. It can be added that, in the absence of a non-arbitrary theory of lexical decomposition, cardinality of attribute value pairs is itself an arbitrary measure of efficiency. To address some of these problems, (Dale & Reiter 1994) adopt (Reiter 1990*a*)'s notion of *local brevity*, a preference for the shorter of any two distinguishing descriptions. Where "the small dog" and "the sleeping female dog" are distinguishing descriptions, local brevity is satisfied by the former expression.

I propose an alternative principle of *economy* based more directly on the lexico-grammatical resources of the language. Economy relies on a syntactic categorization of noun phrase types into distinct preference classes. At the same time, it makes use of entailment to compare two expressions semantically. The three preference classes, in descending order, are: definite pronouns; unmodified noun phrases consisting solely of a determiner and a head noun; and finally, noun phrases with pre- or post-modifiers. Given two expressions E_i

and E_j, E_i is the more economical one if the semantic interpretation of E_j entails the semantic interpretation of E_i, and E_i contains fewer words. The semantic interpretation of a definite description entails the semantic interpretation of a definite pronoun when they have the same grammatical features (gender, number, animacy, definiteness), and so on. Within the class of bare noun phrases, I follow (Dale & Reiter 1994) in assuming a preference for heads that denote basic category information (e.g., *house* is preferred over *domicile*).

Dale's implementation of adequacy (Dale 1989) (Dale 1992) is elegant and simple. The ae-describe relation proposed below incorporates his algorithm. He defines the discriminatory power ($\mathcal{F}$) of an attribute-value pair <A, V> that is true of a discourse entity e in a universe of entities U in terms of the cardinality N of U, and the total number n of entities in U that <A, V> is true of:

$$\mathcal{F}(< A,\, V >,\, U) \; = \; \frac{N-n}{N-1}$$

$\mathcal{F}$ ranges in value from 0 to 1. If $< A_i,\, V_j >$ is true of only one of the entities in the set U, then $\mathcal{F}_{<A_i,\, V_j>}$ is 1, and $< A_i,\, V_j >$ is a distinguishing description of the entity.

The algorithm for constructing a distinguishing description (Dale 1992) of e in U, given a set $\mathcal{P}$ of attribute-value pairs that are true of e, briefly works as follows. U is set to be the entities in the current global focus. Compute $\mathcal{F}$ for each member of $\mathcal{P}$. If all values of $\mathcal{F}$ are 0, no unique description can be constructed. Otherwise, increment the description with the attribute-value pair having the highest value, and reset U to be only those entities from U that the selected attribute-value pair is true of. Repeat this process, terminating when an attribute-value pair with a discriminatory power of 1 has been selected. The selected attribute-value pairs constitute the input to the surface generator.

Centering

(5) a. A woman (e_j) went to the bookstore.

b. Afterwards, she (e_j) gave her aunt (e_k) a new book.

c. She's (e_j) a true bibliophile.

Centering is a model of local focus of attention (Joshi & Weinstein 1981) (Grosz et al. 1983) (Kameyama 1985). Among the motivations for centering is the goal of limiting the domain of inferences pertaining to the interpretation of under-specified expressions, such as definite pronouns. Example (5) (repeated from Fig. 2) illustrates that where the semantics of the utterance and commonsense reasoning do not discriminate among possible referents for an ambiguous pronoun, there is an independent effect of local attentional constraints. Centering predicts that the preferred interpretation of the pronoun in (5c) is the discourse entity corresponding to the woman mentioned in the grammatical subject of the two previous utterances (e_j). Note that in this context, neither the pronoun alone nor the utterance distinguishes between the entities for the woman (e_j) and her aunt (e_k), hence the pronoun is not informationally adequate (cf. *Introduction*).

(6) a. A woman (e_j) went to the bookstore.

b. On her way home, she$_j$ saw her aunt (e_k).

c. She (e_j) looked pale.

In (Kameyama 1985), examples like (6) are presented to illustrate how commonsense reasoning and lexical semantics override centering predictions for default pronoun interpretation. The discourse is structurally parallel to (5). Again, centering would predict that e_j (for "the woman") should be the default interpretation of the pronoun in (6c). But (6c) is interpreted as a continuation of the description of the perceptual event in (6b), where it is the aunt whose appearance is in focus. The discourse entity corresponding to the aunt (e_k) has already been described as the object of the perceptual event in (6b). The predication "*look pale*" describes a perceptual state with a perceived object (here the grammatical subject), and an implicit perceiver. Only one of the two female discourse entities in the context satisfies the presuppositions of the predication in (6c). If the pronoun subject is interpreted to evoke e_j, e_k and e_j maintain the same semantic relations to one another—of perceived and perceiver—thus maintaining the coherence of the discourse. A much greater inference load is required for the alternative interpretation. Under this interpretation, occurrence of a new perceptual event must be inferred in which e_k and e_j have reversed their roles of perceived and perceiver. Use of appropriately weighted abduction rules (Hobbs et al. 1988) would result in the former interpretation.

CORPUS ANALYSIS

In this section I review results derived from an observational study of discourse anaphoric noun phrases in a corpus of spoken monologues in order to propose initial specifications for a generation model. In the first subsection below I describe the corpus data. In the next subsection, I define three classes of informational status based on informational adequacy and economy: discourse anaphoric NPs are either well-specified, under-specified or over-specified. To minimize the degree of hand coding, I heuristically identify a large percentage of the over-specified discourse anaphoric noun phrases in the corpus. I present results showing that over-specified discourse anaphoric noun phrases occur relatively infrequently, and that there is only a weak tendency for them to occur at hypothesized shifts in global focus structure. In the last subsection, I discuss three types of attributes used in discourse anaphoric NPs in this corpus. These play a role when it is necessary to generate less economical expressions.

Data Coding

The corpus consists of ten of the twenty narrations in (Chafe 1980). Chafe recorded and transcribed subjects who had been asked to view the same movie and describe it to a second person. In many respects, the narrators share the same model of what they are to convey. The movie contained seven sequential episodes about a man picking pears. It had a vivid sound track, but no language. The transcriptions (Chafe 1980) represent all lexical and non-lexical articulations (e.g., "uh, um, tsk") in ordinary orthography. Chafe identified three types of prosodic phrases from graphic displays of intonation contours, and used punctuation to distinguish them. A period indicates a phrase terminated by a pitch fall, and a question mark indicates final level or rising pitch. A comma indicates a prosodic phrase whose boundary tone signals continuation with the following phrase. The prosodic phrase numbering shown in column 2 of Fig. 4 has two fields corresponding to two levels of prosodic phrases. The first field is for sequential numbering of prosodic phrases ending in '.' or '?', and the second field is for the continuation phrases ending in ','. Chafe's transcriptions include pause location and duration, but only location is shown here. The ten narrations contain about 7,000 words, just over 1,000 prosodic phrases, and 768 clauses of a type described below.

As part of a long term study of the relationship between linguistic features and discourse structure, (Passonneau & Litman To appear*a*) each discourse anaphoric NP in the corpus was coded for coreference relations by two independent coders, with consensual resolution of conflicts. A surface constituent is considered to be a discourse anaphoric noun phrase (NP) if it occurs in free variation with syntactically prototypical NPs, and corefers with a preceding NP (cf. (Passonneau 1994*b*)). One type of empty category is also included, namely zero pronoun subjects that occur in certain utterance fragments (cf. prosodic phrase 14.1 in Fig. 4) or utterances with (non-embedded) conjoined tensed verb phrases.[1]

For purposes of examining coreference relations across clauses, a clause unit referred to as a functionally independent clause (FIC) was defined (Passonneau 1994*b*). Briefly, an FIC is a tensed clause that is not a verb argument, not a restrictive relative clause, nor one of a set of formulaic "interjection" clauses (e.g., "*You know*", as in PP 15.2 of Fig. 4; for full details cf. (Passonneau 1994*b*)). Utterances (column 3 of Fig. 4) were defined so as to be a sequence of units that completely covers a narrative. (The boxed NPs in Fig. 4

Seg	PP	U	
4	...	...	...
	13.1	27	Because he's (e_4) looking at the girl (e_5).
	14.1	28	[pause] (ZERO) (e_4) Falls over,
5	14.2	29	[pause] uh there's no conversation in this movie (e_6).
	15.1	30	[pause] There's sounds (e_7),
	15.2		you know,
	15.3		like the birds (e_8),
	15.4	31	but there [pause] the humans beings (e_{10}) in $\boxed{\text{it } (e_6)}$ don't say anything.
6	16.1	32	[pause] $\boxed{\text{He } (e_4)}$ falls over,
	16.2	33	and then these three other little kids about his same age come walking by.

Figure 4: Interruption Segment

are explained below.) An FIC_{i+1} begins a new utterance if it is immediately adjacent to the preceding FIC_i. U_{33} in Fig. 4 exemplifies this case. Otherwise the utterance begins at the onset of the prosodic phrase within which the next FIC begins. The utterance can thus begin with a pause and/or non-lexical material as in U_{29}, or a fragment as in U_{31}. Note that U_{30} covers three prosodic phrases (15.1-15.3).

Column 1 of Fig. 4 indicates the sequential number of discourse segments. The discourse segmentation in the corpus was derived from an empirical study described in (Passonneau & Litman 1993) (Passonneau & Litman To appear*a*). Each narrative was segmented by 7 new, untrained subjects.[2] Subjects were given transcripts that had been formatted with one prosodic phrase per line. They were instructed to place segment boundaries between prosodic phrases to indicate when the narrator had finished one communicative task and begun a new one. To focus their attention on the segmentation criterion, subjects were also instructed to label segments with a brief description, in their own words, of the speaker's communicative intention. The size and number of segments per subject per narrative varied widely, from a boundary assignment rate of 5.5% to 41.3% (Avg.=16%), with segment lengths ranging from 1 to 49 prosodic phrases (Avg.=5.9). Despite this variation, the different subjects assigned the same boundary locations far more frequently than would be predicted by chance. The observed number of agreements on boundaries was highly unlikely ($.6 \times 10^{-9} \leq p \leq .114 \times 10^{-6}$, using Cochran's Q (Cochran 1950); cf. (Litman & Passonneau 1995*a*)), hence statistically significant. As discussed in (Litman & Passonneau 1995*a*), boundaries identified by at least 3 subjects reached statistical significance, hence are taken to be empirically validated. The segment shown in Fig. 4 is typical, with the boundary at U_{28},U_{29} identified by 5 subjects and the one at U_{31},U_{32} by 7 subjects. The 5 subjects in agreement on the segment beginning and end all labeled the segment as an interruption about the characteristics of the movie.

As I concluded in the *Background* section, I assume that the current focus space represents a set of discourse entities that is currently relatively more accessible for discourse anaphoric noun phrases. I also assume that part of what relates utterances within a segment is that the speaker has a common intention (or Discourse Segment Purpose (Grosz & Sidner 1986)) while producing them. For present purposes, I hypothesize that a single empirically derived discourse segment corresponds to a focus space (or a resumption focus space; cf. comments below). I make no assumptions about whether these are minimal focus spaces; some may embed other focus spaces. I also make no assumptions about how the focus spaces are related to one another. We will see in the following subsection that there is only weak evidence that speakers use overly informative noun phrases at segment onsets. This accords with results from other studies that there is no single feature associated with segment boundaries from prosody, cue phrases, referring expressions (Passonneau & Litman To appear*a*) or centering (Passonneau To appear). However, segments do seem to establish an appropriate context for producing informationally adequate and economical referring expressions. In this section, I present some data supporting this hypothesis. In the section on evaluation, I demonstrate that an algorithm for generating discourse anaphoric noun phrases replicates the referring expressions in the Pear story narrations better when the context of discourse entities and their known attributes is partitioned into focus spaces corresponding to the empirically derived segments.

Informational Adequacy and Economy

Fig. 5 identifies three possibilities regarding the semantic informativeness of an NP relative to its context, based on adequacy and economy. Context here refers to the containing utterance, and to the set of accessible discourse entities. The first utterance introduces three discourse entities (e_1, e_2, e_3), of type `man'`, `ladder'` and `tree'`, respectively. The boxed pronouns in the first continuation sentence (2a) of Fig. 5 are both adequate and economical, thus well-specified. The first pronoun—"*he*"—is adequate with respect to the accessible discourse entities, because only one of the discourse entities is a male being. The second pronoun is adequate when the semantics of the containing utterance is taken into account: only the ladder can be moved to a different tree. Both pronouns are economical because the only expression with fewer words, a zero pronoun, would be ungrammatical. In the alternative continuation sentence in (2b), the full noun phrases are adequate but not economical, thus they are over-specified. The pronominal NP in (2c) is economical but inadequate: "*it*" could refer either to the ladder or the tree, thus is under-specified.

The three possibilities illustrated in Fig. 5 are not exhaustive. In addition to evoking an existing discourse

	1.	A man (e_1) saw a ladder (e_2) leaning against a pear tree (e_3.) FS = {e_1, e_2, e_3 }
	2a.	Later, $\boxed{\text{he}_1}$ moved $\boxed{\text{it}_2}$ to a different tree.
WELL-SPECIFIED		+adequate; +economical
	2b.	$\boxed{\text{The man}_1}$ moved $\boxed{\text{the ladder}_3}$ to a different tree.
OVER-SPECIFIED		+adequate; −economical
	2c.	$\boxed{\text{It (?)}}$ was tall.
UNDER-SPECIFIED		−adequate; +economical

Figure 5: Three Categories of Informational Status

entity, a discourse anaphoric NP can add information about it. I refer to this case as over-determination. A subsequent reference to the man mentioned in Fig. 5 with the phrase "*the contented pear picker*" would be over-determined. Only one over-determined discourse anaphoric NP occurred in the 10 narratives examined here, although in other modalities or genres, the frequency can be higher.[3] It is outside the scope of the present paper to model the use of over-determined discourse anaphoric NPs, attributive NPs (Kronfeld 1986), or NPs with other functions beyond providing sufficient information to identify an existing discourse entity.

There were 1,233 discourse anaphoric noun phrases in the corpus. Instead of classifying each token, I heuristically identified a set of NPs that were potentially over-specified. After the corpus was tagged for part of speech (cf. n. 1), I sorted the discourse anaphoric NPs in the corpus into the three categories of full NPs (NPs; N=563), independent pronouns (PROs; N=544), and zero pronominals (ZPs; N=126). Note that pronouns and full noun phrases have roughly the same frequency. The syntactically more restricted zero pronominals occur much less often. The three categories are ordered with respect to relative informational content (>): NP > PRO > ZP. This ordering is inversely related to economy. Using the coreference coding and the ordering relation >, I identified all pairs of co-indexed NPs where $NP_2 > NP_1$. The second NP of each pair (NP_2) is potentially over-specified because it contains more information than the antecedent (NP_1) that evoked the same discourse entity. Note that NP_2 is not necessarily over-specified. Where NP_2 and NP_1 evoke an entity e_1, NP_2 may need to have more semantic content than NP_1 if a new entity whose attributes partly overlap the attributes of e_1 has been introduced in the context intervening between NP_1 and NP_2.

The heuristic described above produced 128 discourse anaphoric NPs in the corpus that were potentially over-specified. One of these was in fact over-determined (illustrated in Fig. 6); it provides new information about the discourse entity it evokes in the adjective "*old*". It seems mainly to provide contrast (cf. "*that old man*" in U_{85} vs. *those little boys*" in U_{84}). I evaluated the 127 remaining cases to determine which were in

Seg	PP	U	
17	35.1	84	[pause] You just know that those little boys are going to go back,
	35.2		[pause] to where the pear tree was,
	35.3	85	and you just know that old man 's going to see [pause] [pause] these little boys coming and say "Ha.
	35.4		.. You're the ones who stole the pears."

Figure 6: An Over-determined NP

fact over-specified with respect to the current utterance and the hypothesized focus space.

As noted above, I hypothesize that each discourse segment corresponds to a focus space. For each discourse anaphoric noun phrase evoking a discourse entity 'e,' I take the set of accessible discourse entities to be the union of the current focus space (FS_c) containing the discourse anaphoric NP with the most recent focus space containing 'e' (FS_e). Given the assumptions discussed in the *Background*, consider these two possibilities regarding the structural relation between FS_c and FS_e. They may be in a relation that justifies keeping them distinct (e.g., siblings in a discourse tree), or FS_c may be a resumption of FS_e, in which case they should be considered a single focus space. (Fig. 3 illustrates a resumption.) I take here the more conservative approach of assuming that if the most recent segment containing the discourse entity is not the current segment, then the current segment is a resumption. I do not include any intervening focus spaces if they exist, on the assumption that this would give too weak a test of the utility of the focus spaces when the evaluation results are examined. To sum up, informational adequacy and economy is computed relative to the context consisting of the union of FS_c and FS_e, which may be the same.

For the boxed pronoun "*it*" in U_{31} of Fig. 4, the current focus space FS_5 (co-indexed with the segment number) contains the discourse entities evoked in U_{29}-U_{30} of segment 5. For the sake of clarity, I show the discourse entities here as arguments of a type predicate corresponding to the lexical head of the evoking NP: FS_5 = {movie(e_6), sounds(e_7), birds(e_8)}. FS_5 is both the current focus space and the most recent focus space containing e_6. With respect to this context, the pronoun "*it*" is adequate and economical because the utterance context for U_{31} requires the discourse entity evoked by the pronoun "*it*" to be something that human beings can be "*in*." Intersecting the utterance context with FS_5 reduces the set of discourse entities to a singleton: $FS_5 \cap U_{31}$ = {movie(e_6)}.

The pronoun "*he*" in U_{32} of Fig. 4 presents the case where the current focus space (FS_6) does not contain the relevant discourse entity. FS_4 is the most recent focus space that does. FS_6 is empty because U_{32} is the first utterance of the new segment, so $FS_6 \cup FS_4 = FS_4$. Although only part of the discourse segment for FS_4 is shown in Fig. 4, one discourse entity in FS_4 has already been described as having fallen over: U_{39} is essentially a repetition of U_{28}. Thus the intersection of the utterance context for the relevant pronoun with the accessible discourse entities again reduces to a singleton set: $FS_4 \cap U_{32} = \{boy(e_4)\}$.

Of the 127 potentially over-specified discourse anaphoric NPs mentioned above, only 68 were in fact over-specified. Thus the heuristic used here identifies only 6% (68 of 1,223) of the discourse anaphoric NPs in the corpus to be over-specified. The actual frequency is possibly somewhat higher, but still a relatively small percentage. The first specification for model generating discourse anaphoric NPs will thus be to block over-specified discourse NPs. Later, this this constraint will be relaxed.

To determine whether over-specified NPs occur at segment boundaries, I sorted the 68 over-specified NPs as shown in Table 1. An over-specified NP that occurs in the first utterance of a new discourse segment is classified as a segment onset (column 2). The discourse segments arguably contain intra-segmental shifts of attention associated with changes in temporal aspect, or shifts in discourse reference time (for definitions assumed here, cf. (Kameyama, Passonneau & Poesio 1993)). Over-specified NPs that did not occur at a segment onset, but occurred at a temporal shift, appear in column 3. The fourth column ("Other") includes, e.g., repetitions, repairs, contrastive NPs and unexplained cases. The rows labelled "Same" and "Prev" indicate whether the antecedent of the over-specified NP occurred in the same discourse segment or a previous one. About half were "Same" and half were "Prev". Overall, about 72% of over-specified NPs occurred at a discourse segment onset (29%) or temporal shift (43%), thus there seems to be a tendency for over-specified NPs to correlate with a shift in global focus. Note however, that the reverse implication does not hold, that is, it is not the case that a segment shift is likely to be signaled by an over-specified NP.

Antecedent Segment	Segment Onset	Temporal Shift	Other	Total
Same	N.a.	21 (58%)	15 (42%)	36 (100%)
Prev	20 (63%)	10 (31%)	2 (6%)	32 (100%)
Totals	20 (29%)	29 (43%)	19 (28%)	68 (100%)

Table 1: Over-Specified NPs

Focused Attribute Sets

To generate relatively informative discourse anaphoric NPs, it is necessary to determine how attributes are selected from the information known about a discourse entity. With respect to discourse anaphoric NPs in the Pear stories, NP modifiers are typically what I refer to as *focussed attributes*. Focussed attributes comprise the following three categories of relevance. An attribute can be linked to a *lexical focus* because the lexical item was used in the most recent reference. For example, in Fig. 6, three boys are referred to in U_{85} as "*these little boys*," repeating the prevrious phrase mentioned in the reference to the boys (in U_{84}). (Here "*little*" is an example of the use of an unnecessary modifier; cf. discussion of (Dale & Reiter 1994) in the *Background* section.) Lexical focus occurs in other genres, e.g., technical writing (Reiter, personal communication).

Seg	PP	U	
	3.8	13	and it's just a monotonous kind of thing for him (e_1).
4	4.1	14	[ps] And a man (e_2) comes along with a goat,
	4.2	15	[ps] and the goat obviously is interested in the pears.
	5.1	16	But the man (e_2) just walks by with the goat.
5	6.1	17	And $\boxed{\text{the man } (e_1) \text{ up in the tree}}$ doesn't even notice.

Figure 7: Location Focus

In *location focus*, the focussed attribute set may specify the most recently mentioned location of an entity. The subject NP in U_{17} of Fig. 7 refers to one man as "*the man up in the tree*" to distinguish him from the second man who came by with a goat. The tree is the last mutually known location of the former. Finally, in *event focus*, attributes are in focus that pertain to a key narrative event that the entity has been a participant in. Intuitively, an event is more central to a narrative the more difficult it is to relate the narrative without mentioning that event. Operationally, key events occur more frequently than others both within and across narratives. The main adult character is often described as "*the pear picker*," or as "*the man who was picking pears*" (see U_{108} of Fig. 8), and so on; the other main character is often described as "*the thief*," "*the boy who stole the pears*," "*the boy with the pears*," and so on. Events are key elements of narrative, and location is important in interpreting events. Consequently, event and location focus may be genre-dependent strategies.

Seg	PP	U	
21	63.1	105	[ps] So they (e_1)'re walking along,
	63.2	106	and they (e_1) brush off their pears (e_3),
	63.3	107	and they (e_1) start eating it (e_3).
22	64.1	108	Then $\boxed{\text{they } (e_1)}$ walk by-[ps] $\boxed{\text{the man who was picking the pears } (e_2)}$

Figure 8: Event Focus

How to order the focussed attributes for a discourse entity is a topic for further investigation. Here, I simply assume that the three types of attribute sets mentioned above—where applicable—are in focus. I also assume that the focussed attributes of an entity (FA_e) are updated as the discourse progresses.

ENFORCING ADEQUACY AND ECONOMY: AE-DESCRIBE

The data presented in the preceding section provide specifications for generating informationally cooperative NPs. I presented evidence that discourse anaphoric noun phrases in the spoken language corpus analyzed here are rarely over-specified with respect to the context consisting of the hypothesized global attentional state and the current utterance. Further, this observation is independent of the distinction between pronouns and full-noun phrases. I have not attempted to quantify the frequency of under-specified noun phrases, but I will provisionally assume they are relatively infrequent. In this section, I present an algorithm that constrains discourse anaphoric NPs to be adequate and economical, i.e., well-specified.

In the *Background* section, I reviewed the algorithm for enforcing adequacy and efficiency presented in (Dale 1989) (Dale 1992). The algorithm presented here—ae-describe—enforces adequacy and economy. Ae-describe incorporates a modified version Dale's algorithm for constructing distinguishing descriptions that differs in the following respects.

- The current utterance is part of the input context; it serves as a filter on the discourse entities in the current focus of attention.[4]
- The current focus of attention is the union of the current focus space and the most recent focus space containing the discourse entity to be described.
- Ae-describe enforces economy, rather than efficiency. Economy is defined in terms of a preference ordering based on a combination of syntactic and semantic factors instead of cardinality of attributes.
- Economy is enforced by a generate-and-test method for computing distinguishing descriptions. The three preference classes, ordered by decreasing economy, are:
  - PRO: a third person definite pronoun is generated if it is a distinguishing description;
  - MIN: else a minimal noun phrase where the syntactic criterion is that the attributes must map to a noun phrase consisting solely of a determiner and head noun (no modifiers), and the semantic criterion is that the head must denote a basic category (Reiter 1990*b*).

– FULL: else any full noun phrase expressing a focussed attribute, either a bare NP whose head is more specific those generated in the MIN class (e.g., lexical focus), or an NP with modifiers.

Ae-describe is a 4-place relation among a discourse entity E, a surface NP, the current utterance context λU, and the focus space context FS. By definition, discourse anaphoric NPs must be definite (but cf. (Passonneau 1994*a*)). Ae-describe requires a discourse anaphoric NP to be a distinguishing description (Dale 1992) of E relative to FS $\cap$ λU. For generation, NP is solved for given an instantiation of the remaining three arguments; for understanding, E is solved for. While my focus here is on generation, I touch on understanding at the end of this section and in the conclusion.

Seg	PP	U	
5			...
		27	Um anyway,
			$\boxed{\text{a guy } (e_2)}$ comes by leading a goat (e_3).
6		29	And the goat (e_3)'s aaarrr but and they (e_4) don't talk to each other,
		30	they don't e I don't think they (e_4) even look at each other,
		31	and $\boxed{\text{the guy } (e_2)}$ walks by.
		32	And you watch the goat,
			disappearing all the way,
		33	and then then you're back to the man (e_1) in the tree.

FS_6:	{goat(e_3), men(e_4), guy(e_2), pear_picker(e_1) ...}
FS_6: $\cap$ U_{31}	{ guy(e_2), goat(e_3), pear_picker(e_1) }

Figure 9: Full NP: Lexical Focus

For generating the full noun phrase *"the guy"* in U_{31} of Fig. 9, the input arguments of $ae{-}describe$ are:

(1) $ae{-}describe(e_2,\ \_NP, \lambda X\ walks\_by'(X),\ FS_5 \cup FS_6)$

The utterance context is assumed to be a feature structure co-indexed with any existing discourse entities. By using the utterance context as part of the input in solving for NP, given information that appears anywhere in the current utterance can filter entities from the discourse context before the discriminatory power (Dale 1989) of attribute-value pairs is computed. New information about an entity in the utterance is not mutually known, and thus will have no discriminatory power (Dale 1989).

The relevant components of attentional state are shown at the bottom of Fig. 9. FS_6 represents discourse entities for the goat (e_3) that was introduced in the previous segment (segment 5), the set of men (who don't talk to each other; e_4) consisting of the man picking pears (e_1; introduced in an earlier utterance not shown here), the man who walked by with the goat (e_2), and the pear picker discussed in FS_5 (e_1). For mnemonic purposes, Fig. 9 represents each discourse entity as the argument of a predicate that specifies a discriminatory type, rather than representing a set of attribute-value pairs for each discourse entity. I assume that as a result of resolving the plural pronoun "*they*", FS_6 contains a representation of a set entity (`men'`(e_4)) as well as the two male entities in the set (`guy'`(e_2) and `pear_picker'`(e_1)). FS_6 is thus the union of current focus space with the most recent focus space containing the discourse entity e_3.

The result of filtering the current focus of attention with the utterance context eliminates the set entity, yielding $FS_6 \cap U_{31}$ with three entities, as shown in Fig. 9. Ae-describe first tries to generate a pronoun, but fails. The semantically acceptable pronoun "*he*" fails to distinguish between `guy'`(e_2) and `pear_picker'`(e_1). Next ae-describe tries to generate a minimal noun phrase (MIN). The attribute value pairs lexicalized as what (Dale 1992) and (Reiter 1990*b*) refer to as the basic type of e_2—"*man*"—will be selected in order to generate a noun phrase with a common noun as head and no modifiers. The resulting description will not distinguish between e_2 and e_1, so the next step tried by ae-describe will be to select attribute value pairs belonging in the focussed attributes of e_2. These attributes include the lexical FA_{e_2}: "*guy*". As noted in the preceding section, I do not attempt to control how a given focussed attribute is selected. Note however, that the phrase "*the man in the tree*" in U_{33} of Fig. 9 would be generated by an analogous process, except that here the most recently mentioned location of e_1 is the selected focussed attribute.

Seg	U	
1	1	Okay.
		Um the s the scene (e_1) opens up with um you see a tree (e_2),
		k kay?
	2	And there's a ladder (e_3) coming out o of the tree (e_2),
	3	and there's a man (e_4) at the top of the ladder (e_3),
	4	you can't see him (e_4) yet.
2	5	And then $\boxed{\text{it } (e_1)}$ shifts,
	6	and you see him (e_4),
		plucking a pear (e_5) from the tree (e_2).

Figure 10: Pronoun

Generating the boxed pronoun in Fig. 10 follows a process analogous to the generation of the full noun phrase described above. The intersection of the utterance context in U_5 (λ X `shifts'`(X)) with the accessible discourse entities in $FS_2 \cup FS_1$ yields a singleton set consisting of `scene'`(e_1); of the available discourse entities, only scenes are described as shifting. Ae-describe first tries to generate a pronoun. The pronoun "*it*" is a distinguishing description (Dale 1992) in this context.

An analogous process applies to understanding discourse anaphoric noun phrases. For example, given the pronoun in U_5 of Fig. 10, the arguments of ae-describe would be as follows: the entity variable E would be uninstantiated, NP would be instantiated to "*it*", and the utterance and discourse context would be as described above. Given a distinguishing description, there is guaranteed to be exactly one solution to E. However, the search problem increases with the size of the context. Partitioning the search space into focus spaces controls the search through the discourse model to some degree. Integrating centering with ae-describe as described in the next section adds a further degree of control to the search process.

THE NEED FOR CENTERING CONSTRAINTS

Violations of Economy and Adequacy

Ae-describe constrains discourse anaphoric expressions to be informationally well-specified, meaning both adequate and economical. However, as noted in the *Background* section, centering predicts that speakers can use under-specified pronouns. An under-specified pronoun, though adequate in the sense of communicating the intended meaning to the hearer, violates informational adequacy. In this section, I summarize relevant components of the centering model to show when centering predicts an under-specified pronoun. In addition, I show how centering can explain certain over-specified full noun phrases. I begin by illustrating both types of violation of informational constraints.

Fig. 11 illustrates an excerpt from the Pear corpus. Here the speaker discusses four boys: the boy who has fallen off his bike (e_2), and the three boys (e_1) who help him "*get all situated again*" (the set e_1 consists of e_3, e_4 and another unspecified boy). Despite the need to discriminate among several boys, all but one of the three masculine singular pronouns in the excerpt are unambiguous within their utterance contexts. For example, only one boy in the prior context (not shown here) has been riding a bike, so only this boy (e_2) is in a state that satisfies the preconditions of the assertion made in U_1: "*he's fallen off his bike.*" U_8 describes a second boy (e_3) picking up a rock; only this boy satsifies the preconditions of U_9, which describes this boy as throwing the rock out of the road.

The analysis of U_{10} appears to leads us off track, but raises the pertinent issue of where it is within an utterance that discriminating semantic information occurs. Ae-describe does not take this into account,

Seg	PP	U	
7	...	...	...
	18.1	1	[pause] And they (e_1) see that he's (e_2) [pause] fallen off his bike,
	18.2	2	[pause] and his pears [pause] have scattered,
	18.3	3	[pause] and uh [pause] they (e_1) walk over to help him (e_2),
	18.4	4	[pause] they (e_1) g [pause] gather all the pears
		5	and put them in the basket,
	18.5	6	[pause] and [pause] one (e_3) of the guys (e_1),
	18.6	7	helps him (e_2) brush off the dust,
	18.7	8	[pause] and another guy (e_4) picks up the rock,
	18.8	9	and he (e_4) throws it out of the road,
	18.9	10	[pause] and [pause] he (e_2) gets all situated again,
	18.10	11	and $\boxed{\text{he } (e_2)}$ takes off,

FS_7: $\{boys(e_1), boy(e_2), fell(e_2), boy(e_3), brushed\_off(e_3, e_2), boy(e_4), threw\_rock(e_4), \ldots\}$

$FS_7 \cap U_{11}$ $\{ boy(e_2), boy(e_3), boy(e_4) \}$

Figure 11: Under-Specified Pronoun

whereas centering effectively does. The subject pronoun in U_{10} is informationally adequate because only the boy who fell (e_2) is in a position to "*get all situated again.*" Yet in the absence of the full prosodic cues from the speech signal, a garden path effect is produced. The first available interpretation when reading the text is that the subject pronoun in U_{10} is intended to corefer with the subject pronoun of the preceding utterance, i.e., is used to refer to the boy who threw the rock (e_4). This false interpretation can only be rejected when processing the final verb phrase in U_{10}, "*gets all situated,*" and especially the anaphoric adverbial "*again.*" When we turn to the discussion of Fig. 12 we will see a similar example where I argue that a full noun phrase subject is used instead of a pronoun precisely to avoid such a garden path effect.[5] The integrated model I present in a later section is designed to block examples such as the pronoun subject in U_{10} of Fig. 11. However, instead of blocking such examples altogether, a more complete model could predict the need for the distinctive prosody found here, namely the long pauses before and after the initial "*and,*" as well as a distinctive pitch contour.[6]

Turning now to the key example in Fig. 11, the pronoun in U_{11} is informationally under-specified. The predication "*take off*" can be used for locomotion on bicycle or on foot, so any one of the three boys referred to here (e_2, e_3, or e_4) is a potential referent. This is illustrated in Fig. 11 by the intersection of the presumed focus space (FS_7) with the utterance context ($FS_7 \cap U_{11}$). Again, there is a tendency to interpret this pronoun subject as coreferential with the pronoun subject of the preceding utterance. In contrast to the

Seg	U	
2	4	And he (e_1) fills his thing with pears,
	5	and ZERO (e_1) comes down
	6	and there's a basket he (e_1) puts them in.
3	7	And you see passerbyers on bicycles and stuff go by.
	8	A-nd then a boy (e_2) comes by,
		on a bicycle,
	9	the man (e_1) is in the tree,
	10	and $\boxed{\text{the boy } (e_2)}$ gets off the bicycle,
	11	and ZERO (e_2) looks at the man (e_1),

FS_3:	{man(e_1), in_tree(e_2), boy(e_2), on_bicycle(e_2), ...}
$FS_3 \cap U_{10}$	{ boy(e_2) }

Figure 12: Over-Specified Full NP

garden path effect found when reading U_{10}, here the default interpretation is correct because it is the boy on the bicycle (e_2) who takes off conforms to the actual events in the movie.

Given that speakers produce utterances with under-specified references in ordinary spontaneous speech, we need a processing model that can interpret them. It is all the better if the model is also efficient. In the next section (*Centering Constraints*), I explain how the centering model provides a correct and efficient interpretation strategy for under-specified pronouns such as the one in in Fig. 11. I will also explain how to adapt centering in order to generate such pronouns, thus to simulate actual human discourse.

Fig. 12 illustrates the second motivation for centering discussed here, a violation of informational economy. As in the preceding example, the violation of an informational constraint is communicatively appropriate, and in fact insures more effective communication. In U_{10} of Fig. 12, the subject NP ("*the boy*") is over-specified. The utterance context is sufficiently specific to discriminate between the two accessible male discourse entities (the man e_1 and the boy e_2; cf. the singleton member of the set $FS_3 \cap U_{10}$). The boy (e_2) is mutually known to have been on a bicycle in the previous discourse. Temporal processing (cf. (Kameyama et al. 1993)) supports the inference that the boy is still on the bicycle after U_9. Thus a definite pronoun in U_{10}, an utterance describing someone getting off a bicycle, would be well-specified, and ae-describe would generate the definite pronoun "*he*". However, a pronoun would initally be interpeted as corefering with the subject of the preceding utterance, thus producing a garden path effect. The interpretation that the speaker is referring to the man (e_1), would only be rejected after processing "*bicycle*." Because centering predicts the garden path interpretation, it can be used to block the use of a pronoun in such contexts.

10 [pause] and [pause] he (e_2) gets all situated again,

Cfs: [e_2]

Cb: e_2

Cp: e_2

11 and $\boxed{\text{he (X)}}$ takes off,

Cfs: [X]

Cp: X

Cb: Y

Center continuation: $Cb(U_{11}) = Cb(U_{10}) = Cp(U_{11})$

$X = Y = e_2$

Figure 13: Centering Constructs: Example 1

In sum, informational adequacy and economy need to be relaxed in the two cases identified above. In the next section, I briefly review aspects of centering that can be used to relax informational constraints in precisely these two types of discourse context.

Centering Constraints

Here I present only specific aspects of centering to account for the two types of examples illustrated above. I rely primarily on the recent summary article in (Grosz, Joshi & Weinstein 1995), although versions of centering with distinct constructs and rules can be found in (Kameyama 1986) (Brennan, Friedman & Pollard 1987) (Walker, Iida & Cote 1994) and (Passonneau 1991), and elsewhere. For an overview of recent empirical studies, I refer the reader to (Prince, Joshi & Walker To appear).

Centering constrains the inferences that relate adjacent utterances.[7] More specifically, centering constrains *realization* of discourse entities.[8] (Grosz et al. 1995) argue that a combination of syntactic, semantic, and pragmatic factors must be taken into account in characterizing whether an expression U *directly realizes* a discourse entity e, where e is the interpretation assigned to U. A discourse entity can also be *realized* indirectly in the interpretation of an utterance. They conclude that specification of the discourse entities *realized* in an utterance "depends on the semantic theory one adopts." While most of the centering literature deals with discourse entities evoked by singular definite noun phrases, (Grosz et al. 1995) state that the theory should be extended to handle plural and quantified noun phrases, indefinites, and in addition, that "events and other entities that are more often directly realized by verb phrases can also be centers."[9] At first, I will restrict my attention to singular definite noun phrases. Later, for the example in Fig. 12, I will discuss centering and indirectly realized temporal entities.

Utterances U_{10} and U_{11} from Fig. 11 are repeated in Fig. 13 in order to illstrate centering constructs. The discourse entities realized in one utterance comprise a partial domain of entities that might be realized in the next utterance. These entities are referred to as the *forward-looking centers (Cfs)* (Grosz et al. 1995), and are partially ordered to reflect their likelihood of being re-mentioned. The precise means for ordering the Cfs is not of concern here.[10] Fig. 13 shows a discourse entity e_2 directly realized, or evoked, by the definite pronoun in U_{10}. Ignoring temporal and situational entities for the moment, the Cf list for U_{10} contains only this discourse entity. An utterance may also contain a *backward-looking center (Cb)*, linking it to the previous utterance. The Cb is obligatory or optional, depending on the version of centering. According to (Grosz et al. 1995), the Cb of an utterance U_{i+1} is "the most highly ranked element of Cfs(U_i) that is realized in U_{i+1}." Here, e_2, the only discourse entity evoked in U_{10}, is not in Cfs(U_9). Let us assume for purposes of illustration that the relevant previous utterance is not U_9 (cf. note 7), but rather an earlier utterance where e_2 is directly realized. On this assumption, e_2 is Cb(U_{10}). The more highly ranked a discourse entity is in Cfs(U_i), "the more likely it is to be Cb(U_{i+1})." Here I follow (Brennan et al. 1987) in referring to the most highly ranked member of the Cfs as the *preferred center (Cp)*. Thus Cp(U_{10}) is also e_2.

Given an utterance with known Cfs, Cp and Cb, centering provides hypotheses about how discourse entities in the subsequent utterance will be realized. For Fig. 13, for example, centering predicts that Cp(U_{10}) is the most likely candidate for Cb(U_{11}), thus constraining the interpretation of the pronoun subject in U_{11} to be e_2. In addition, (Grosz et al. 1995) claim that maintaining the same entity as Cb across adjacent utterances produces the most coherent discourse. They introduce three centering transitions to constrain the realization of the current Cb (cf. Fig. 15 and accompanying discussion below). In center continuation, claimed to be the most coherent of the three transitions, Cb(U_i), Cb(U_{i+1}) and Cp(U_{i+1}) are the same discourse entity. Prior to assigning an interpretation to the pronoun subject in U_{11}, let us represent the discourse entity it realizes as X. As in U_{10}, whatever discourse entity X gets bound to is the sole member of Cfs(U_{11}), hence is Cp(U_{11}). Cb(U_{11}), represented here as Y, is defined to be the highest ranked member of Cfs(U_{10}) that is realized in U_{11}, thus Cb(U_{11}) is e_2. By center continuation, the default binding for both X and Y is e_2, independent of other contextual factors. Since nothing in the contextual representation of e_2 or in the semantics of *take off* produces logical inconsistencies, the default binding predicted by centering leads to an acceptable interpretation of the pronoun.

We have now seen one example of how centering provides hypotheses for interpreting definite pronouns in an utterance U_{i+1}, given an interpretation of the preceding utterance U_i. In the case of understanding,

Rule 1:
If any element of $Cf(U_n)$ is realized by a pronoun in U_{n+1}, then the $Cb(U_{n+1})$ must be realized by a pronoun also.
$(\exists$ e (e $\in$ Cfs(U_n) $\wedge$ pronoun(realize_in(e,U_{n+1})))) $\leadsto$ pronoun(realize_in(Cb,U_{n+1}))
Contrapositive:
not(pronoun(realize_in(Cb,U_n))) $\rightarrow$ ($\forall$ e (e $\in$ Cfs(U_{n+1}) $\leadsto$ not(pronoun(realize_in(e,U_{n+1})))))

Figure 14: Rule 1

the surface form of U_{i+1} is a given. Now let us see how centering can be used to constrain the form of expressions in an utterance U_{i+1}, given an interpretation of U_i and a semantic representation of the content to be expressed in U_{i+1}. In order to use centering transitions to constrain surface form, we need an alternative definition of $Cb(U_{i+1})$, one framed in terms of its semantic and pragmatic relation to the prior utterance, rather than in terms of its realization in the utterance. A notable example of such an alternative is found in Dale (Dale 1992). For his domain of recipe generation, Dale (Dale 1992) defines the discourse center to be the entity that results from each next recipe operation that is described. After an utterance of *Stir the rice*$_1$, rice$_1$ is the discourse center. Dale tracks the discourse center across utterances, allowing the discourse center to be realized by a pronoun or to be elided. In addition, no other entities can be realized by a pronoun unless the discourse center is. This latter constraint implements a centering rule proposed in (Grosz et al. 1983), and included as Rule 1 in (Grosz et al. 1995). In Fig. 14, I show the (Grosz et al. 1995) version of Rule 1 along with my semiformal representation, including the contrapositive, which I will make use of later. (Note that *pronoun* in this and the other rules means *third person definite pronoun.*) Following (Grosz et al. 1995), I take all the centering rules to be defeasible inferences ($\leadsto$).

Centering Transitions:

Continuation:	$Cb(U_i) = Cb(U_{i-1})$	$Cb(U_i) = Cp(U_i)$
Retention:	$Cb(U_i) = Cb(U_{i-1})$	$Cb(U_i) \neq Cp(U_i)$
Shifting:	$Cb(U_i) \neq Cb(U_{i-1})$	$Cb(U_i) \neq Cp(U_i)$

Rule 2: Transition preferences (>)
$\{cont(U_{i-1}, U_i) \wedge cont(U_i, U_{i+1})\} > \{reten(U_{i-1}, U_i) \wedge reten(U_i, U_{i+1})\} > \{shift(U_{i-1}, U_i) \wedge shift(U_i, U_{i+1})\}$

Figure 15: Centering Transitions and Rule 2

I cannot yet give a precise, complete definition of Cb for use in generating narratives. Instead, I use the existing narratives to hypothesize what the Cb of each utterance is, based on the centering rules presented in (Grosz et al. 1995) and on the conventions of narrative discourse. The data presented in the evaluation section is a partial test of these hypotheses. Fig. 15 depicts the *continuation* transition mentioned above,

Rule 3:
$(\mathrm{Cb}(\mathrm{U}_i) = \mathrm{Cb}(\mathrm{U}_{i+1})) \leadsto \mathrm{pronoun}(\mathrm{realize\_in}(\mathrm{Cb},\mathrm{U}_{i+1})))$

Figure 16: Rule 3

along with two other transition types: *retention* and *shifting*. By Rule 2 from (Grosz et al. 1995), shown at the bottom of the figure, sequences of utterances that exhibit center continuation are preferred over those that exhibit retention, which are in turn preferred over those that exhibit shifting. Following Rule 2, I assume that a pair of utterances U_i and U_{i+1} within a discourse segment should exhibit continuation. I also assume that coherence within a narrative segment requires situations mentioned in adjacent utterances to be in a narrative relation (cf. (Lascarides & Oberlander 1992)). Here I present possible narrative relations in order to motivate a definition of Cb appropriate for generating narrative, rather than as a theory of narrative per se. I use the following four narrative relations, ordered from most specific to most general:

- main situation described in U_1 *causes* main situation described in U_{i+1}
- main situation described in U_1 *enables* main situation described in U_{i+1}
- main situation described in U_1 *occasions* main situation described in U_{i+1}
- main situation described in U_1 *precedes* main situation described in U_{i+1}

In intuitive terms, I take the Cb for generating narrative to be the most salient entity common to the main situations described in a pair of successive utterances that are in a narrative relation. Further, I assume that salience of entities in a situation is a partial order reflected to some degree in thematic structure (Jackendoff 1990). Thus agents and actors are relatively more salient; the time evoked by the main tense is relatively less salient. If the Cb of an utterance is selected during sentence planning, after the thematic structure has been determined, then thematic structure can be used to determine the Cb. If the Cb of an utterance is selected prior to sentence planning, other criteria must be used to evaluate relative salience of entities.[11]

A final centering rule made use of here is shown in Fig. 16 as Rule 3, predicting that when U_i and U_{i+1} have the same Cb (continuation or retention), the Cb is (defeasibly) realized by a third person definite pronoun. This rule was originally proposed in (Grosz et al. 1983), but does not appear in (Grosz et al. 1995). Instead, (Grosz et al. 1995) note that the Cb need not be realized as a pronoun, but that "it appears such uses are best when the full definite noun phrases that realize the centers do more than just refer." Handling non-referential effects is beyond the scope of this paper, so I retain Rule 3 from (Grosz et al. 1983).

I assume that the process of selecting and ordering the situations mentioned in successive utterances of a narrative takes place during content planning, and makes use of narrative relations such as those described above. The centering constraints take effect during sentence planning. Here, the situations described in U_{10} and U_{11} of Fig. 11 are that e_2 gets all situated again, and that e_2 takes off. Thus e_2 is hypothesized to be the Cb of U_{11} on the assumption that it is the most salient entity common to the situations described in U_{10} and U_{11}.

Taking e_2 to be $Cb(U_{11})$ for generating U_{11} has two consequences. First, by Rule 2, there is a defeasible preference to realize e_2 so as to satisfy the continuation transition. We already have that $Cb(U_{11}) = Cb(U_{10})$, so continuation will occur if $Cb(U_{11}) = Cp(U_{11})$. Much empirical work suggests that in English, the Cp of an utterance is often the grammatical subject (cf. citations in (Grosz et al. 1995); (Passonneau 1991). Second, by Rule 3, there is a preference for e_2 to be realized as a pronoun. The actual utterance satisfies both these constraints on surface form. In particular, it is conformance to Rule 3 that gives rise to the violation of informational adequacy. In the integrated model described in the following section, centering constraints on Cb realization are given priority over informational constraints for generating discourse anaphoric noun phrases.

Let us turn now to a centering account of the violation of economy that occurs in U_{10} of Fig. 12. Again, the first task is to determine $Cb(U_{10})$. (Grosz et al. 1983) (Grosz et al. 1995) claim that every utterance within a discourse segment has a Cb; whether a segment internal utterance optionally has a Cb is not specified. Both (Kameyama 1986) and (Passonneau 1993) propose that an utterance need not have a Cb, independently of whether it is segment initial. In (Passonneau 1995), I argued that U_9 and U_{10} in Fig. 12 are examples of utterances that have null Cbs because neither utterance contains a discourse anaphoric NP that realizes a discourse entity from the previous utterance in the segment. Here I take an alternative position that the Cb of U_9 is *realized* in the utterance, though indirectly. This follows from the hypothesis regarding the characterization of Cb—independent of its realization—in terms of narrative relations.[12]

U_9 in Fig. 12 occurs within the second empirically derived segment of the discourse. The first utterances in segment 2 describe the setting in which the man introduced in the preceding segment is picking pears. The narrative relation between U_9 and U_{10} is rather complex, and involves U_{11} as well. Later in the same segment, it is important to know where the man is located when the boy looks at him in order to infer that he is far enough away for the boy to be unobserved while stealing a basket of pears. The situations described in U_9 and U_{10} help set up the later description of the theft by specifying the relative positions of the man and the boy. The situation in U_9 "*occasions*" the situation that is described jointly in U_{10} and

U_{11}. Hobbs (Hobbs 1985) defines the occasion relation as one where the first argument describes a change of state, and the final state of which can be inferred from the second argument. By the same reasoning used to explicate the previous example, let us assume that Rule 2 applies, and that U_9 and U_{10} (as well as previous utterances within the segment) have the same Cb. The only discourse entity or discourse entities in common to both situations involve the space-time coordinates. Presumably, the time and location could be said to be *realized* by the utterances, given an appropriate theory of dynamic semantic interpretation (Grosz et al. 1995). For example, in the framework outlined in (Kameyama et al. 1993), both utterances would have the same *discourse reference interval*: an under-specified component of the logical forms of the utterances that is resolved incrementally against the context. Let us provisionally assume that it is the discourse reference interval, t_1, that is the Cb of all the utterances in segment 2 through U_{10}.

By Rule 3, there is a preference for t_1 (Cb(U_{10}) to be realized as a pronoun. However, in English, the discourse reference time is often not directly realized in an utterance. An adverb like "*now*" can potentially realize the discourse reference time, but typically not a third person definite pronoun, thus preventing the outcome predicted in Rule 3. Given that Cb(U_{10}) is not directly realized in the utterance, it is not Cp(U_{10}). As such, preference for center continuation cannot apply. However, the next most preferred transition—retention—can apply. Finally, the contrapositive of Rule 1 applies: since Cb(U_{10}) is not a pronoun the utterance must contain no pronouns. In the integrated model, centering contraints override the informational ones, thus achieving the effect of blocking the pronoun in examples like Fig. 12.

INTEGRATED MODEL

Ae-describe requires a discourse anaphoric noun phrase to be informationally adequate and economical with respect to the current focus space. It applies to all discourse anaphoric noun phrases independently of local (utterance-by-utterance) attentional constraints. Centering imposes certain constraints on how the Cb of the current utterance will be realized, and on whether any pronouns can occur in the current utterance, independently of any informational constraints that might apply. Because centering operates over a more local context than ae-describe, it applies first. Because centering applies first, centering constraints have priority in cases of conflict with informational constraints.

Conflicts between centering and ae-describe can arise in the two types of violations of informational constraints described above. An under-specified pronoun can occur when centering rule 3 applies, requiring the current Cb to be a definite pronoun regardless of informational constraints. An over-specified noun phrase can occur when rule 1 applies to an utterance where the Cb is not directly realized, or is not realized as a definite pronoun. I will illustrate how the integrated model operates in these two types of violations.

The Cfs of the prior utterance are omitted from the discussion of these examples because they play no role here. The input to the centering module consists of the discourse entity that is the Cb of the current utterance, along with the Cb of the prior utterance. The output is a set of constraints on how the Cb should be realized. For U_{11} in Figs. 11, we have these three equalities:

- $\mathrm{Cb}(U_{10}) = \mathrm{boy}(e_2)$
- $\mathrm{Cb}(U_{11}) = \mathrm{boy}(e_2)$
- $\mathrm{Cb}(U_{11}) = \mathrm{Cb}(U_{10})$

By Rule 3, we derive that if $\mathrm{Cb}(U_{11})$ should be realized as a pronoun. I assume that a separate sentence planning mechanism determines the argument structure of the sentence, and that e_2 is likely to be realized as a noun phrase argument of a verb of motion. The defeasible consequent of Rule 3 would consistent with this outcome of sentence planning.

The input to ae-describe for the same utterance consists of the set of discourse entities currently in focus intersected with the utterance context. The utterance context selects for a singular, animate entity, thus pruning all but the entities corresponding to each of the three boys from the context:

$$\begin{aligned} FS_7 &= \{\mathrm{boys}(e_1), \mathrm{boy}(e_2), \mathrm{bike}(e_{10}), \mathrm{boy}(e_3), \mathrm{boy}(e_4), \mathrm{pears}(e_{11}), \ldots\} \\ U_{11} &= \lambda\, X\ \mathrm{takes\_off}(X) \\ FS_7 \cap U_{11} &= \{\mathrm{boy}(e_2), \mathrm{boy}(e_3), \mathrm{boy}(e_4)\ \} \end{aligned}$$

The only pronoun that can felicitously be used to evoke e_2 ("*he*") would be ambiguous in this context, so ae-describe fails at the attempt to generate a pronoun. Similarly, it fails to generate a simple noun phrase using a lexical item denoting the basic type `boy` as the head. Ultimately, ae-describe requires that the noun phrase evoking e_2 express a focussed attribute that applies only to e_2. This conflicts with the centering constraint that the evoking noun phrase should be a definite pronoun, but the centering constraint has priority. Consequently, an under-specified pronoun will be generated.

Turning now to the example where economy is violated, the relevant input to centering will be the Cbs of U_8 and U_9. In the previous section, I argued the temporal entity t_1 is $\mathrm{Cb}(U_9)$. A similar argument applies for U_8 and U_7, giving the following equations as input to the centering module:

- $\mathrm{Cb}(U_8) = t_1$

- $\text{Cb}(\text{U}_9) = \text{t}_1$
- $\text{Cb}(\text{U}_9) = \text{Cb}(\text{U}_8)$

As in the preceding example, I rely on certain assumptions regarding output constraints on U_9. I assume that a separate sentence planning mechanism determines that the default expression for t_1 will be a grammatical tense or temporal adverbial rather than a discourse anaphoric NP. This assumption in combination with Rule 1 imposes the constraint that U_9 will not contain any definite pronouns, independent of informational constraints imposed by ae-describe.

For this example, the input to ae-describe consists of the discourse entities in focus in segment 3 intersected with the current utterance context:

FS_3	{people(e_{10}), bicycles(e_{11}), boy(e_2), bike(e_3), man(e_1), tree(e_4 ...}
U_{10}	λ X gets_off(X, bike(e_3)
$\text{FS}_3 \cap \text{U}_{10}$	{boy(e_2) }

Given the filtered context set consisting of $\{\text{e}_2\}$, ae-describe would succeed when it attempts to generate a pronoun that is a distinguishing description of e_2. However, the prior centering constraint requires that the utterance contain no definite pronouns. Consequently, the next preference class is chosen by ae-describe (MIN), leading to a minimal noun phrase whose lexical head denotes the basic type `boy`.

EVALUATION

Results

To evaluate the processing model presented here, I used a reserved set of narratives not included in the corpus analysis described in the section on *Corpus Analysis*. For each narrative, I abstracted away from the linguistic structure to create an input representation from which to compute output constraints on all the discourse anaphoric NPs. The target output is the observed distribution of definite pronouns (PRO), minimal NPs (MIN), and full NPs (FULL) with or without focussed attributes. A comparison of the output generated by the algorithm to the actual output gives a measure of the degree to which the algorithm conforms to observed human behavior. Three testing conditions are used. First, ae-describe is evaluated with and without a partitioning of the discourse context into empirically derived focus spaces. Then, using the focus spaces, the integrated model (ae-describe with centering) is evaluted.

Fig. 17 illustrates three utterances from one of the reserved narratives and the relevant data coding for the discourse anaphoric noun phrases that occur in the second two utterances.[13] To simplify the implementation and the pre-processing of the input, the utterance context is not used to filter the accessible discourse entities. Each narrative has on average 100 utterances, each of which would have required individual pre-processing to derive an appropriate semantic input. As shown in Fig. 17, the onset of a new segment is indicated by an attribute value pair indicating the new segment number (omitted under condition 1). The data for each utterance is represented as a 3-element list: the first item is the utterance number; the second item is an attribute value pair representing the discourse entity that is the current Cb; and the third item is a list representing the input for the discourse anaphoric NPs in the utterance. Utterance 8 has two discourse anaphoric NPs evoking the man and the ladder that were first mentioned in U_7. The input for a discourse anaphoric NP is a number uniquely indexing the evoked discourse entity, followed by a list of attribute value pairs describing the discourse entity. Discourse entity 1 has four attributes: one for the type, and three for the three types of focussed attributes including the location, lexical and event attributes. Event attributes were defined for the significant event of each of 7 episodes in the movie narrative, and the participants in these events (cf. section on *Focussed Attributes*). Discourse entity 2 has no focussed event attribute.

Seg	PP	U	
1	...	...	...
	7.1	7	And a man,
	7.2		is picking pears.
	8.1		In the trees.
	9.1		On a ladder.
2	10.1	8	And he (e_1) comes down,
	10.2		from the ladder (e_2),
	10.3	9	and he's (e_1) wearing an apron,

```
(SEG 2)
(8   (CB (ID 1))   (((ID 1)   ((TYPE man)
                              (LEX "man")
                              (LOCATION      ((LOC-TYPE in ) (ARG ((ID 3) (TYPE trees)))))
                              (EVENT         ((EVENT-TYPE pick) ...))))
                   ((ID 2)    ((TYPE ladder)
                              (LEX "ladder")
                              (LOCATION      ((ID 3) (TYPE trees)))))))
(9   (CB (ID 1))   (((ID 1)   ((TYPE man)
                              (LEX "man")
                              (LOCATION      ((ID 3) (TYPE trees)))
                              (EVENT         ((EVENT-TYPE pick) ...))))))
```

Figure 17: Data Coding of Input

Narrative	Total	ae-describe/-FS	ae-describe/+FS	Integrated
5	68	31 (46%)	51 (75%)	62 (91%)
11	124	77 (62%)	86 (70%)	112 (90%)
15	127	49 (39%)	81 (64%)	91 (72%)

Figure 18: Evaluation under Three Conditions

Table 18 shows the results: the number and percentage of correct outputs for each condition. Column 1 indicates the narrative and the column 2 the total number of discourse anaphoric NPs in the narrative. The remaining three columns are for the three testing conditions. The first testing condition establishes a baseline for evaluating the contribution of the global attentional state. Under this condition, the context argument is the set of discourse entities previously evoked in any utterance prior to the current one (column 3, ae-describe/-FS). Condition 2 (column 4, ae-describe/+FS) uses the current focus space as defined in preceding sections: the union of the set of discourse entities mentioned in the current segment (up to the current utterance) with the set of discourse entities in the most recent segment in which the relevant discourse entity was last evoked. Condition 3 uses the integrated model (centering with ae-describe; column 5, Integrated).

Table 18 shows wide variation across the three narratives, but a consistently strong increase in performance for conditions 2 and 3. The increase of column 4 over column 3 in Table 18 shows that the hypothesized focus spaces derived from the empirically derived discourse segments lead to a far greater proportion of correct outputs. Note that it is likely that an even greater percentage of correct outputs would occur under conditions 1 or 2 given an implementation that makes use of the utterance context to filter the current focus space. The increase of column 5 over column 4 shows that the integrated model produces an even greater improvement. The average improves from 49% correct in condition 1 to 84% correct for condition 3.

Discussion

The implementation results support the conclusions derived from the corpus analysis regarding informational adequacy and economy. Most discourse anaphoric NPs are adequate and economical relative to the current hypothesized focus space. The choice of attributes to express is also guided by the focus structure to the degree that focussed attributes are selected over other elements of a discourse entity's description. The implementation results also bear out the hypothesized role of centering in allowing violations of adequacy

Seg	PP	U	
7	...	...	...
	18.5	6	[pause] and [pause] one (e_3) of the guys,
	18.6	7	helps him (e_2) brush off the dust,
	18.7	8	[pause] and another guy (e_4) picks up the rock,
	18.8	9	and he (e_4) throws it out of the road,
	18.9	10	[pause] and [pause] $\boxed{\text{he } (e_2)}$ gets all situated again,
	18.10	11	and he (e_2) takes off,

Figure 19: Under-Specified Pronoun

and economy pertaining to the expression of the current Cb. Given the simplified implementation presented here, it is also likely that centering improves the results by compensating for not using the utterance to filter the context used by ae-describe. In the integrated model, ae-describe and centering provide sometimes complementary, sometimes overlapping, and sometimes conflicting constraints. For example, only ae-describe provides constraints for discourse anaphoric noun phrases other than the current Cb, whereas only centering provides constraints pertaining to grammatical role. In many cases, the form constraints provided by centering and ae-describe for the current Cb are compatible, but in some cases they conflict, in which case centering takes priority.

Careful readers might recall an apparent counterexample to the model argued for here. In the *Background*, I discussed a case where a pronoun subject of U_{10} (boxed NP) in Fig. 19 does not corefer with the subject pronoun of the preceding utterance. This is an apparent counter example to the intuition that centering should block such pronouns in order to avoid the garden path effect that seems to occur with non-coreferential definite pronouns in grammatical subject of adjacent utterances (cf. the discussions of Fig. 11-12 above). However, there are certain key differences between the non-coreferential subject pronoun that occurs in Fig. 19 and the avoidance of this configuration in examples like Fig. 12.

Utterance 10 in Fig. 19 differs from utterance 10 in Fig. 12 with respect to the inferred local and global attentional state, and certain prosodic cues that may correlate with these differences. One difference is that the entity evoked by the pronominal subject in Fig. 19 is presumably the current Cb, whereas the entity evoked by the full NP subject in Fig. 12 is not the Cb of its utterance. The evoked entity in Fig. 19 (e_2; corresponding to the boy who fell off the bike) has been in focus throughout the empirically derived discourse segment; the segment describes the boy's accident and the various ways in which the other boys assist him. In contrast, the evoked entity in Fig. 12 (e_2; same character) has just been introduced into the discourse two utterances prior. As discussed above, the segment is about the setting, and the relation of e_2

to another character in the setting. Second, the utterance in Fig. 19 begins with a long pause, followed by a cue word, followed by another long pause. The long pauses in Fig. 19 may signal a discontinuity with the preceding utterance that helps the addressee (and reader) infer a less cohesive relationship with the preceding utterance. Finally, U_{10} in Fig. 19 seems to occur at a discourse pop, a resumption of a superordinate or previously suspended focus space. Empirical work on the relation of global discourse structure to linguistic cues (Passonneau & Litman 1993) (Litman & Passonneau 1995*a*) (Grosz & Hirschberg 1992) (Hirschberg & Grosz 1992) (C.H.Nakatani et al. 1995) presents evidence that pauses, prosodic contours, cue phrases and discourse anaphoric noun phrases do correlate with segment boundaries, though less frequently and in a more complex way than predicted in previous literature. Although neither of the two algorithms from (Litman & Passonneau 1995*a*) for locating segment boundaries locates one here, one of the seven subjects who segmented this narrative did locate a boundary here.

None of the differences between Fig. 19 and Fig. 12 cited above is definitive, but together they suggest a refinement in the hypothesis regarding the relation between ae-describe and centering proposed here. In the model I have presented, I claim that during the generation process, centering constraints have priority over adequacy and economy. It is likely that speakers are free to rely on an alternative prioritization, but only when other factors guide the reasoning process, such as those discussed above in regard to Fig. 19.

CONCLUSION

I have presented an analysis of discourse anaphoric NPs in a corpus of narrative monologues showing that pronouns and phrasal NPs are rarely over-specified. Future research should indicate to what degree this generalization applies to other genres and modalities. Previous work on centering predicts conditions under which an under-specified pronoun can be used, but says little about how centering fits into global discourse processing. I have outlined a model that integrates the attentional constraints of centering with a module that employs a hypothesized global focus space to enforce two informational constraints, adequacy and economy. This model predicts that informational constraints must also be relaxed, given certain centering conditions, in order to allow over-specified noun phrases. In the integrated model, centering constraints have priority over informational constraints during the generation process.

The understanding process involves a distinct relationship between centering and informational constraints. The fact that a sequence of utterances with non-coreferential definite pronoun subjects (e.g., (6c) and Fig. 19) produces a garden path effect suggests that centering constraints apply first during understanding, but can be overridden. If they apply first during understanding, that would explain the experimental results that subjects assign an interpretation to pronominal subjects almost immediately (Hudson-D'Zmura 1988). But the ultimate interpretation assigned to U_{10} in Fig. 19 (cf. also (6) argues that in case of conflict during understanding, commonsense reasoning has priority. This accords with an argument made in (Kameyama et al. 1993). There we also explored the interaction between centering and Gricean reasoning, proposing a model of temporal centering for constraining the interpretation of simple past and past perfect sentences. We argued that the effect of a structurally superordinate temporal operator (e.g., past is superordinate to perfect in the past perfect) is analogous to the effects of more prominent grammatical role in centering (e.g., subject vs more oblique roles). We captured the effects of linguistic structure on local attentional state by defeasible inference rules, but argued that commonsense reasoning could always override the more general inferences from grammatical structure.

Centering provides one mechanism for relaxing the requirement that an NP (either pronominal or phrasal) should be informationally adequate and economical. Another mechanism would be needed to relax informational constraints at shifts in focus structure, so as to account for the weak tendency of over-specified NPs to occur at global shifts of attention (Table 1). However, further investigation is needed to determine how to integrate local and global discourse processing. Some of the examples presented here suggest that the contextual dependencies captured by the use of focused attributes might constrain the relation of each new utterance to the global discourse model. For example, the segment onset in Fig. 8 (U_{108}) contains two NPs, one of which is presumably the same as the Cb of the preceding utterance. Maintaining the same Cb relates U_{108} and its focus space (FS_{22}) to the most recent focus space FS_{21}. But the object NP expresses attributes last mentioned in segment 17, thus relating U_{108} to the earlier focus space FS_{17}. If the global structure is a tree, the relation of U_{107} to both segments 21 and 17 might indicate how high up in the tree to locate the new focus space. Alternatively, an investigation of such relations might provide evidence about the nature of global structure.

Notes

[1] Zero pronouns were identified by hand-checking the output of an automatic tool. The narratives were first tagged for part of speech. The tool proposed zero pronoun whenever a tensed verb was found following a conjunction or comma.

[2]All 20 narratives from (Chafe 1980) were segmented, using a total of 140 subjects.

[3]In the basketball reports analyzed in (Robin 1994) over-determined NPs occur frequently.

[4]Cf. (Dale & Haddock 1991) for an extension of Dale's (Dale 1989) algorithm to handle n-ary predicates.

[5]Di Eugenio (Eugenio 1990) discusses another interaction between centering and avoidance of garden path effects pertaining to pro-drop in Italian. In Italian, gender marking on the verb disambiguates the subject. Di Eugenio argues that because gender marking occurs early enough in the utterance to prevent a garden path effect, null pronouns can be used where centering would otherwise predict a stressed pronoun.

[6]The author has recordings of the narratives.

[7]In (Grosz et al. 1983), centering is restricted to utterances that are adjacent within a discourse segment, but, as noted in (Passonneau To appear), the most coherent intra-segment centering transitions seem to apply at some segment boundaries. Joint work by the author with Diane Litman (Passonneau & Litman To appear*a*) indicates that segment boundaries are not signalled consistently either by the same speaker or across speakers. Modifications to the present model pertaining to segment boundaries await further study.

[8]Technical reasons that are not relevant here motivate a distinction between a speaker's intended referent and the discourse entity evoked by the speaker's use of a referential expression in (Passonneau To appear).

[9]Cf. (Kameyama et al. 1993) for applications of centering to temporal or situational entities.

[10]Various criteria have been proposed for computing this partial order, including properties pertaining to the grammatical role or surface form of the realizing expression, or to parallelism of such features across adjacent utterances.

[11]Generation systems typically distinguish between a content selection phase and a process of sentence planning for determining lexical items, linear order, grammatical relations and so on (Robin 1994). The

implementation described in the evaluation section does not include full sentence planning.

[12]With respect to predicting the over-specified NP, my two approaches amount to notational variants.

[13]The input data has been simplified somewhat for expository purposes.

References

Allen, J. (1987), *Natural Language Understanding*, The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Brennan, S. E., Friedman, M. W. & Pollard, C. J. (1987), A centering approach to pronouns, *in* 'Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics', Stanford, CA, pp. 155–162.

Chafe, W. L. (1980), *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*, Ablex Publishing Corporation, Norwood, NJ.

C.H.Nakatani, Hirschberg, J. & Grosz, B. J. (1995), Discourse structure in spoken language: Studies on speech corpora, *in* 'AAAI 1995 Spring Symposium Series: Empirical Methods in Dicourse Interpretation and Generation', pp. 106–112.

Cochran, W. G. (1950), 'The comparison of percentages in matched samples', *Biometrika* **37**, 256–266.

Dale, R. (1989), Cooking up referring expressions, *in* 'Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics', pp. 68–75.

Dale, R. (1992), *Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes*, MIT Press, Cambridge, MA.

Dale, R. & Haddock, N. (1991), Generating of referring expressions involving relations, *in* 'Proceedings of the European ACL'.

Dale, R. & Reiter, E. (1994), 'Computational interpretations of the Gricean maxims in the generation of referring expressions'. Forthcoming.

Eugenio, B. D. (1990), Centering theory and the italian pronominal system, *in* 'Proceedings of the 13th International Conference on Computational Linguistics', Helsinki.

Grice, H. P. (1975), Logic and conversation, *in* 'Syntax and Semantics, Vol. 3: Speech Acts', Academic Press, New York, pp. 41–58.

Grosz, B. & Hirschberg, J. (1992), Some intonational characteristics of discourse structure, *in* 'Proc. of the International Conference on Spoken Language Processing (ICSLP)'.

Grosz, B. J. (1977), The Representation and Use of Focus in Dialogue Understanding, PhD thesis, University of California, Berkeley.

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, Massachusetts, pp. 44–50.

Grosz, B. J., Joshi, A. K. & Weinstein, S. (1995), 'Centering: A framework for modeling the local coherence of discourse', *Computational Linguistics* **21.2**, 203–226.

Grosz, B. J. & Sidner, C. L. (1986), 'Attention, intentions and the structure of discourse', *Computational Linguistics* **12**, 175–204.

Hearst, M. A. (1994), Multi-paragraph segmentation of expository text, *in* 'Proceedings of the 32nd Annual Meeting of the ACL', Las Cruces, NM, pp. 9–16.

Heim, I. (1983), File change semantics and the familiarity theory of definiteness, *in* Bauerle, Schwarze & von Stechow, eds, 'Meaning, Use and Interpretation of Language', Walter de Gruyter, Berlin.

Hirschberg, J. & Grosz, B. (1992), Intonational features of local and global discourse structure, *in* 'Proceedings of the Darpa Workshop on Spoken Language'.

Hobbs, J. R. (1985), On the coherence and structure of discourse, Technical Report CSLI-85-37, Center for the Study of Language and Information.

Hobbs, J. R., Stickel, M., Martin, P. & Edwards, D. (1988), Interpretation as abduction, *in* 'Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics', Buffalo, NY, pp. 95–103.

Hudson-D'Zmura, S. B. (1988), The Structure of Discourse and Anaphor Resolution: The Discourse Center and the Roles of Nouns and Pronouns, PhD thesis, University of Rochester.

Jackendoff, R. (1990), *Semantic Structures*, MIT Press, Cambridge, MA.

Joshi, A. & Weinstein, S. (1981), Control of inference: Role of some aspects of discourse structure—centering, *in* 'Proceedings of the International Joint Conference on Artificial Intelligence', Vancouver, B.C., pp. 385–387.

Kameyama, M. (1985), Zero Anaphora: The Case of Japanese, PhD thesis, Stanford University.

Kameyama, M. (1986), A property-sharing constraint in centering, *in* 'Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics', New York, pp. 200–206.

Kameyama, M., Passonneau, R. & Poesio, M. (1993), Temporal centering, *in* 'Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics', Columbus, OH, pp. 70–77.

Kamp, H. (1981), A theory of truth and semantic representation, *in* J. Groenendijk, T. Janssen & M. Stokhof, eds, 'Formal Methods in the Study of Language, Part I', Mathematisch Centrum, Amsterdam, pp. 277–322.

Karttunen, L. (1976), Discourse referents, *in* J. McCawley, ed., 'Syntax and Semantics, Vol. 7: Notes from the Linguistic Underground', Academic Press, New York.

Kronfeld, A. (1986), Donnellan's distinction and a computational model of reference, *in* 'Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics', New York, pp. 186–191.

Lascarides, A. & Oberlander, J. (1992), 'Temporal coherence and defeasible knowledge', *Theoretical Linguistics* . To appear.

Litman, D. & Passonneau, R. (1995*a*), Combining multiple knowledge sources for discourse segmentation, *in* 'Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics', Cambridge, MA, pp. 108–115.

Litman, D. & Passonneau, R. (1995*b*), Developing algorithms for discourse segmentation, *in* 'Working Notes of AAAI Spring Symposium Series on Empirical Methods in Discourse Interpretation and Generation', Palo Alto, CA, pp. 85–92.

Marslen-Wilson, W., Levy, E. & Tyler, L. K. (1982), Producing interpretable discourse: The establishment and maintenance of reference, *in* R. J. Jarvella & W. Klein, eds, 'Speech, Place and Action', John Wiley and Sons Ltd., New York, pp. 339–378.

Moore, J. D. & Pollack, M. E. (1992), 'A problem for RST: The need for multi-level discourse analysis', *Computational Linguistics* **18**, 537–544.

Morris, J. & Hirst, G. (1991), 'Lexical cohesion computed by thesaural relations as an indicator of the structure of text', *Computational Linguistics* **17**, 21–48.

Passonneau, R. J. (1991), Some facts about centers, indexicals and demonstratives, *in* 'Proceedings of the 29th Annual Meeting of the ACL', Berkeley, pp. 63–70.

Passonneau, R. J. (1993), Getting and keeping the center of attention, *in* R. Weischedel & M. Bates, eds, 'Challenges in Natural Language Processing', Cambridge University Press. Also available as Tech. Report CUCS-060-90, Dept. of Computer Science, Columbia University.

Passonneau, R. J. (1994*a*), Frame shifts: Perlocutionary meaning and discourse reference, *in* 'Conference on Focus and Natural Language Processing', Kassel, Germany.

Passonneau, R. J. (1994*b*), Protocol for coding discourse referential noun phrases and their antecedents, Technical report, Columbia University.

Passonneau, R. J. (1995), Integrating Gricean and attentional constraints, *in* '14th International Joint Conference on Artificial Intelligence', Montreal, pp. 1267–1273.

Passonneau, R. J. (To appear), Interaction of the segmental structure of discourse with explicitness of discourse anaphora, *in* E. Prince, A. Joshi & M. Walker, eds, 'Proceedings of the Workshop on Centering Theory in Naturally Occurring Discourse', Oxford University Press.

Passonneau, R. J. & Litman, D. (1993), Intention-based segmentation: Reliability and correlation with linguistic cues, *in* 'Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics', Columbus, OH, pp. 148–155.

Passonneau, R. J. & Litman, D. (To appear*a*), 'Discourse segmentation by human and automated means'. Submitted for publication.

Passonneau, R. J. & Litman, D. (To appear*b*), Empirical analysis of three dimensions of spoken discourse: Segmentation, coherence and linguistic devices, *in* E. Hovy & D. Scott, eds, 'Interdisciplinary Perspectives on Discourse', Springer Verlag, Berlin.

Polanyi, L. (1988), 'A formal model of discourse structure', *Journal of Pragmatics* pp. 601–638.

Prince, E., Joshi, A. & Walker, M., eds (To appear), *Proceedings of the Workshop on Centering Theory in Naturally Occurring Discourse*, Oxford University Press.

Reichman, R. (1985), *Getting Computers to Talk Like You and Me*, MIT Press, Cambridge, Massachusetts.

Reiter, E. (1990*a*), The computational complexity of avoiding conversational implicatures, *in* 'Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics', pp. 97–104.

Reiter, E. (1990*b*), Generating appropriate natural language object descriptions, PhD thesis, Harvard University.

Robin, J. (1994), Revision-based generation of natural language summaries providing historical background: Corpus-based analysis, design, implementation and evaluation, Technical Report CU-CS-034-94, Computer Science Department, Columbia Universtity, New York, NY. PhD. Thesis.

Walker, M. A. (1995), 'Limited attention and discourse structure'. Ms.

Walker, M. A., Iida, M. & Cote, S. (1994), 'Japanese discourse and the process of centering', *Computational Linguistics* **20**, 194–231.

Webber, B. L. (1978), A formal approach to discourse anaphora, Technical Report 3761, Bolt Beranek and Newman Inc.

Webber, B. L. (1988), 'Tense as discourse anaphor', *Computational Linguistics* **14**, 113–122.

Webber, B. L. (1991), 'Structure and ostension in the interpretation of discourse deixis', *Language and Cognitive Processes* pp. 107–135.