Interaction of Discourse Structure with Explicitness of Discourse Anaphoric Noun Phrases

Rebecca J. Passonneau
Bellcore
445 South Street
Morristown, NJ
and
Columbia University
New York, NY 10027
beck@bellcore.com

Abstract

Centering is a processing model that relates the local utterance-by-utterance context, inference, and discourse anaphoric reference. As such, it potentially serves as a point of departure for a general theory of language use. However, its predictive force has been restricted to a small class of speech act constituents: unstressed or reduced forms of third person definite pronouns. One question for extending centering is, how does local utterance processing relate to the global discourse context? Another open question is, how does centering interact with the global context in constraining the surface form of referring expressions other than pronouns. These are empirical questions. I present the results of two empirical studies directed at these larger questions.

The studies presented here make use of a corpus of spoken narratives that have been annotated with multi-utterance segments hypothesized to be constituent units of global structure. First I look at the correlation of centering transitions with the empirically derived segments. The correlation is poor, suggesting that centering transitions do not directly reflect segmental structure. Then I examine the informativeness of discourse anaphoric noun phrases relative to the same segmentation data. The results suggest that the global informational context corresponding to the discourse segment in which utterances occur should be factored into predictions about centering transitions. I discuss how centering could be integrated with global processing to provide a more uniform model of discourse anaphoric reference. I conclude by hypothesizing how centering and informational constraints relate to inferences about relations among segments.

1 Introduction

The centering principle formulated in Grosz et al. (Grosz, Joshi, & Weinstein 1983) has the potential to serve as a point of departure for a general theory of language use. However, its predictive force is restricted to a very small class of speech act constituents, namely unstressed or reduced forms of third person, definite pronouns (cf. (Kameyama 1985)). Centering models how the local context of the previous and current utterances constrains the surface form used to evoke the most salient discourse entity of the current utterance. One question for extending centering is, how does local utterance processing relate to global discourse context? Another open question is, how does the role of global context in constraining the form of referring expressions interact with centering? Relatively few claims have been offered regarding these questions, largely because of a lack of knowledge about how observable factors like the form of referring expressions in a discourse interact with theoretically posited constructs like centering and global structure.

Here I approach the problem of extending centering by viewing it as a constraint on the intersection of three factors: pragmatic constraints on the form of noun phrases (NPs), cohesive relations among adjacent utterances, and the global structure of discourse. Section 2 is a brief review of claims regarding the relation of centering to global discourse structure, and to the surface form of NPs. In section 3, I describe the data and in section 4 I present two corpus analyses. First I look at the correlation of centering transitions to empirically derived segmentations of five narratives in the corpus. The correlation is poor, suggesting that centering does not directly reflect global structure. Then I examine the relative informativeness of NPs in ten narratives of the corpus, arguing that the contrast between definite pronouns and definite NPs is relative to context. In section 5, I briefly summarize a generation model (described more fully in (Passonneau Forthcoming)) designed to account for the observations from section 4. The model enforces informational *adequacy* (Dale 1992) and *economy* (cf. related notions of efficiency (Dale 1992) and local brevity (Reiter 1990)). The use of centering in the integrated model makes it possible to relax informational constraints in key contexts. In the conclusion, I discuss the significance of the integrated model for addressing the interaction between local and global discourse processing.

2 Background

2.1 Centering

Centering is a model of local attentional state that constrains the surface form of certain noun phrases, and that potentially links the current utterance to the preceding one. The objects that centering operates on are discourse entities. Briefly, a discourse entity is a representation of an object in the discourse model whose attributes are derived from the semantic structure of co-indexed NPs and of the utterances in which they occur, as discussed in Webber (Webber 1978). I will henceforth assume the reader to be familiar with the notions of discourse entity and discourse model. As discussed in the introductory chapter to this volume, one of the discourse entities evoked by an NP in an utterance U_i may be the *backward-looking center* (Cb) of U_i (Grosz, Joshi, & Weinstein 1983), representing the current local focus of attention. Alternatively, the Cb of U_i (Cb_{U_i}) might not be explicitly mentioned (realized) in the utterance. The discourse entities mentioned in U_i are ordered, e.g., by increasing obliqueness of grammatical role (Kameyama 1985) (Brennan, Friedman, & Pollard 1987), so as to represent the likelihood that they will be mentioned in the subsequent utterance. This ordered set of discourse entities in U_i constitutes the *forward-looking centers* (Cfs) of U_i.

(1) a. $Patty_i$ took $Carmella_j$ to the bookstore.
b. Afterwards, $Carmella_j$ gave her sister $Rachel_k$ a new book.
c. $She_?$'s a true bibliophile.

Grosz et. al (Grosz, Joshi, & Weinstein 1986) observe that whether the Cb is maintained or shifted across adjacent utterances contributes to the relative coherence of the local context. They introduce three centering transitions, ordered by the degree of coherence, whose primary function is to constrain the identity of the current center. Center continuation is claimed to be the most coherent: the Cb of U_i and U_{i-1} are the same, and the Cb of U_i is also the highest ranked member of the Cfs of U_i. I follow (Brennan, Friedman, & Pollard 1987) in refering to the highest ranked member of Cfs as the *preferred center* (Cp). Center retention differs from continuation in that Cb_{U_i} and Cp_{U_i} are not the same. Shifting, said to be the least coherent transition, is where the current Cb differs both from the previous Cb and the current Cp. Alternative interpretations of an ambiguous definite pronoun are ordered in part by whether the resulting utterance transition is continue, retain or shift.

Example (1) illustrates that where semantic selectional constraints and commonsense reasoning do not distinguish between possible referents for an ambiguous pronoun, there is an independent effect of local attentional constraints on the use of definite pronouns. Given a continue transition from (1b) to (1c), "*she*" refers to Carmella; given a shift, "*she*" refers to Rachel (or to someone else). The preference ordering of transition types accords with the intuition that the preferred interpretation of the pronoun in example (1c) is Carmella.

2.2 Global Discourse Structure

(2) a. And he he comes back,
b. and gives him his hat,
c. and so he gives him some pears,
d. and so they walk off, eating these pears.
e. And then they pass by where the farm laborer is,

f. and while $\boxed{\text{this}}$ is going on,
g. the farm laborer realizes that a whole basket's missing.

A variety of phenomena pertaining to discourse reference illustrate the need to posit global discourse structure. For example, Webber (Webber 1991) accounts for what she refers to as discourse deixis in terms of global structure. She looks at examples like this one, taken from the corpus of narrative monologues analyzed in section 3. The demonstrative pronoun "*this*" in (2f) refers to a complex event derived from the sequence of utterances (2a) through (2c). Following Linde (Linde 1979) and Polanyi (Polanyi 1988), Webber (Webber 1991) takes discourse structure to be a tree, and argues that only units on the right frontier of the tree can support the inferences needed for discourse deixis.

(3) a. Wanda (w), June (j) and Sally (s) all went to school together.
b. They (w, j, s) were and have remained good friends.

c. In fact, June is now Wanda's accountant.
d. Last week, they (w, j) got started on Mary's taxes.
e. They'll need to meet frequently for the next month.

f. Sally works in the same huge firm as Wanda,
g. but she volunteers at a reading clinic every evening with June.
h. They (s, j) enjoy discussing their teaching methods.

i. Lately, Sally and Wanda have both had to work late.
j. They (s, w) have gotten to know the night watchman well.
k. They like his friendly manner.

m. Anyway, $\boxed{\text{the women ((s, w) } \| \text{ (w, j, s))}}$ are glad to see one another so often.

Definite pronoun interpretation can also depend on global structure by constraining accessibility to previously evoked entities. Grosz (Grosz 1977) gives examples of definite pronouns whose antecedents are distant in terms of utterances, but not in terms of the hypothesized global structure. The constructed discourse in (3) illustrates another type of example. The discourse is about Wanda, June and Sally (w, j, s), and every subset of two or three members is evoked by a definite plural pronoun somewhere in the discourse. For visual clarity, the discourse is divided into the utterance sequences that focus on the various subsets of Wanda, June and Sally. The global focus of attention apparently constrains the accessibility of discourse entities for inferring the set referent of "*the women*" in (3.m). As indicated, the sets (s, w) and (w, j, s) seem to be possible referents, but not the other two subsets of two women, even though the sentence in

(3.m) refers to the frequency of meeting, and it is only June and Wanda who have been described as meeting frequently. The discourse tree model of attentional state would predict that entities mentioned in (3a)-(3b) and (3i)-(3k) are on the 'open' right frontier, hence more accessible than entities in 'closed' parts of the tree (Polanyi 1988).

The relation of centering to global discourse structure is unspecified. In part, this reflects indeterminacy about global structure, as noted in the next paragraph. Let us first turn briefly to Grosz and Sidner's (Grosz & Sidner 1986) model of discourse structure and what they say about its relation to centering. Grosz and Sidner (Grosz & Sidner 1986) posit a tri-partite structure of discourse arising from relations among discourse intentions (their intentional structure), which are paralleled by relations among focus spaces (their attentional state), which are in turn paralleled by relations among discourse segments (their linguistic structure).[1] They argue that utterances group together into segments, and that each segment corresponds to a unit of intentional structure they refer to as a Discourse Segment Purpose (DSP). Associated with the segment is a focus space containing representations of the current DSP and the discourse entities that are in global focus. Grosz and Sidner say little about whether global structures can be computed from centering, or vice versa. They (Grosz & Sidner 1986) are also careful to distinguish between the local phenomenon of the Cb of an utterance, and the global focus for the segment that the utterance is a part of. Thus they (Grosz & Sidner 1986) note that the Cb, but not the segment intention, can shift during a segment.

Theories of global discourse structure typically address the question of how (the hearer's interpretation of) a new utterance increments (the speaker's and hearer's beliefs about) the model of the discourse. A variety of views are represented in (Linde 1979) (Polanyi 1988) (Webber 1988) (Webber 1991) (Mann, Matthiessen, & Thompson 1992) (Hobbs 1985) (Lascarides & Oberlander 1992), to mention just a few works, depending on whether utterances are viewed as actions or as propositions, and on the nature of the relation between the surface structure of the utterance and the abstract model representing the hearer's understanding of the discourse. Moore and Pollack (Moore & Pollack 1992) distinguish between intentional relations and informational relations, such as elaboration, causation, evidence, and so on, discussed in (Polanyi 1988) (Mann, Matthiessen, & Thompson 1992) (Hobbs 1985). Moore and Pollack take issue with Grosz and Sidner's (Grosz & Sidner 1986) claim that dominance and satisfactions-precedence relations among DSPs are sufficient for capturing the hierarchical structuring of discourse arising from relations among utterances, to the exclusion of such substantive relations. They present a sample discourse whose interpretation gives rise to conflicting trees of intentional and informational structure. Given Grosz and Sidner's (Grosz & Sidner 1986) hypothesis that relations among focus spaces are determined by the discourse tree structure, Moore and Pollack's argument raises the question of whether the tree structure of informational relations also constrains accessibility of focus spaces , and how one might empirically address questions about the interaction of discourse structure

[1]As in (Passonneau 1993) and elsewhere, I refer to Grosz & Sidner's (Grosz & Sidner 1986) linguistic structure as segmental structure in order to avoid confusion with lexico-grammatical and prosodic structure.

and attentional state.

2.3 Surface Form of NPs

Grosz et al. (Grosz, Joshi, & Weinstein 1983) claim that the local context represented in centering has a greater effect on definite pronouns than on definite phrasal NPs, while the reverse is true with respect to the global context. Grosz and Sidner (Grosz & Sidner 1986) make a similar claim that there are different constraints on definite pronouns and reduced definite NPs within versus across segments. For example, the say that because *pronouns contain less explicit information than definite descriptions* mechanisms like centering *are needed to account for what may and may not be pronominalized.* Dale (Dale 1992) argues that the evidence they (Grosz & Sidner 1986) present for these claims is weak, except in the case of discourse pops. A pop occurs at an utterance following an embedded segment, thus "popping" back up to the embedding segment (Grosz 1977). Dale (Dale 1992) concurs that if occurrence of a discourse pop has been clearly signaled, then a pronoun in the utterance resuming the embedding segment will unambiguously re-evoke an entity in global focus.

(4) a. Patty$_i$ and Carmella$_j$ went to the bookstore.
b. Afterwards, Carmella$_j$ saw her sister Rachel$_k$.
c. She$_k$ looked pale.

Another factor influencing the use of definite pronouns versus definite noun phrases is that the relative inexplicitness of a pronoun can be compensated for in cases where the semantics of the utterance context is sufficiently informative. According to Fillmore (Fillmore 1971), verbs denoting perceptual events are deictic: the argument structure of the verb corresponds to a particular point of view on the event. Kameyama (Kameyama 1986) shows that a pronoun in a discourse describing a sequence of perceptual events is not ambiguous in contrast to a pronoun in a discourse with parallel sentence structure describing events with no deictic bias. Thus, in (4), which is parallel to (1) above, the subject pronoun in the last sentence is taken to corefer with the preceding object, rather than the preceding subject, in contrast to (1). Furthermore, alternate readings seem less available than in (1).

3 The Corpus and Coding Features

The question of how the surface form of referring expressions, centering, and global discourse structure interact is an empirical one. It is also very complex, and can only be investigated incrementally, assembling pieces of evidence from which to construct new questions. In this section, I describe two companion analyses of a corpus of spoken narratives directed at this general question. As described in the next subsection, the corpus had already been coded for a number of contextual features for purposes of another study (Passonneau & Litman 1993). The coding features included properties of noun phrases and utterances that, with certain additions described below, can be used to compute centering data structures. In addition, the corpus was coded for aspects of global structure derived from an empirical study of discourse segmentation. Using these two classes of features as input, I first examine how centering transitions between adjacent utterances correlate with the empirically derived discourse segments (§4.1). Then in section 4.2, I summarize results of another study in which I investigated the correlation between the discourse segmentation in the corpus and relative informativeness of NPs.

The corpus consists of ten randomly selected narrations from Chafe's Pear stories (Chafe 1980). Chafe's corpus was created by recording subjects who were asked to view a movie and describe it to a second person. The movie, specially constructed for this purpose, contained seven sequential episodes about a man picking pears. It had a vivid sound track, but no language. Chafe's (Chafe 1980) transcriptions are described briefly below.

3.1 NP Coding and Centering

As part of a general study of the relationship between reference and discourse structure (Passonneau & Litman 1993) (Litman & Passonneau 1995a) (Passonneau & Litman 1996), the corpus was coded for features pertaining to the surface form of noun phrases (NPs), their referential value, and their location in the discourse. The coding features are documented in (Passonneau 1994). Two coders coded each narration. Here, I briefly describe the NP coding features. I also describe how they were used to compute centering data structures for five of the narratives.

The ten narratives contained 1,814 NPs (including zero pronouns; cf. (Passonneau 1994)) that strictly co-referred with a preceding NP. Their referents are semantically heterogeneous, including properties, events, and abstract entities, as well as concrete individuals. A data structure for each NP records its surface form, its referential index, and its location. An NP's referential index is unique if the NP does not corefer with any other NPs; otherwise it is assigned the same index as the NPs it corefers with. Coding of the NP's location is described in the next paragraph. NPs in five narratives were also coded for grammatical role.

One element of the location coding is the sequential position of the NP in the discourse. The second is the sequential position of the containing utterance. Identification of utterance units involves syntactic, prosodic and performance factors. Chafe's (Chafe 1980) transcriptions distinguish three types of prosodic phrases that were identified using graphic displays of intonation contours extracted from the acoustic signal. In Chafe's transcriptions, a period indicates a phrase terminated by a pitch fall, a question mark indicates a phrase terminated by a pitch rise, and a comma indicates phrasal units with neither.[2] The syntactic clause unit used here was roughly any tensed clause that was not a verb argument, not a restrictive relative clause, and not one of a small set of

[2]Cf. Figs. 1-3, 5-7.

highly formulaic clauses that I refer to as interjection clauses (e.g., "*You know*" with no clausal argument; for full details cf. (Passonneau 1994)). Here, utterances are defined to be a non-overlapping sequence of units that completely covers the discourse. Briefly, a new utterance begins with a new syntactic clause if it is immediately adjacent to the preceding clause. Otherwise it begins at the onset of the prosodic phrase where the next clause begins. There can be portions of an utterance that intervene between syntactic units within a prosodic phrase, such as sentence or word fragments, and non-lexical articulations. Utterance coding is illustrated in Fig. 1, discussed below.

Two versions of centering were computed; A as in (Brennan, Friedman, & Pollard 1987) and B as in (Kameyama, Passonneau, & Poesio 1993). In both versions, the Cf list for an utterance U_i (Cf_{U_i}) contains the discourse entities evoked in U_i. They are ordered by surface form of the associated NP, with definite pronouns ranked higher than phrasal NPs, and by increasing obliqueness of grammatical role. Here, Cb_{U_i} is treated as obligatory for Version A.[3] It is the highest ranked member of Cf_{U_i} that also appears in $Cf_{U_{i-1}}$, else the highest ranked member of Cf_{U_i}. The preferred center (Cp) is the highest ranked member of Cf_{U_i}. The transition types are:

1.1 CON continue: $Cb_{U_i} = Cb_{U_{i-1}} = Cp_{U_i}$

1.2 RET retain: $Cb_{U_i} = Cb_{U_{i-1}} \neq Cp_{U_i}$

1.3 SHIFT1 shift-1: $Cb_{U_i} \neq Cb_{U_{i-1}} = Cp_{U_i}$

1.4 SHIFT2 shift: $Cb_{U_i} \neq Cb_{U_{i-1}} \neq Cp_{U_i}$

In Version B, Cb_{U_i} must have been realized as a pronoun, and is the highest ranked referent in Cf_{U_i} that exhibits property sharing (Kameyama, Passonneau, & Poesio 1993). Property sharing holds if Cb_{U_i} appears in the Cf list for the previous utterance ($Cf_{U_{i-1}}$), and in both utterances it was mentioned as a grammatical subject (subject property sharing), or in both utterances as a non-subject (non-subject property sharing). Otherwise, Cb_{U_i} is null. The transition types are:

2.1 RET1 retain (subject property sharing): $Cb_{U_i} = Cb_{U_{i-1}}$

2.2 RET2 retain (non-subject property sharing): $Cb_{U_i} = Cb_{U_{i-1}}$

2.3 EST establish: not RET1, not RET2, but Cb_{U_i} is a member of $Cf_{U_{i-1}}$

2.4 NULL null: Cb_{U_i} is null

Fig. 1 illustrates a narrative excerpt, the NP coding for the excerpt, and the centering data structures. In addition to prosodic phrase units, Chafe's (Chafe 1980) transcriptions

[3] In (Brennan, Friedman, & Pollard 1987), constraint 1 states that *there is precisely one Cb*; in subsequent personal communication with one of the co-authors, I learned that this constraint has been relaxed.

represent all lexical articulations including repeated and incomplete words and phrases; non-lexical articulations such as "uh, um, tsk"; vowel lengthening as indicated by '-'; and location and duration of pauses in seconds (bracketed numbers, or ".." for very short pauses). Prosodic phrase numbers (column 2 of the narrative excerpt) have two fields, the first for phrases with sentence final intonation ("." or "?"; cf. above) and the second for intervening phrasal units.[4] Note that utterance 6 (column 1 of narrative excerpt) ends with an NP fragment and repair ("*uh uh a don uh a goat*") and utterance 7 begins with the next prosodic phrase. The NP coding table below the excerpt shows the sequential position of the NPs from the excerpt (column 1), their utterance numbers (column 2), referential indices (column 3), and grammatical roles (here subject, direct object, or prepositional object).[5] Note that the repair NP (*uh a goat*) is not coded for grammatical role, and doesn't figure in the centering data structures.

The table at the bottom of Fig. 1 shows the centering data structures. The Cf lists are the same in both versions; the referents are ordered with respect to the obliqueness of the grammatical role of the associated NP. In Version A, Cb_{U_7} is the same as Cb_{U_6} and Cp_{U_7}, and the transition to U_7 is CON by rule 1.1 above. In Version B, Cb_{U_6} is null because neither 12 nor 13 is evoked in Cf_{U_5} (not shown). Cb_{U_7} is null because property sharing does not apply. The transition to U_7 is NULL by rule 2.4 above.

3.2 Segmentation Study and Coding.

As described in (Passonneau & Litman 1993), linear segmentation of the narratives was performed by untrained subjects using a naive notion of intention as the segmentation criterion. Each narrative was segmented by 7 subjects; except in a few cases, no subject segmented more than one narrative. Subjects were given transcripts, and instructed to place segment boundaries wherever the speaker had completed one communicative task and begun a new one. They were restricted to placing boundaries between prosodic phrases. To focus their attention on the criterion, subjects' were also instructed to label each segment with a brief description of the speaker's intention. The instructions explained intention in common sense terms and by example (cf. (Passonneau & Litman To appear)). The size and number of segments per subject per narrative varied widely.

Subjects assigned boundaries at different rates, but all subjects assigned boundaries relatively infrequently. I refer to the number of subjects placing a boundary at any interphrasal location (ranging from 0 to 7) as *level of agreement*. Cochran's Q (Cochran 1950) quantifies the probability of the observed number of times each level of agreement occurs in the data, given the observed rate at which each subject assigns a boundary. Given the relative infrequency and different rates of boundary assignment, agreements of 4 or more are highly improbable in this data. Cochran's Q (Cochran 1950) gives extremely low probabilities ($.114 \times 10^{-6} < p < .6 \times 10^{-9}$). A partition of Cochran's

[4] The distinction roughly corresponds to complete versus intermediate prosodic phrases, as in (Beckman 1991) (Pierrehumbert 1980).

[5] For multi-clause utterances, Cf ordering parallels the degree of clause embedding.

U_n	Pros.Phr.	
6	4.02	[1.0 and [.45] and um...] along comes (a $\text{man}_{12})_{NP_{20}}$ with (a $\text{donkey}_{13})_{NP_{21}}$.
	5.01	Uh uh a don uh (a $\text{goat}_{13})_{NP_{22}}$.
7	6.01	[.45] And $(\text{he}_{12})_{NP_{23}}$ comes along.. by.. you know,
8	6.02	[.4] $(\text{ZERO}_{12})_{NP_{24}}$ passes him_4.

NP Coding				
NP id	U_n	Surface NP	Index	Gram. Role
20	6	a man	12	subj
21	6	a donkey	13	prep obj
22	6	a goat	13	*n.a.*
23	7	he	12	subj
24	8	zero	12	subj
25	8	him	4	dir obj

Centering Transitions		
U_n	Version A	Version B
U_6	Cfs: [12, 13] Cb: 12; Cp: 12	Cfs: [12, 13] Cb: NULL
U_7	Cfs: [12] Cb: 12; Cp: 12 Tran: CON ($\text{Cb}_{U_7} = \text{Cb}_{U_6}$, $\text{Cb}_{U_7} = \text{Cp}_{U_7}$)	Cfs: [12] Cb: [12] Tran: EST ($\text{Cb}_{U_7} \in \text{Cfs}_{U_6}$; subject property)
U_8	Cf: [12, 4] Cb: 12; Cp: 12 Tran: CON ($\text{Cb}_{U_8} = \text{Cb}_{U_7}$, $\text{Cb}_{U_8} = \text{Cp}_{U_8}$)	Cfs: [12, 4] Cb: 12 Tran: RET1 ($\text{Cb}_{U_8} = \text{Cb}_{U_7}$; subject property sharing)

Figure 1: Utterances, NPs, and Centering

Q by number of subjects (N) was consistently significant for N $\geq$ 4. Thus we took agreement among at least 4 subjects as a threshold for empirically validated boundaries. Boundaries that fewer subjects agreed on are not necessarily invalid, but are not used here.

Fig. 2 in section 4.1 below illustrates an excerpt with two segment boundaries. The boundary at (11.04,12.01) was identified by 5 subjects; the one at (12.03,13.01) by 4 subjects. There were 3 subjects in agreement on the intervening segment, who labeled it variously as:

- tells what really happened[6]
- then states actual situation of boy wanting a whole basket of pears
- boy wanted one basket

[6]This label contrasts with the label for previous segment, "predicts what will happen in story".

4 Two Corpus Analyses

4.1 Centering Transitions and Segmentation

In this section, I examine how centering transitions correlate with segmentation in the coded Pear corpus. I test Grosz et. al's (Grosz, Joshi, & Weinstein 1986) proposal regarding the preference for maintaining the same Cb across utterances by looking at the relative frequency of the various transition types. I also examine the correlation of transition types with segment boundaries. Positive results would be promising support for the hypothesis that centering could be used to mediate between local and global discourse processing in natural language processing systems. Understanding systems could potentially use centering as a source of knowledge for inferring global speaker intentions. Generation systems could potentially use centering to constrain lexico-grammatical choice both within and across segment boundaries. However, the results conflict with predictions about the relative preference for the transition types. Also, there is relatively poor correlation between centering and segment boundaries in this corpus. Although the negative results presented here do not by themselves disconfirm the hypotheses about centering transitions and global structure, at the least they reflect the inherent difficulties of empirical investigation. After presenting the results, I conclude with a brief discussion of some of these difficulties.

The two centering algorithms were implemented in Perl, and tested on the five narratives coded for centering. The input to the algorithms consists of the NP and utterance features described in §3 above, and illustrated in the middle of Fig. 1. The output consists of the Cfs for the utterance, the Cb, and the centering transition, as illustrated at the bottom of Fig. 1.

Table 1 correlates centering transition types at utterance locations with locations of empirically validated boundaries. There were relatively few SHIFT2 in Version A; similarly for RET2 transitions for Version B. Consequently, SHIFT2 was collapsed with SHIFT1 (Version A), and RET2 with RET1 (Version B). The rows labelled "Total %" indicate the relative frequency across transition types in each narrative. On average, Version A has 44% CON, 6% RET and 50% SHIFT. Version B has 27% RET, 21% EST and 52% NULL. In this corpus, the most frequent transition types are the ones predicted to represent the least coherence across utterances, and predicted to be the least frequent, namely SHIFT (Version A) or NULL (Version B). RET, the second preferred transition in Version A, is extremely infrequent, whereas the two more preferred transitions in Version B (RETAand EST) are about equally frequent.

The percentages in parentheses across the first row of Table 1 for each narrative represent the proportion of boundary utterance pairs (B) to all utterance pairs (N+B) for each transition type (columns 3-5), and for the narrative as a whole (column 6). The percentages show that in general, SHIFT and NULL transitions are more strongly associated with boundaries than the other transition types. That is, SHIFT or NULL transitions occur with boundaries more often (about a quarter of the time on average)

Version A					
Narr.	Status	CON	RET	SHIFT	TOTAL
01	B	3 (.08)	0 (0)	19 (.30)	22 (.25)
	N	35	8	45	88
% of Transitions		.35	.07	.58	1.00
02	B	4 (.10)	0 (0)	15 (.29)	19 (.19)
	N	38	6	37	81
% of Transitions		.42	.06	.52	1.00
03	B	0 (0)	0 (0)	7 (.35)	7 (.17)
	N	19	2	13	41
% of Transitions		.46	.05	.49	1.00
08	B	2 (.10)	0 (0)	3 (.19)	5 (.14)
	N	19	1	13	32
% of Transitions		.54	.03	.43	1.00
09	B	1 (.03)	2 (.33)	5 (.14)	8 (.11)
	N	31	4	31	66
% of Transitions		.43	.08	.49	1.00
Version B					
Narr.	Status	RET	EST	NULL	TOTAL
01	B	2 (.12)	1 (.04)	19 (.29)	22 (.25)
	N	15	26	47	88
% of Transitions		.16	.24	.60	1.00
02	B	2 (.07)	1 (.05)	16 (.31)	19 (.19)
	N	27	18	36	81
% of Transitions		.29	.19	.52	1.00
03	B	0 (0)	0 (0)	7 (.33)	7 (.17)
	N	11	9	14	34
% of Transitions		.27	.22	.51	1.00
08	B	2 (.13)	0 (0)	3 (.20)	5 (.14)
	N	13	7	12	32
% of Transitions		.41	.18	.41	1.00
09	B	0 (0)	0 (0)	8 (.19)	8 (.11)
	N	17	16	33	66
% of Transitions		.23	.22	.55	1.00

Table 1: Correlation of Centering Transitions and Segment Boundaries

Narr.	Ver.	REC	PRE	ERR
01	A (SHIFT)	.86	.30	.44
	B (NULL)	.86	.29	.45
02	A (SHIFT)	.79	.29	.41
	B (NULL)	.84	.31	.39
03	A (SHIFT)	1.00	.35	.32
	B (NULL)	1.00	.33	.34
08	A (SHIFT)	.60	.19	.41
	B (NULL)	.60	.20	.38
09	A (SHIFT)	.63	.14	.46
	B (NULL)	1.0	.19	.45
Avg.	A (SHIFT)	.78	.25	.41
	B (NULL)	.86	.26	.40

Table 2: Recall, Precision, Error Rate for NULL and SHIFT

than the other transition types do.[7] However for SHIFT, this pattern is less pronounced on narrative 8 and does not hold for 9. At the same time, comparing the B and N values shows that the likelihood of a boundary given a SHIFT or NULL is not particularly high: most transitions are SHIFT or NULL, and most SHIFT or NULL occur at non-boundaries. This is illustrated in a different way in Table 2.

Table 2 presents the recall, precision, and error rate for SHIFT and NULL as a means of identifying segment boundaries. Recall is the ratio of the number of times a boundary is correctly predicted to the total number of boundaries; perfect recall is 1. Precision is the ratio of the number of times a boundary is correctly predicted to the total frequency of the transition type (SHIFT or NULL); perfect precision is 1. Error rate is the ratio of the number of incorrect associations between a transition type and an utterance pair (i.e., SHIFT/NULL at a non-boundary, or no transition at a boundary) to the total number of utterance pairs (i.e., boundaries plus non-boundaries); perfect error rate is 0. On average, the two versions of centering perform quite comparably. For both versions, recall is fairly good but precision and error rate are very poor. This tradeoff reflects the generally high frequency of SHIFT and NULL in the data. Also, both versions perform better on narratives 1,2 and 3 than on 8 and 9.

Neither version of centering correlates well with the segments in the coded Pear corpus. CON and RET of Version A occur relatively infrequently at boundaries, but they also occur less frequently overall than predicted in (Grosz & Sidner 1986). The same is true for RET and EST of Version B. SHIFT or NULL occur too frequently to be predictive of segment boundaries. It might be argued that the segments in the coded Pear corpus simply occur at too large a granularity of structure for centering transitions to correlate

[7]For SHIFT, CON and RET, the averages are 25%, 6% and 7%, respectively; for NULL, RET and EST, the averages are 26%, 9% and 2%. Note that there is a lot of spread around the averages, i.e., a lot of variation across narratives.

Seg	U_n	PrPhr	
6	28	11.01	And you think "Wow,
		11.02	this little boy's$_i$ probably going to come and see the pears,
	29	11.03	[.4? and [.3]] he$_i$'s going to take a pear or two,
		11.04	and then.. go on his way."
7	30	12.01	[1.3 [.75] U-m] but $\boxed{\text{the little boy}_i}$ comes,
			(Cb_{V_A}=i; Cb_{V_B}=NULL)
	31	12.02	[1.8?.. a-nd u-h [1.0]] he$_i$ doesn't want just a pear,
	32	12.03	he$_i$ wants a whole basket.
			($Cb_{V_A,\ V_B}$=i)
8	33	13.01	[.55] So $\boxed{\text{he}_i}$ puts the- [.2] bicycle down,
			($Cb_{V_A,\ V_B}$=i; TRAN$_A$=CON; TRAN$_B$=RET)
	34	13.02	and he$_i$ [.25] you wonder how he$_i$'s going to take it with this.

Figure 2: NULL and SHIFT at segment boundaries

well, given that SHIFT and NULL transitions occur much more often than the segment boundaries that 4 subjects agreed on. Since some subjects assigned relatively more boundaries, I also examined the correlation of SHIFT and NULL with cases where any subject posited a boundary. The general outcome was that precision showed an improvement, but recall and error rate were significantly worse.

Another problem is suggested by the boundaries shown in Fig. 2. Note that the utterances at both segment onsets begin with a pause, and the first lexical item in both cases is a word that can be used as a discourse marker, or cue word (e.g., "*but, so*"). Hirschberg & Litman (Hirschberg & Litman 1993) found that an utterance initial cue word is likely to function as a discourse marker. Note also that the centering transitions at the second boundary are the type least likely to be correlated with boundaries: CON (Version A) or RET1 (Version B). Evidence suggests that if the pause and cue word information were considered here, performance could improve.

In (Litman & Passonneau 1995b), we discuss the need to incorporate multiple cues to enhance an algorithm designed to automatically identify the empirically validated boundaries in the Pear corpus. Unlike centering, it uses global as well as local context. It was originally tested on ten narratives rather than five (Passonneau & Litman 1996); the average recall of .50 was lower than the results for centering of about .75-.85 shown in Table 2, precision was comparable (.30 vs. .25), and the error rate of .19 was much better than the centering error rate of about .40. In (Litman & Passonneau 1995b), we find that if the input also includes information about the distribution of cue words, pauses, and types of prosodic phrases, precision can be markedly improved without degrading error rate, and only modest degradation of recall. An algorithm derived through machine learning that iteratively learns on 90% of the corpus of ten narratives and tests on the reserved 10% achieves the following scores on boundaries agreed on by at least 4 subjects: recall of .30, precision of .71, and error rate of .10 (Litman & Passonneau 1995b). Human performance on the same boundaries yielded a much better recall of .73 but a much poorer precision of .55; error rate was the same.

As suggested by the discussion of the second boundary in Fig. 2, the lack of correlation of centering with segmentation in the coded Pear corpus is potentially a consequence of the need to consider multiple knowledge sources when computing coherence of utterance transitions. This may be particularly true for spoken discourse, where prosody provides

Seg	U_n	PrPhr	
3	9	3.03	[1.6? and [.5]] anyway,
		3.04	he$_j$ comes down with a load of pears,
	10	3.05	and he$_j$ [.25] puts them into the basket,
	11	3.06	[.6] and then he$_j$'s..going back up into the tree,
	12	3.07	it's like he$_j$'s..been doing this all day,
	13	3.08	and.. it's just a monotonous kind of thing for him$_j$.
4	14	4.01	[.75] And a man$_k$ comes along with a goat,
	15	4.02	[.7] and the goat obviously is interested in the pears
	16	5.01	But the man$_k$ just..walks by with the goat.
5	17	6.01	And $\boxed{\text{the man}_j \text{ up in the tree (NP}_{36}\text{)}}$ doesn't even notice.

Figure 3: Context Illustrating use of Phrasal NP to Avoid Ambiguity

an additional source of information. But one of the differences between the two versions of centering examined here suggests another factor: how to interpret the relative coherence of using definite pronouns versus phrasal NPs. The two versions differ on Cb of U_{30} in Fig. 2, hence on how coherent the transition to U_{30} is. Version B requires the Cb to be realized by a definite pronoun (Kameyama 1985). In Version B, U_{30} has a NULL Cb because the boxed subject NP in U_{30} is phrasal rather than pronominal. In Version A, the referent of this NP is the Cb of U_{30}. In section 4.2, I investigate local and global discourse factors constraining the choice between a pronoun and a phrasal NP. From this analysis, I abstract specifications for an algorithm for generating discourse anaphoric NPs. The algorithm is described and evaluated in (Passonneau Forthcoming). Below in section 5, I briefly summarize the interaction between centering and informational constraints in the algorithm, and present the evaluation results.

4.2 Surface Form and Segmentation

In this section, I address the issue of how much information an NP encodes, relative to the current segment context. In contexts where multiple discourse entities have the same semantic attributes, an NP intended to discriminate one of those entities from the others may need to encode relatively more specific semantic content. With Version B of centering, some NULL transitions correspond to the shift from a pronoun to a full NP (cf. transition at U_{33} in Fig. 2). In evaluating whether shifting to a more informative NP independently correlates with segment transitions, the effects of the informational context must be isolated. Both new segments in Fig. 2 begin with an utterance whose subject NP corefers with a definite pronoun subject in the preceding utterance. As noted above, in one case the subject is a phrasal NP (U_{30}) and in the other case (U_{33}), the subject is a definite pronoun. Note that a pronoun or a zero could have replaced the phrasal NP with no loss of information. The subject of U_{17} in Fig. 3 is a contrasting case of a phrasal NP, "*the man up in the tree*", (also at a segment onset) that is informationally necessary. There are two men to distinguish between: man$_j$—in the tree, and man$_k$—with the goat. Either the pronoun "*he*" or the bare NP "*the man*" would have been ambiguous between the two men.

Here I characterize four possibilities regarding the semantic informativeness of an NP relative to its context. Three of them pertain to what Dale (Dale 1989) refers to as

1	A man_1 saw a $ladder_2$ leaning against a pear $tree_3$.		
2a.	Later, $\boxed{\text{he}_1}$ moved $\boxed{\text{it}_2}$ to a different tree.		
	well-specified	+adequate +efficient −increasing	
2b.	$\boxed{\text{The man}_1}$ moved $\boxed{\text{the ladder}_3}$ to a different tree.		
	over-specified	+adequate −efficient −increasing	
2c.	$\boxed{\text{It (?)}}$ was tall.		
	under-specified	−adequate −efficient	
2d.	$\boxed{\text{The contented pear picker}_3}$ was almost done for the day.		
	over-determined	+adequate +efficient +increasing	

Figure 4: Categories of Informativeness relative to context

informational adequacy and efficiency: the speaker should be sufficiently informative to unambiguously identify the intended referent (adequacy), and the speaker should be no more informative than necessary (efficiency). The boxed NPs in (2a) of Fig. 4 illustrate NPs that are both adequate and efficient, or what I refer to here as well-specified. It is clear what the two pronouns refer to; less informative forms (zero pronouns) would be ungrammatical. The phrasal NPs in (2b) are adequate but not efficient, thus over-specified. The pronominal NP in (2c) is inadequate, or underspecified. *It* in (2c) could refer either to the ladder or the tree (efficiency does not apply to inadequate NPs). A fourth possibility is that an NP may perform two functions, to identify the referent and to add information about it, as in (2d). I refer to this case as over-determined. In Fig. 4, the feature `+/- increasing` applies to the category `+adequate` to distinguish between over-specified and over-determined NPs. Below I discuss the correlation of segmentation with over-determined and over-specified NPs in the Pear corpus.

As noted in §3.1, there were 1,814 discourse anaphoric NPs in the corpus. Of these, only 1 was over-determined. Table 3 depicts a typical chain of coreferential NPs corresponding to a major protagonist in the Pear story (same as man_j in Fig. 3). Each NP is shown in its sequential order along with its location coding (NP sequence number, utterance number, prosodic phrase number and segment number, in columns 1-4 respectively). A '+' in column 6 of Table 3 indicates that the NP was over-specified, and a '++' that it was over-determined. The symbol "SEG" indicates whether the utterance containing the NP was a segment onset.

NP no.	U_n	Pros.Phr.	Seg	NP form	+/-Over-Spec.
4	2	1.03	1	a man	
9	3	2.01	2	him	
10	4	2.01	2	he	
12	5	3.01	2	he	
17	7	3.02	2	he	
19	8	3.04	3	he	
20	9	3.05	3	he	
23	10	3.06	3	he	
25	11	3.07	3	he	
29	12	3.08	3	him	
36	16	6.01	4	the man up in the tree	-; SEG
38	17	7.01	4	he	
41	18	7.02	4	zero	
44	19	8.02	5	he	+; SEG
49	20	8.04	5	he	
50	20	8.04	5	he	
54	21	9.01	5	he	
56	21	9.01	5	he	
57	22	9.02	5	he	
59	24	9.02	5	he	
102	39	17.03	9	the man up in the tree	+; SEG
105	40	18.01	9	he	
107	41	18.02	9	he	
112	43	19.02	10	the man	
114	44	21.01	10	the man	+
199	78	35.03	17	that old man	++
206	79	38.01	18	the old man	+; SEG
208	80	38.03	18	he	
210	81	39.01	18	he	
211	82	39.01	18	he	
213	83	40.01	18	he	
217	86	40.03	18	he	
219	87	41.01	18	he	
220	88	41.01	18	he	
222	89	41.02	18	he	
224	90	42.02	18	he	
236	95	44.01	19	the old man	+
238	96	45.01	20	the old man	+
241	96	45.02	20	he	
256	100	51.01	20	he	

Table 3: A Chain of Discourse Referential NPs, Narrative 2

NP_{36} in Table 3, an over-specified NP, corresponds to the boxed NP in the last utterance of Fig. 3 above. The only over-determined NP in the corpus is NP_{199} of this narrative ("*that old man*"), which introduces the new attribute "old". In every other case, the semantic content of discourse anaphoric NPs duplicates information expressed earlier in the discourse. Of the 6 over-specified NPs, 4 correlate with a segment onset, including one case of a pronoun where a zero pronoun was permissible (NP_{41}). The correlation of informativeness with segmentation illustrated in Fig. 3 is tabulated for the whole corpus in Table 4.

For Table 4, I sorted the 1,814 discourse anaphoric NPs into phrasal NPs (PhrNPs; N=1,012), explicit pronouns (PROs: definite, indefinite, demonstrative; N=697), and zero pronominals (ZPs; N=105). These 3 classes form an ordering of semantic explicitness ($\prec_{expl}$): PhrNP $\prec_{expl}$ PRO $\prec_{expl}$ ZP. All pairs of co-indexed NPs were collected such that for each pair (NP_1, NP_2), NP_2 was the very next NP to corefer with NP_1, and $NP_2 \prec_{expl} NP_1$, or both NPs were phrasal. Cases where NP_2 was syntactically obligated to be more explicit than NP_1 were eliminated from consideration (e.g., NP_1 was a zero subject, and NP_2 was not a subject). This heuristic identifies potentially over-specified definite pronouns whose antecedent is a zero, but will fail to identify over-specified definite pronouns whose antecedent is a definite pronoun. However, zero pronouns present different issues because they do not occur freely.

On the one hand, zero pronouns are syntactically constrained. On the other hand, factors other than informational constraints apparently govern the use of a zero pronoun. Many definite pronouns occur where a zero would be syntactically permissible and well-specified, as in the subject NPs in U_{29}, U_{31}, and U_{32} of Fig. 2. However, zeros would sound less acceptable than pronouns in U_{31} and U_{32}. For these reasons, in evaluating informational adequacy and efficiency, the use of a definite pronoun instead of a zero was not considered to be an instance of potential over-specification. With this qualification in mind, Table 4 shows that only 127 NPs were potentially over-specified; 58 occurred within the same segment (row 1 total) and 69 occurred in distinct segments (row 2 total).

Potentially over-specified NPs were sorted into four mutually exclusive categories—well-specified, segment onset, attentional shift, and other. If the entity evoked by the NP was mentioned within the current segment, the context considered consists of the current segment up through the current utterance. We saw in example (4) above (section 2) that in some contexts, the proposition expressed in an utterance can disambiguate a referring expression. If the entity was last evoked in an earlier segment, then the context considered includes this segment in addition. A potentially over-specified NP is classified as well-specified if a less explicit form would have resulted in ambiguity or vagueness. If a potentially over-specified NP is not well-specified, but occurred in the first utterance of a new segment, then it was classified as a segment onset. The segments in the coded Pear corpus 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)). The third category, attentional shift, consists of these cases. The fourth catch-all category includes cases where the utterance containing NP_2 is a repetition or repair of an earlier utterance, where the NP is contrastive, and other uncategorizable cases.

Table 4 indicates that most potentially over-specified NPs are either well-specified (38% of within segment cases; 53% of across segment cases), or occur at an empirically verified segment onset or a hypothesized attentional shift. Over a third (36%) of the within-segment cases correlate with an intra-segmental attentional shift. Nearly one third of over-specified NPs that occur in a different segment from the previous NP in the chain occur at a segment onset (29%). A smaller proportion (12%) correlate with an attentional shift within the current segment. However, note that it is not typically the case that a segment onset contains over-specified NPs. That is, an over-specified NP is somewhat predictive of a segment onset, but not vice versa.

In sum, NPs that re-evoke existing entities are rarely over-specified (68/1814, or 3.7%)

	Well-Specified	Segment Onset	Attentional Shift	Other	Total
Within segments	22	-	21	15	58
%	38%	-	36%	26%	100%
Across segments	37	20	8	4	69
%	53%	29%	12%	6%	100%
Totals	59	20	29	19	127
%	46%	16%	23%	15%	100%

Table 4: Factors Correlating with Potentially Over-Specified NPs

or over-determined (1/1814). But when over-specified NPs occur, they seem to correlate with segmental structure in the coded Pear corpus. Of the 68 over-specified cases (columns 2-5), 20 (20/68, or 30%) correlate with segment onsets independently identified by naive subjects, and 29 (29/68, or 42%) appear to correlate with intra-segmental attentional shifts. In this corpus, an over-specified NP is more likely than not to correlate with an attentional shift (49/68, or 72%). However, the reverse implication does not hold, that is, a segment shift is not likely to be signaled by an over-specified NP. Thus in this corpus, neither centering, nor the occurrence of over-specified NPs, consistently signal segment shifts.

4.3 Discussion

We have seen that the relative frequency of centering transitions does not reflect the predicted preference order (Grosz, Joshi, & Weinstein 1986). SHIFT (Vers. A) and NULL (Vers. B) are the least preferred transitions, but occur most often. Due to their relatively high frequency compared with the infrequency of segment boundaries, they correlate poorly with segmentation. In addition, the relative infrequency of CON (Vers. A) or RET (Vers. B) shows that the Cf ordering of entities in an utterance is not a good predictor of the Cb of the next utterance. Although the two versions of centering perform comparably, they differ in how they use the input surface features of grammatical role and pronominalization. This leads to different analyses for the same input. For example, because version A is treated here as requiring every utterance to have a Cb, many Cbs are not expressed by pronouns (29% on average). Because in Version B the Cb is null in the absence of property sharing, many utterances with definite pronouns have a null Cb (18% on average). In both cases, centering accounts for relatively few of the pronouns in the corpus. Whether additional information could be incorporated in centering in order to simultaneously compute globally relevant inter-utterance transitions and constrain pronominalization is an area for future investigation. However, one important result of the second data analysis is that the distinction between pronouns and phrasal NPs is not absolute, but is relative to context. In section 5, I present results showing that centering nevertheless contributes significantly to the performance of an algorithm for generating discourse anaphoric NPs.

The most striking result of the data analysis is that discourse anaphoric NPs are typically well-specified. The algorithm based on these studies that I present in (Passonneau Forthcoming) exploits the interdependence among informational constraints, centering, and global discourse structure. I present results of an evaluation of how well the algorithm simulates the observed human choices of pronominal versus full discourse anaphoric noun phrases in a test sample from the Pear corpus. The evaluation results demonstrate the utility of the empirically derived focus spaces for defining the global informational context, and of centering for defining the local context. In the next section, I briefly summarize the algorithm and present the evaluation results from (Passonneau Forthcoming).

5 Generation Algorithm

5.1 Centering Component

Centering has often been implemented as a means of constraining the interpretation of discourse anaphoric expressions in an utterance U_i, given an interpretation of the preceding utterance U_{i-1}. In the case of understanding, the surface form of U_i is a given. In using centering for generation, the surface form of U_i is not given; constraints on the surface form of the utterance are the target output. In this case, we need an alternative means of identifying the Cb of U_i, one framed in terms of its semantic and pragmatic relation to U_{i-1}. 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. This is a domain specific characterization that does not extend to narrative.

For the generation algorithm summarized here, I hypothesized a Cb for each utterance of the narrative corpus examined here, based on the centering rules presented in (Grosz, Joshi, & Weinstein 1995), and on the conventions of narrative discourse. 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. As discussed in (Passonneau Forthcoming), two successive utterances that describe situations S_i and S_j are in a narrative relation if S_i causes, enables, occasions, or temporally precedes S_j. Salience of entities in a situation is a partial order reflected to some degree in thematic structure (Jackendoff 1990).

The input to the centering component consists of the discourse entities that were the center and Cfs of U_{i-1}, and the entity that is to be the center for U_i. The output consists of defeasible constraints on the surface form and grammatical role of Cb(U_i), and possibly of other entities in U_i. For example, if Cb(U_i) is constrained not to be a definite third person pronoun, then no other entities in U_i can be realized as a pronoun (Grosz, Joshi, & Weinstein 1983) (Grosz, Joshi, & Weinstein 1995).

5.2 Informational Constraints: ae-describe

Centering posts constraints on the realization of the current center, and possibly on the realization of other entities in the current utterance relative to the center. The semantic content of a discourse anaphoric NP, and certain corollary constraints on its syntactic form, are handled after centering by the 4-place relation ae_describe. AE_describe relates a discourse entity, the current focus space context, the current utterance context, and the NP that realizes the discourse entity. It requires a discourse anaphoric NP to be adequate, in the sense of (Dale 1992). To enforce adequacy, it uses Dale's algorithm for constructing a distinguishing description. Given a set of discourse entities in a context C, and attributes that are true of each entity in C, a distinguishing description of an entity e is a set of attributes A that discriminates e from all other entities in C. One difference from Dale's algorithm (Dale 1992) is that context C is first filtered by the current utterance context (but cf. (Dale & Haddock 1991) where the generation algorithm is extended to handle discourse entities mentioned in modifiers of an NP).

AE_describe also constrains a discourse anaphoric NP to be economical, which is similar to Dale's (Dale 1992) notion of efficiency, and Reiter's (Reiter 1990) notion of local brevity. Efficiency requires the cardinality of the set of attributes in a distinguishing description to be as small as possible. Dale and Reiter (Dale & Reiter 1995) point out that efficiency is computationally intractable, and empirically incorrect. Local brevity is defined as a preference for the shorter of any two distinguishing descriptions (e.g., for "*the small dog*" over "*the sleeping female dog*" in a context with multiple dogs, only one of which is small, and simultaneously asleep and female). Economy, as described in (Passonneau Forthcoming), differs from efficiency and local brevity in that it is used to generate definite pronouns as well as full NPs, and it relies simultaneously on semantic and syntactic ordering of attributes. The most economical syntactic category generated by ae_describe is a definite pronoun (PRO). The next most economical syntactic category is a minimal noun phrase consisting solely of a determiner and a common noun head (MIN). Within the category MIN, nouns that encode basic categories are preferred, as in (Dale & Reiter 1995). Thus "*cannibal*" would be preferred over "*anthropophagi.*" The third category is FULL, for NPs that have modifiers. AE_describe makes use of an ordering on attributes corresponding to whether they map to PRO, MIN or FULL NPs. Within each subset of attributes, it attempts to compute a distinguishing description, and moves on to the next set in case of failure.[8]

The evaluation results presented below pertain to a simplified implementation of ae_describe that does not use the utterance context. For the focus space context, it uses focus spaces derived from the empirical study of segmentation of the Pear stories

[8] AE_describe also makes use of a focus ordering on attributes, derived from analysis of usage in the Pear stories. For details, cf. (Passonneau Forthcoming).

Seg	U	
2	4	And he_1 fills his thing with pears,
	5	and $ZERO_1$ comes down
	6	and there's a basket hee_1 puts them in.
3	7	And you see passerbyers on bicycles and stuff go by.
	8	A-nd then a $boye_2$ comes by,
		on a $bicycle_3$,
	9	the man_1 is in the tree,
	10	and $\boxed{\text{the boy}_2}$ gets off the $bicycle_3$,
	11	and $ZERO_2$ looks at the man_1,

Figure 5: Over-Specified Full NP

described in (Passonneau & Litman 1996). When the discourse entity that ae_describe is generating an NP for has been mentioned in the current segment, than the focus space context consists of all the discourse entities mentioned in the current segment up to the current utterance. If the most recent mention of the discourse entity is in a segment S_j prior to the current segment S_k, then there are a variety of possible relations between S_j and S_k. For purposes of establishing a base line evaluation, I assume that both S_j and S_k are currently in focus. This simplifies the implementation, and corresponds to the following minimal assumptions about S_j and S_k. If the current utterance is a resumption of S_k (Grosz & Sidner 1986), then the new focus space is $S_j \cup S_k$, and any intervening focus spaces are no longer in focus. Else, if the current utterance is the onset of an entirely new segment, then I assume that S_j remains in focus until the current utterance has been processed, and that the most recent context for the discourse entity—S_k—must be accessed in processing any new mention of that entity. It is possible that other previous focus spaces are also in focus, but lacking any global model of the relation among segments that would control which focus spaces to select, I take the minimal approach of ignoring other focus spaces. As the results in the evaluation section indicate, this approach works rather well.

5.3 Integration of Centering and Informational Constraints

Centering constraints and the results of ae_describe can conflict. Centering allows a pronoun that realizes the Cb to be under-specified (cf. example (1) in section 2 above), whereas the output of ae_describe is always well-specified. In (Passonneau Forthcoming), I argue that the boxed NP in U_{10} of Fig. 5 does not realize the current center, and is thus constrained by centering to be realized by an expression other than a definite pronoun. In this utterance, a definite pronoun would be well-specified, and would thus be generated by ae_describe. However, a definite pronoun would at first be assigned a reading in which it corefers with the subject of the preceding utterance, thus producing a garden path effect. Because of such examples, centering constraints take priority in cases of conflict. I will illustrate by briefly describing how the integrated algorithm handles the boxed NP in U_{10} of Fig. 5.

Utterances 9 and 10 of Fig. 5 are presumed to be in a narrative relation. The first utterances in segment 2 of Fig. 5 describe the setting in which the man introduced in the preceding segment is picking pears. 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. Thus I assume U_9 and U_{10} are in an occasion relation (Hobbs 1985). Because they are in a narrative relation to one another, the Cb of U_{10} will be the thematically most salient entity in the situation described in U_{10} that is also part of the situation in U_9. In (Passonneau Forthcoming), I argue that the only common entity is the discourse reference time. The discourse reference time is rarely realized as a definite pronoun, and cannot be in this case.[9] As a result of applying rule 1 from (Grosz, Joshi, & Weinstein 1995), no entities in U_{10} can be realized by a definite pronoun. After centering has posted this constraint, ae_describe attempts to generate descriptions for the two entities in U_{10}, $\texttt{boy}_2$ and $\texttt{bicycle}_2$. The category PRO is blocked, leaving MIN and FULL in the syntactic preferences for the output of ae_describe. In both cases, the output is MIN, a minimal noun phrase whose head is a common noun expressing a basic category.

5.4 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). 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 evaluated.

Table 6 shows the results: the number and percentage of correct outputs for each condition. Column 1 indicates the narrative and 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 the preceding section: the union of the set of discourse entities mentioned in the current segment (up to the current utterance) with the set of

[9] Note the awkwardness of a sentence like "*It's when the boy gets off the bicycle.*" if it were to follow U_9.

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 6: Evaluation under Three Conditions

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 6 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 6 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.

The implementation results support the conclusions derived from the corpus analysis regarding informational constraints. Most discourse anaphoric NPs are well-specified relative to the current hypothesized focus space. The implementation results also bear out the hypothesized role of centering in allowing under-specified or over-specified NPs in certain restricted cases. 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.

6 Conclusion

The two empirical studies reported in section 4 provide a rich view of the use of discourse anaphoric NPs in a corpus of spoken narratives. Unsurprisingly, the results add some understanding of, while raising many questions about, the interrelation among surface form of discourse anaphoric NPs, local utterance processing, and global discourse structure. First I review the results and limitations, which pertain mainly to constraints on surface form of NPs. I show how the *ae_describe* relation among discourse entities, surface NPs and focus structure accounts for the main empirical generalizations. Then I review how centering in combination with *ae_describe* could address some of the limitations. I conclude with areas for future research that would address the relation between local and global processing.

Seg.	U	
21	105	[.6] So $(\text{they}_1)_{NP_{255}}$'re walking along,
	106	and $(\text{they}_1)_{NP_{256}}$ brush off $(\text{their pears}_3)_{NP_{257}}$,
	107	and $(\text{they}_1)_{NP_{258}}$ start eating $(\text{it}_3)_{NP_{259}}$.
22	108	Then $\boxed{(\text{they}_1)_{NP_{260}}}$ walk by- [.35] $\boxed{(\text{the man who was picking the pears}_2)_{NP_{261}}}$

Figure 7: Excerpt from Narrative 1

6.1 Centering and Global Structure

Centering transitions from two versions of centering, version A from (Brennan, Friedman, & Pollard 1987) and version B from (Kameyama To appear), are ordered in terms of relative coherence. The results presented in Tables 1 and 2 showed that centering is poorly correlated with segmentation in the coded Pear corpus. First, the transitions predicted to be least coherent occur very frequently. Maintaining the hypothesis that the transition types as proposed in (Brennan, Friedman, & Pollard 1987) (Kameyama To appear) reflect coherence would require justification for the view that all the speakers in the corpus were relatively incoherent in comparison with other speakers in informal discourse, or that spoken language is relatively incoherent. Second, the distribution of transition types does not correlate well with segments in the corpus.

The results might indicate that the data structures of centering need redefinition. Yet a third version of centering using a different means for ordering the Cfs, and exploiting the contrast between pronouns and phrasal NPs differently, might perform better. But evidence from other studies (Litman & Passonneau 1995b) of the Pear corpus suggest that information from referring expressions alone is insufficient to predict global structure. A more serious problem is that both the theoretical and empirical characterization of segmental structure may need revision. The segments in the coded Pear corpus are empirically derived, and represent highly significant agreement across subjects on the general intention behind and location of certain segments. At best, they represent certain well-defined shifts of attention. That over-specified NPs correlate both with segment onsets and intra-segmental temporal shifts suggests that segments may be too coarse-grained. But, as Passonneau and Litman (Passonneau & Litman To appear) argue, it is likely that there is at best only a rough correlation between the semantic units of discourse and sequences of utterances. To mention only a few problems in relating the two, a single utterance might contribute to more than one segment, or a shift from one unit to another might occur across multiple utterances.

6.2 Gricean Informational Constraints: *ae_describe*

Table 3 shows a chain of discourse anaphoric NPs for one discourse entity in one narrative. It illustrates a pattern that holds in general for discourse anaphoric NPs in the coded Pear corpus; most discourse anaphoric NPs are pronouns. The heuristic for identifying potentially over-specified discourse anaphoric NPs yielded 127 out of 1,814 discourse anaphoric NPs in the corpus. Of these, only 68 were in fact over-specified (cf. Table 4), showing that when a full NP is used discourse anaphorically, it is often necessary in order for the NP to be informationally well-specified. The data in Table 4 indicate that over-specified NPs may be one way to signal an attentional shift, given that 39% of the NPs in Table 4 correlate with a segment onset, or an intra-segmental shift of attention.

6.3 Future Work

The *ae_describe* relation insures that all NPs are informationally well-specified, relative to the current utterance and the current focus space. However, in restricted cases informational constraints need to be relaxed. As discussed in section 5, I integrate centering with *ae_describe* in order to allow two types of relaxation: the current Cb can be under-specified, or if the current Cb is not realized as a pronoun, other NPs in the utterance can be over-specified. In reference to Fig. 5, I discussed an utterance (U_{10}) containing an over-specified phrasal NP, where *ae_describe* would generate a definite pronoun. However, a pronoun would lead to a garden path effect: a definite pronoun subject would be inferred to corefer with the subject of the preceding utterance, resulting in an unintended interpretation.

The results from Table 4 summarized above suggest that informational constraints also need to be relaxed at segment onsets, or perhaps more generally, at shifts in focus structure. How to do so is a matter for future research, particularly given our lack of knowledge about how global structure is signaled. One possibility suggested by some of the examples presented here is that the particular attributes mentioned in a semantically rich discourse anaphoric noun phrase help constrain the relation of each new utterance to the global discourse model. For example, U_{17} in Fig. 3 (the onset of segment 5) illustrates a discourse pop (to a focus space corresponding to segment 3). It contains one NP; the evoked entity is not mentioned in the intervening segment 4, and the attributes of the entity were last mentioned in segment 3. These factors could potentially serve as a cue to close a focus space correlated with segment 4, and reopen the one associated with segment 3. Similar reasoning can be illustrated for U_{108}, a segment onset in Fig. 7. U_{108} contains two NPs, one of which continues (or retains) the Cb of U_{107}. This NP would thus provide a link of the current utterance, and its focus space, to the most recent focus space FS_{21}. The other NP expresses attributes last mentioned in segment 17, providing a link from the current utterance to a much earlier focus space. If the global structure is a tree, the links from 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 links might provide evidence about the nature of global structure, such as whether it is a tree or a lattice. In either case, I believe this is a fruitful arena for future research.

References

[Beckman 1991] Beckman, M. 1991. Notes on prosody. Ms.

[Brennan, Friedman, & Pollard 1987] Brennan, S. E.; Friedman, M. W.; and Pollard, C. J. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 155–162.

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

[Cochran 1950] Cochran, W. G. 1950. The comparison of percentages in matched samples. *Biometrika* 37:256–266.

[Dale & Haddock 1991] Dale, R., and Haddock, N. 1991. Generating of referring expressions involving relations. In *Proceedings of the European Association of Computational Linguistics.*

[Dale & Reiter 1995] Dale, R., and Reiter, E. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science.*

[Dale 1989] Dale, R. 1989. Cooking up references. In *Proceedings of the 27th Annual Meeting of the Association of Computational Linguistics*, 68–75.

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

[Fillmore 1971] Fillmore, C. J. 1971. Santa Cruz lectures on deixis. Technical report, Indiana University Linguistics Club, Bloomington, IN.

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

[Grosz, Joshi, & Weinstein 1983] Grosz, B. J.; Joshi, A. K.; and Weinstein, S. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Association of Computational Linguistics*, 44–50.

[Grosz, Joshi, & Weinstein 1986] Grosz, B. J.; Joshi, A. K.; and Weinstein, S. 1986. Towards a computational theory of discourse interpretation. Ms.

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

[Grosz 1977] Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding.* Ph.D. Dissertation, University of California, Berkeley.

[Hirschberg & Litman 1993] Hirschberg, J., and Litman, D. 1993. Empirical studies on the disambiguation of cue phrases.

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

[Jackendoff 1990] Jackendoff, R. 1990. *Semantic Structures.* Cambridge, MA: MIT Press.

[Kameyama, Passonneau, & Poesio 1993] Kameyama, M.; Passonneau, R.; and Poesio, M. 1993. Temporal centering. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 70–77.

[Kameyama 1985] Kameyama, M. 1985. *Zero Anaphora: The Case of Japanese.* Ph.D. Dissertation, Stanford University.

[Kameyama 1986] Kameyama, M. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 200–206.

[Kameyama To appear] Kameyama, M. To appear. Intrasentential centering. In Prince, E.; Joshi, A.; and Walker, M., eds., *Proceedings of the Workshop on Centering Theory in Naturally Occurring Discourse.* Oxford University Press.

[Lascarides & Oberlander 1992] Lascarides, A., and Oberlander, J. 1992. Temporal coherence and defeasible knowledge. *Theoretical Linguistics.* To appear.

[Linde 1979] Linde, C. 1979. Focus of attention and the choice of pronouns in discourse. In Givon, T., ed., *Syntax and Semantics: Discourse and Syntax.* New York: Academic Press. 337–354.

[Litman & Passonneau 1995a] Litman, D., and Passonneau, R. 1995a. Combining multiple knowledge sources for discourse segmentation. In *Proceedings of the 33rd Association for Computational Linguistics.*

[Litman & Passonneau 1995b] Litman, D., and Passonneau, R. 1995b. Developing algorithms for discourse segmentation. In *Working Notes of AAAI Spring Symposium Series on Empirical Methods in Discourse Interpretation and Generation.*

[Mann, Matthiessen, & Thompson 1992] Mann, W. C.; Matthiessen, C. M.; and Thompson, S. A. 1992. Rhetorical structure theory and text analysis. In Mann, W. C., and Thompson, S. A., eds., *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text.* Amsterdam: J. Benjamins Pub. Co.

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

[Passonneau & Litman 1993] Passonneau, R. J., and 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.*

[Passonneau & Litman 1996] Passonneau, R. J., and Litman, D. 1996. Discourse segmentation by human and automated means. Submitted for publication.

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

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

[Passonneau 1994] Passonneau, R. J. 1994. Protocol for coding discourse referential noun phrases
and their antecedents. Technical report, Columbia University.

[Passonneau Forthcoming] Passonneau, R. J. Forthcoming. Using centering to relax Gricean informational constraints on discourse anaphoric noun phrases. *Language and Speech.*

[Pierrehumbert 1980] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation.* Ph.D. Dissertation, M.I.T.

[Polanyi 1988] Polanyi, L. 1988. A formal model of discourse structure. *Journal of Pragmatics* 601–638.

[Reiter 1990] Reiter, E. 1990. *Generating appropriate natural language object descriptions.* Ph.D. Dissertation, Harvard University.

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

[Webber 1988] Webber, B. L. 1988. Tense as discourse anaphor. *Computational Linguistics* 14:113–122.

[Webber 1991] Webber, B. L. 1991. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 107–135.