

WHAT DOES *THAT* REFER TO? NON-PERSONAL PRONOUN ANAPHORA IN ENGLISH CONVERSATIONAL DISCOURSE

RANDALL B. SPARKS

This paper presents an account of the use of the demonstrative pronoun *that* based on an analysis of an informal English conversation. Anaphoric *that* was found to pattern differently from other pronouns in several interesting ways. Anaphoric *that* differs from *it* in the animacy of its referents in that most of the referents of *that* are non-physical objects, such as claims or propositions. Many of the referents of *that* must be inferred, because they have not previously been introduced into the discourse. Furthermore, some of the referents of *that* are partially cataphoric. The referent is a complex proposition, part of which is present as given or inferable information in the discourse, and part of which appears in the text after the use of *that*. Some cases of cataphoric *that* involve a process of reconstruction of the reference of the anaphor in which the speaker revises the identity of the referent of *that* after its use. It is argued that an adequate account of these features of anaphoric *that* must take into account both the text itself and the processes by which speakers and hearers produce and interpret anaphors in real time.*

1. INTRODUCTION. A great deal of attention has been focused on anaphora in recent research within the fields of linguistics, cognitive psychology, and artificial intelligence. This is understandable, as anaphora constitutes a prototypical example of the sort of problem that must be tackled in order to understand how human language works: How is it that people are able to refer to things and resolve or interpret such references, and what are the linguistic devices that are employed to perform this communicative function?

Most studies of anaphora can be divided into those that deal with intra-sentential, antecedent-bound anaphora,¹ and those that deal with general anaphora, which is usually, though not exclusively, inter-sentential. This paper is of the second type, as it involves the analysis of all referential uses of *that* which appeared in a naturally occurring conversation.

One can characterize the general (inter-sentential) accounts of anaphora as following one of three general approaches. First, there are text-based studies, primarily in the linguistics literature, which examine naturally produced texts, either conversational or written, in order to identify the textual correlates of the use of different anaphoric devices (such as pronoun vs. full NP vs. zero-anaphora). These textual correlates include such things as the distance separating an anaphor from its antecedent, the nature and amount of intervening material, and such structural features as narrative or rhetorical units, the boundaries separating them, and the hierarchical relationship among them (see Givón 1983, Fox 1987a,b).

Second, there are experimental studies, which appear primarily in the psychology literature. These commonly involve assessments of the processing difficulty in resolving anaphoric references, usually measured by response or reading times.² This involves a comparison of the processing difficulty of anaphoric expressions under two (or more) different conditions, either different types of anaphors in the same environment, or the same type in different environments. The results generally indicate that certain of the conditions facilitate anaphoric processing while others result in increased processing difficulty (cf. Marslen-Wilson & Tyler 1980; Marslen-Wilson et al. 1982; Tyler & Marslen-Wilson 1982; Walker & Yekovich 1987; Garrod & Sanford 1983; Clark & Sengul 1979; Dell et al. 1983;

* The following notational conventions are used in the examples:

- // Indicates the point at which the current speaker's utterance is overlapped by the beginning of the next speaker's utterance.
- [Joins two lines at the point where an overlap occurs, as when two people begin speaking simultaneously.
- (???) Indicates an unintelligible utterance or part of an utterance.
- (()) Material in double parentheses indicates non-linguistic events, such as laughter, coughing etc.
- (.) One, two, or three dots inside single parentheses indicates 'short', 'medium', and 'long' pauses, ranging from about 1/3 to 1 full second. Longer pauses are indicated by the number of seconds in single parentheses, e.g. (3.5).
- : indicates a lengthened syllable.

No attempt was made to capture subtleties in pronunciation in the transcription. Standard English spelling was used.

¹ Most studies of intra-sentential anaphora examine 'antecedent-bound' anaphora. That is, they analyze the (usually formal) constraints that exist between an anaphor and an antecedent that occurs at some earlier point in the same sentence. The more general approach to anaphora adopted in this paper may (and often does) involve an antecedent in the text, but it need not, as in the case of cataphoric expressions and referring expressions that require inference on the part of the hearer in order for the reference to be resolved.

² Some of the experimental studies fall into the intra-sentential group, others deal with general anaphoric processing.

McKoon & Ratcliff 1980; Bosch 1988).

Third, in addition to work in linguistics and cognitive psychology, anaphora has been studied within the framework of research on artificial intelligence (AI). This work can be roughly characterized as formalizing and implementing some of the results from the linguistic and psychological studies in the form of computer programs, either to test linguistic and psychological theories of linguistic processes or for possible use in natural language processing software. Examples of this type of work that are most typical of AI programs often include anaphora handling as one component of a larger natural language processing system.

Some research straddles the boundaries between linguistics, psychology, and AI, involving models of specific anaphoric processes, with less concern for their direct utilization in application systems (cf. Grosz 1977, Grosz & Sidner 1985). This latter type of work offers an opportunity to test and analyze the predictive power of various accounts of how anaphora works.

Each of these different research strategies has its strengths and weaknesses. AI programs can provide a rigorous test of a particular account of anaphoric use, but only with types of data that can easily be handled as input. This currently excludes unrestricted, naturally occurring human language – exactly the sort of data that must ultimately be accounted for. Text-based linguistic studies, on the other hand, deal with real, naturally produced discourse. But this work generally has focused primarily on the textual correlates of anaphora, that is, the *PRODUCT* of the processes involved in the production and resolution of anaphoric expressions, rather than on the processes that produce the text.³ Experimental studies address these processes more directly, but, for the most part, are necessarily limited to short, made-up sentences or texts as data, rather than texts of naturally occurring discourse.

Given these strengths and weaknesses, it should be a goal of research on anaphora (as well as other language phenomena) to find ways to combine the advantages offered by each approach. That is, the goal should be to develop an explanation of the phenomenon of anaphora that is characterized by the following three features. First, it should account for the types of complexities presented by naturally produced conversational discourse. Second, it should be presented in terms of the processes by means of which such discourse is produced (i.e., the account should be consistent with empirical results on the representation of the meaning of texts, memory use during comprehension, etc.) Finally, where possible, it should be explicit and specified in sufficient detail to allow at least partial implementation in a computational system that can be used to verify those aspects of the account that prove to be accurate and to point out the areas in which the account (perhaps unexpectedly) proves to be inadequate, thus providing possible directions for future investigation.

This paper takes some initial steps toward this goal. No experiment or computer system will be presented. Rather, this paper analyzes a naturally produced conversational discourse and identifies properties of anaphoric processes that account for the patterns of anaphora found in the text. In other words, this paper addresses the following questions. By examining the product (the text), what can we determine about the processes that produced it? What constraints on a theory of anaphoric processes can we identify by examining instances of anaphora in the text of a conversational discourse?

The decision to take a processing perspective on the analysis of a discourse was motivated to a large extent by the fact that a fully text-oriented perspective – the more common approach in studies of discourse – was found to be less than adequate in accounting for some of the anaphoric patterns found. The results of this study show several interesting differences between personal-pronoun anaphora and *'that'*-anaphora. These differences suggest that an account of the processing involved in the production and comprehension of discourse is essential to an adequate understanding of such linguistic phenomena as reference and anaphora.⁴

2. METHOD OF ANALYSIS. The method employed in this study was the following. A dinner conversation among four adults and one child was tape-recorded. Approximately one hour of this conversation was transcribed to produce the text, which will be referred to henceforth as 'the text' or 'the dinner text'. The transcribed text is from one continuous hour of conversation, with the exception of one break of approximately 10 minutes. From the text, a database was developed, with one record for each mention,⁵ or appearance in the text, of an anaphoric use of *that*. For each mention, the values of a number of associated features relevant to patterns of anaphora were recorded, including

³ This is not a criticism as much as a characterization of the current state of research. It makes sense to develop an account of what's there (in a text) before going on to account for how it got there.

⁴ A similar point is made by Givón (1989:234): "In dealing with pragmatics of definiteness and anaphoric reference, one deals essentially with communicative tasks in discourse. Atomic, Platonic notions such as 'identifiable', 'known', 'referred to earlier' etc. go only a short distance toward accounting for the wealth of structural devices used in the grammar of anaphora and definite reference in human language, let alone for the use of grammar in natural communication. A speaker-hearer based functional-pragmatic account is both more realistic and more far reaching in explaining the facts of structure and function."

⁵ The term 'mention' is from Du Bois 1987.

animacy, definiteness, information status, and certain semantic and pragmatic features of the referent. The contents of this database and the results of the analysis of *that*-anaphora are described below.

3. THAT-ANAPHORA IN CONVERSATIONAL DISCOURSE. A detailed account of personal pronoun anaphora is given for conversational and for written expository English texts in Fox 1987a, and for written narrative English in Fox 1987b. The basic pattern found for conversational texts was the following. The first mention of a referent within a sequence, or discourse unit, is accomplished with a full NP.⁶ After that, a pronoun is used through the remainder of that sequence. The use of a full NP to refer to the same referent is one way in which a speaker can signal that he or she understands the preceding sequence to be closed off and a new sequence to have begun. The basic patterns in written discourse are similar. In expository texts, the first mention of a referent is a full NP, after which a pronoun is used as long as that referent is 'active' and 'controlling', and intervening material is 'non-complex' and continues to refer to that referent. A full NP tends to be used for the same referent at the beginning of a new rhetorical unit. Similarly, in written narratives, a pronoun is used after an initial full NP. If another character's goals and actions are introduced, or a new narrative unit begins, the original referent will next appear as a full NP. These basic patterns may be complicated somewhat in the context of multiple, same-gender referents. In this case, a referent may continue to be referred to with a pronoun if grammatical role continuity is maintained. That is, a subject pronoun in one clause may be used to refer to the same referent as the subject of the preceding clause.

The studies in Fox 1987a,b are limited to anaphora with personal pronouns, as is the case with most work on anaphora. This is to be expected, since the prototypical function of anaphora is generally considered to be that of 'tracking' various participants through a discourse. This paper, on the other hand, focuses on the anaphoric use of the demonstrative pronoun *that* in a conversational discourse. In this conversational text, it was found that the use of what I will call '*that*-anaphora' differs from personal pronoun anaphora in a number of ways.⁷ Some of these differences are quite straightforward. There is a difference in the animacy of referents, for example; most referents of personal pronouns are human, while most referents of anaphoric *that* are non-animate.

3.1. STRUCTURE OF THE THAT-ANAPHORA DATABASE. A description of the features encoded in the database for each occurrence of anaphoric *that* in the text are presented in this section.

TEXTUAL MENTION STATUS is a feature used as part of an alternative perspective on the common notion of **INFORMATION STATUS**. This alternative was motivated by the fact that most formulations of information status are somewhat vague, combining both text-based and processing-based features. The most common terms for information status, **OLD** and **NEW**, do not clearly distinguish between the text-based characteristic of whether or not a referent has previously (or recently) appeared in the text and the processing-based feature of being 'new' conceptually for the hearer. Of course, these two often coincide, but they do not always do so. Several more-careful definitions of information status, such as those of Du Bois 1987 and Chafe 1987, describe information status primarily in such processing-based terms as 'active', 'semi-active', and 'accessible'. In this study, the processing-based features **FRAME-ACTIVATED** and **ACCESS-ACTIVATED** were used (see below). **ACCESS-ACTIVATED** corresponds most closely to the notion of 'active'. So, in this study, in an effort to maintain a distinction between text- and process-based features, the general notion of information status was split into its textual aspect (**TEXTUAL MENTION STATUS**), and its processing aspects (**FRAME-ACTIVATION** and **ACCESS-ACTIVATION**).

Furthermore, in an attempt to make a somewhat finer distinction than that between previously mentioned textually ('old') versus not previously mentioned ('new'), the range of values of textual mention status was defined to include (1) **NEW** for no previous mention, (2) **RECENT** for a previous mention within the last 20 clauses, (3) **DISTANT** for a last previous mention more than 20 clauses earlier, and (4) **PARTIAL**, for cases in which the referent is a complex proposition, only part of which has appeared explicitly at some previous point in the text.

The other features related to information status that were coded for each example of *that*-anaphora were **FRAME ACTIVATION**, **ACCESS ACTIVATION**, and **REFERENTIAL DISTANCE**. These are discussed in §§ 3.4 and 3.5 below.⁸

⁶ In the study of anaphora in conversation, Fox 1987a uses the analytical techniques of conversational analysis, in which discourse units such as sequences are understood in terms of 'turns' taken by participants in the discourse and 'adjacency pairs', sets of two (or sometimes more) turns which 'go together' by virtue of carrying out two complimentary parts of a common action, such question and answer, offer and acceptance, announcement and assessment, etc. See Fox 1987a, Chap. 2; Levinson 1983, Chap. 6; and Schegloff 1979, 1981.

⁷ Much of what can be said about the anaphoric use of *that* likely applies to the anaphoric use of *this* (and *these*, *those*) as well, making most of the patterns found for *that* relevant to non-personal pronoun anaphora in general. However, there are important differences between *this* and *that*. For example, *that* appears much more frequently as an anaphor than does *this* in the dinner text; there are only 12 such instances of anaphoric *this*, compared with 141 for *that*. This same difference in the frequency of occurrence of *this* and *that* was found by Passonneau 1989.

⁸ Referential distance, as measured by the number of clauses intervening between antecedent and anaphor, did not prove to be a useful measure in accounting for the patterning of anaphoric *that*. For example, in some cases, the referential

CATAPHORIC REFERENCE indicates that the referent appears in the text at some point after the anaphor, at least in part. A PARTIALLY CATAPHORIC REFERENCE occurs when the referent is complex, and part of it appears before, and part after, the anaphor.

INFERRED REFERENCE indicates that the referent does not appear (explicitly) in the text, and must, at least in part, be inferred by the hearer.

ANIMACY of referents in the text was coded as (1) human, (2) (non-human) animate, (3) inanimate physical objects, or (4) non-physical objects – such as concepts, propositions, statements, claims etc.

SEMANTIC INTERFERENCE refers to the situation where other potential referents of an anaphor intervene between the previous mention of a referent and the mention accomplished by that anaphor. For each example of *that*-anaphora in which there was an identifiable antecedent, it was recorded whether or not there was another semantically compatible, and thus potentially interfering, referent that either intervened between the occurrence of *that* and its antecedent, and/or occurred in the same discourse unit as the *that* anaphor (see § 3.7).

In the following sections, the ways in which the features recorded in the database were found to affect *that*-anaphora in the dinner text are discussed.

3.2. ANIMACY OF REFERENTS. A large percentage of personal pronouns in discourse refer to humans (*she, he, they* etc.), while references to humans with *that* are exceedingly rare. In the dinner text there were no references to human or other animate beings with *that*.⁹ This difference has at least one consequence for how anaphors appear in a text. Because human referents often play central roles (are highly topical) in conversational (and much written) discourse, personal pronouns commonly occur in extended sequences, with the same pronoun referring to the same referent repeatedly in a sequence of clauses over an relatively lengthy portion of discourse.

That, on the other hand, because it does not generally refer to human (or most animate) referents, is less likely to refer to something that is maintained as a topic or active participant in the events being described, and thus is also less likely to appear repeatedly in a string of contiguous clauses referring to the same referent.¹⁰ Thus the common notion that anaphora serves to 'track referents through' a discourse is somewhat less applicable in the case of *that*-anaphora.

A specific measure of the tendency of *that* to appear in this type of string of repeated references to the same referent, as opposed to the tendency of personal pronouns to do so, was not made. However, there were very few cases in which *that* referred to the same referent in more than two or three adjacent clauses. In contrast, there were a number of cases in which *that* was used to refer to one thing in one clause, something else in the next clause, and then back to the first or on to yet another referent in the following clause. For example, in ex. 1, the first instance of anaphoric *that* refers to the proposition that 'Longmont voted to be long distance from Denver.' In the second instance it refers to the inferred concept 'living where it is a local call to Denver'. In the next line, the third instance of *that* refers again to Longmont's voting to be long distance to Denver. In the next line, L pops back to the same reference she made in the previous line to 'living where it is a local call to Denver'. (Here, L and J seem to carrying on somewhat separate conversations, rather than cooperating in constructing a single one.) The next (5th) instance of anaphoric *that* again reverts back to referring to the vote, as it does in the 6th instance as well.

In all of the examples, the relevant instance of *that* appears in italics, the line in which it appears is marked with an arrow (⇒) in the left margin.

- (1) R: *It's just weird because Longmont is a lot closer than Arvada but Arvada is a local call and Longmont // is long distance*
 L: *Mhm, isn't that .. yeah*
 R: *and that's because Longmont*
 ⇒ M: *Longmont // voted for that years ago.*
 R: *Longmont voted on it sometime.*

distance is quite long – long enough that the use of a pronoun would not be expected solely on the basis of distance – and yet the use of a pronoun as anaphor is not at all problematic. See Givón 1983 for a comprehensive work on referential distance. Since the publication of this work, the various deficiencies of referential distance in accounts of anaphora have been recognized (e. g. by Fox 1987a).

⁹ A study of anaphoric use of it would be comparable to this study in this regard.

¹⁰ *That* also occurs less frequently in anaphoric chains than *it* (Passonneau [Schiffman] 1984), although this is not due to a difference in the range of potential referents of *that* as opposed to *it*, but rather has to do with the fact that *that* is often used to signal a change in focus, thus making it less suitable to a long sequence of references to the same entity (same topic).

(2) J: *Mhm, Yeah I know what it is it's 30 to 40 // dollars a month.*
 ⇒ R: *Is that*
 L: *(???) spent 38.*
 ⇒ R: *Is that on top of insurance or is that the insurance // cost?*
 ⇒ L: *That's that's the insurance.*
 ⇒ R: *The whole thing? That's cheap.*
 ⇒ L: *We'll that's the portion we pay of course.*

The referents of (*this* and) *that* lie within a near vs. far conceptual space. This space can be established by the physical context, in which case *that* (or *this*) can refer to concrete objects, either inanimates or animates. Or it can be established by the discourse and social context, in which case the referents of *that* (or *this*) may be to such NPOs as a statement or claim just made, an implication of something just stated, or some concept or idea being discussed or implied in the conversation. In the dinner text, it appears that this is a major function of the anaphoric use of *that*. A number of the referents of *that* are claims or statements made just previously by one of the participants in the conversation, as in examples 3, 4, and 5 below.¹²

- (3) R: *And .. join the army maybe and then you ..*
 ? : *go to school*
 [*studied .. for a some kind of job, and then you took a job.*
You know you notice how they used to say I took a took a job?
 L: [*Mhm,* [*((laugh))*
 R: *with a // (???) ?*

¹² In the corpus studied by Linde 1979, every instance of a reference to a preceding statement was accomplished using *that*.

PRONOUN	HUMAN		ANIMATE		INANIMATE		NON-PHYSICAL		TOTAL
	N	%	N	%	N	%	N	%	
<i>she</i>	34	100	0	0	0	0	0	0	34
<i>her</i>	12	100	0	0	0	0	0	0	12
<i>he</i>	38	100	0	0	0	0	0	0	38
<i>him</i>	6	100	0	0	0	0	0	0	6
<i>his</i>	4	100	0	0	0	0	0	0	4
Total <i>she...his</i>	94	100	0	0	0	0	0	0	94
<i>they</i>	96	85	11	10	6	5	0	0	113
<i>them</i>	10	67	0	0	2	13	3	20	15
<i>their</i>	17	100	0	0	0	0	0	0	17
<i>it</i>	0	0	1	1	97	51	93	49	292
<i>that</i>	0	0	0	0	25	18	116	82	141

TABLE 1. Animacy of pronoun anaphors.

- J: ((*laugh*))
 ⇒ R: *You don't say that anymore.*
- (4) J: *if I only used if it's single or something I used a doctor once a year that would be an arm and a leg.*
 (1.5)
 J: *But uh // (???)*
 L: *But your (???) been getting company insurance anyway and it would be at least that much.*
 J: *Yeah that's right and so but we with a child we go there all the time.*
- (5) J: *Mhm, Yeah I know what it is it's 30 to 40 // dollars a month.*
 R: *Is that*
 L: *(???) spent 38.*
 R: *Is that on top of insurance or is that the insurance // cost?*
 L: *That's that's the insurance.*
 R: *The whole thing? That's cheap.*
 L: *We//ll that's the portion we pay of course.*
 J: *Mhm,*
 R: *Yeah.*
 (1.0)
 ⇒ R: *Yeah // that's right.*
 M: *The company // (???)*
 L: *(???)*
 ⇒ M: *That's good. Of course that's a bigger company.*

3.3. TEXTUAL MENTION STATUS. As shown in Table 2, 19 of 141 (13%) of the occurrences of anaphoric *that* in the dinner text were references to textually new referents. Of these, 10 were to entities present in the physical context of the conversation. These cases appear to function as one would expect. The referent is textually new, but is available for pronominal reference by virtue of its presence in the physical context, as in 6 and 7.

TM STATUS	N	%
new	19	13
recent	73	52
distant	2	1
partial	47	33
Total	141	100

TABLE 2. Textual mention status of anaphoric *that*.

- (6) J: *We asked you twice not to do that.*
 C: *mm mm that.*
 ⇒ J: *You can eat that.*

- (7) C: (turning crank on toy car)
 ⇒ L: *Is that okay? Is he*
 M: *Yeah. It's okay.*
 M: *If he turns it he won't be able to turn the flywheel anymore.* ((laugh))

The remaining nine cases, in which *that* was used to refer to a textually new referent that was not present in the physical context, are the most interesting cases. In 8, M has asked R whether the tape recorder was turned on. This reminds everyone that the conversation is being recorded, resulting in concern about how people are talking. In the utterance in which *that* is used, R is trying to reassure everyone that the point of recording the conversation is not to catch people making grammar 'mistakes' by claiming that *linguists aren't like THAT*. But what does *that* refer to here? Its referent is obviously not in the immediate physical context. Nor is it simply some referent or statement that has appeared previously in the text. Rather, the referent of *that* is a proposition that must be inferred from what has been said: something like 'the idea that linguists are people who criticize other people's grammar and diction the way elementary school teachers tend to'.

- (8) M: *Did you turn it on?*
 C: *Meat.*
 R: *Turn what on?*
 C: *Yeah yeah // meat meat meat meat meat meat*
 M: *(laugh)*
 L: *We'll be glad*
 R: *Yeah.*
 (???)
 R: *Mhm, You're not supposed to think about it th//ough*
 C: *me//at meat meat meat meat m//eat meat meat meat meat meat meat meat*
 (laughter) [
 L: *I'd forgotten all about it .. re//ally*
 J: *(laugh)*
 J: *Yeah I'm sitting here wondering if my grammar and diction is uh uh where it should be //*
 (laugh)
 R: *Honestly though (???)*
 (???)
 [
 L: *We were living with Jim's mom .. in the .. being an elementary school teacher and everything.*
 (..) ⇒ R: *See linguists aren't like that at all.*
 L: *Well I bet that (-) th//ey're probably not but,*
 R: *They never get on your case.*
 L: *Elementary school teachers do.*
 J: *(laugh)*

In ex. 9, M would like to have some mustard passed to her. J realizes this, and asks if M wants some. In M's reply, she uses *that* to refer to something unstated, apparently that it is okay with her if J has some first, or if there is some other delay in getting the mustard to her. *That's alright* is a common, formulaic expression. But *that* still refers to something here, again a proposition that must be inferred from context. In 8, the discourse context (i.e. what has just recently been said) is the source for the inference. In 9, the inference is made based on the physical and especially social context.

- (9) M: *Let's see I li//ke uh Poupon*
 J: *C, what would you care for.*
 M: *Poupon.*
 J: *(???) Oh hey you want some*
 ⇒ M: *Yeah (???) but that's alright.*

In the entire dinner text, 63 (45%) of the 141 instances of *that*-anaphora referred to propositions or concepts that had to be inferred, at least in part. Cases in which the referent was only partially inferred are like those in example 10. Here the immediately preceding discourse (textual) context contains material that is clearly part of the referent, but not all of it. Rather, the referent is a complex proposition, one that cannot be fully associated with any single

word, phrase, or utterance in the text. In 10, the discussion has been about some horses that graze in a nearby field, and then it turns to an accident involving some other horses in which a woman in a pick-up truck hit some horses that had gotten loose.

- (10) R: *A couple horses got killed on 287*
 M: *Mhm,*
 R: *a few days ago.*
 ⇒ L: *I saw that. That poor woman. Gee.*

Here it is clear that the first statement by R is a part of the referent of *that* in the last line, but it does not constitute the entire referent. Rather, the full referent is a proposition that must, in part, be inferred. That is, the *that* does not refer to the statement itself (it doesn't mean the L saw R's statement), nor to the event itself (it doesn't mean that L saw the accident itself). Rather, what is meant is that L has seen a newspaper article or TV news report about the two horses getting killed on 287. In the entire text, there were 47 instances (33%) like this, in which the referent had a textual mention status of 'partial' (see Table 2). Thus, it appears to be relatively common for *that* to have a referent that cannot fully be accounted for from a purely text-based analysis. Rather, the identity of the referent can only be determined by taking into account the psychological process of inference. Only part of the information needed to determine the referent is available in the text itself.

3.4. FRAME ACTIVATION. One of the motivations for defining the feature TEXTUAL MENTION STATUS for this study was to distinguish between text- and process-oriented features in the analysis of discourse. As a result, much of what is captured in the common notion of INFORMATION STATUS was defined in purely textual terms, such as TEXTUAL MENTION STATUS. However, many definitions of information status also include some process-oriented features, such as 'activation,' which appear to play a role in anaphora. For example, Chafe 1987 provides a more detailed definition of the relatively informal information status terms OLD and NEW INFORMATION in terms of levels of cognitive activation such as ALREADY ACTIVE, PREVIOUSLY INACTIVE, and ACCESSIBLE/PREVIOUSLY SEMI-ACTIVE. In the psychological literature on anaphora as well, this sort of cognitive status of referents is used to analyze the use of anaphora. In order to examine the role of this sort of cognitive, process-oriented factor in the use of anaphoric *that*, the features FRAME ACTIVATION and ACCESS ACTIVATION were used.¹³ Prince 1981 uses the term EVOKED for this type of reference.

In a study of the role of frames, or scripts, in the processing of anaphora, Walker & Yekovich 1987 found that concepts more central to a script evoked during comprehension, though not mentioned explicitly, are more likely to be included in the representation of the text in memory, and consequently are more accessible (accessed more quickly) for the purpose of resolution of anaphora. Concepts that are brought into working memory, or 'activated', in this way should also be available for reference using such reduced anaphoric devices as pronouns, just as if they had been activated with an explicit reference. In order to see whether such references could be attributed to the activation of the referent through its role in a frame evoked by the material in the discourse, the feature FRAME-ACTIVATED was included in the database. Following the use of Walker & Yekovich, each occurrence of anaphoric *that* was evaluated as having a referent that was CENTRAL to an evoked frame, PERIPHERAL to such a frame, or NOT FRAME ACTIVATED.

Examples in which frame activation is the only means by which a referent is made available for anaphoric reference should appear in the text as anaphors with a referential distance of zero, a textual mention status of NEW or PARTIAL, and a frame activation of CENTRAL or PERIPHERAL. No clear examples of this sort of example were found in the dinner text. In all the cases where there was frame-activation of a referent, there was also some other possible source of activation, such as an explicit previous mention, a partial or implicit previous mention, or the physical context. For example, in the portion of the discourse shown below in 11, *that* refers to some food in the immediate physical context. Even though the food plays a central role in the 'eating dinner' frame, its availability for pronominal reference in this case cannot be solely attributed to the frame, since it appears in the physical context.

- (11) L: *You can drink it*
 |
 J: *We asked you twice not to do that*
 C: *mm mm that.*
 ⇒ J: *You can eat that.*

¹³ The inadequacy of a simple, text-based notion of old vs. new information, and the importance of examining language processing in the analysis of reference is also discussed by Givón 1988.

Perhaps part of the reason that the frame-activation feature did not prove particularly helpful in this analysis is that I was fairly conservative in attributing the presence of a frame to a portion of discourse in which a *that*-anaphor appeared. There were a total of only 30 of 141 appearances of anaphoric *that* in which a frame was determined to be present, and only nine of these were cases in which the referent was central to the frame. Of course, assigning this feature is a subjective judgement, but it seems likely that some sort of frame or set of frames is active for nearly all situations in which anaphora occurs. A more complete account of the role of frames in anaphora will require further work. It is one thing to think of a situation in which a frame clearly applies, and then construct a text around that frame. It is a much more tedious and difficult task to determine all of the frames that may be active at each point in a long, spontaneously produced discourse.

3.5. ACCESS ACTIVATION is another process-oriented feature. The referent of a *that* anaphor was said to be access-activated if it would be predicted to be present as an argument of a proposition in working memory at the point when the anaphoric *that* was encountered, according to the sort of model of working memory management during comprehension presented in van Dijk & Kintsch 1983 or Fletcher 1981. This feature was also assigned according to subjective judgement, rather than in a rigorous fashion – to have done so for the entire dinner text would have been impractical. As one would expect, nearly all of the cases in which access-activation was not predicted at the point of occurrence of an anaphoric *that*, the referential distance was zero and the textual mention status NEW, or the referential distance was negative (a case of cataphoric reference). As with frame-activation, a more complete account of the role of access-activation will require a great deal more work, including the testing of explicit models that predict when a referent should be active in working memory.

3.6. PARTIAL CATAPHORIC REFERENCE. In addition to referents that must be inferred and those that appear only partially in the text, there were a number of examples in which *that* referred in part cataphorically to following material. By 'partially cataphoric' is meant that, as in the examples above of 'partial' textual mention status, some of the propositional referent of an instance of *that* occurs in the preceding text, but some other part follows. Most of the instances of partially cataphoric reference combine each of three factors: (a) the referent of *that* appears only partially in the text (PARTIAL textual mention status), (b) the entire proposition can be determined only through an inference process, and (c) the *that* also refers, in part, cataphorically to immediately following material. Partially cataphoric reference can be viewed as a subtype of the sort of (partially) inferred reference discussed above. At the point at which a partially cataphoric expression appears, it must be processed by the hearer without any knowledge of the information contained in the upcoming mention of the referent, as it hasn't yet appeared. Thus the referent must be inferred, based only on what has appeared so far (and what the hearer knows about the context, the world etc.).

In ex.12, the conversation is about how L grew up on a farm in a small community in Iowa, and how R had once visited a college roommate in a small town nearby. When J uses *that* anaphorically, it refers in part to a complex proposition regarding what things are like on a farm in Iowa, but after *that* appears, J adds more to the meaning, and reference, of *that*.

- (12) L: *Well and it has it has the county seat square too so // (???) little towns*
 J: *Mhm, Mhm, it's real cute.*
 R: *And that's where you grew up in Iowa?*
 L: *Mhm,*
 (2.0)
 R: *How was it you were near um what is it Marshalltown? Was that it? // Is that (???)*
 L: *Yeah. It's not too far from Marshalltown.*
 (1.0)
 L: *You know somebody there?*
 R: *One of my .. college roommates .. was from a farm .. near Marshalltown.*
 L: *M//mm.*
 J: *Mmm. ((laugh)) That's what L's like .. right near a farm near some communities.*
 (1.0)
 ⇒ J: *I think that's what farming in America is like. You're just not too close to anything.*

In most of these examples, as in this case, one gets the impression that the speaker somehow felt, after using *that* anaphorically, that the reference may not have been entirely clear, and that a kind of clarification or restatement was in order. In this way, the speaker can 'repair,' or 'reconstruct,' the reference accomplished by an anaphoric

device.¹⁴ (See Schegloff 1979, Fox 1987c.)

In 13, the conversation is about how L&J's child C talks:

- (13) J: *It's kind of like we listen to C speak, a/nd he'll say um sit in the chair don't.*
 L: *Mhm, Mhm,*
 L: *((laugh))*
 J: *And we understand that that means he doesn't want to sit in the chair.*
 L: *Mh//m.*
 ⇒ J: *Just cause we know that's where he puts the negatives, // at the end.*

J's use of *that* is partially anaphoric, in that it refers to something that can be inferred from previous material (putting the negative at the end can be inferred from the example utterance *sit in the chair don't*), and partially cataphoric, referring to J's following statement *at the end*. Here again one gets the impression that *at the end* was added to clarify the reference of *that*.

Similarly, in 14 an anaphoric use of *that* refers in part to a proposition that can be more or less inferred from preceding material, but a following correction or addendum is made to modify or clarify the reference intended. Here, the talk is about the problem of communicating something in a written letter, without the advantage of face-to-face contact.

- (14) J: *There's more than just the written printed word. There's the body posture and then there's the tone?*
 R: *Right.*
 J: *You know that you can't read and uh (???) parts // and*
 R: *Mhm,*
 R: *(???)*
 ⇒ L: *That's one thing I notice when I write (???) as I in my mind I'm thinking the tone, then realize that they're not getting that at all.*

What does this *that* refer to? One could argue that it simply refers to J's statement in the first two lines. But it also refers cataphorically to what follows, where L redefines the reference accomplished by the *that*. In terms of the processing involved in producing and comprehending this type of reference, it appears that, at the point of uttering the *that*, the speaker has something in mind that s/he feels the hearer can at least partially identify (through inference), but that may need further elaboration or even correction. In fact, the speaker is likely monitoring the hearers' reactions (both verbal and non-verbal), and may adjust the amount and type of elaboration accordingly. For the hearer, partially cataphoric references are essentially of the same type as fully inferred references, in that inference processes must be used at the point at which a cataphor appears, as the upcoming mention of the referent is unavailable. It may be the case that, if the hearer can (immediately) recognize the referring expression as cataphoric, he or she is able to temporarily defer the assignment of a referent until the needed information is supplied by the speaker – holding an empty seat for the late-arriving referent, as it were. In this case, a cataphoric expression serves as a kind of promise by the speaker to fill in the information he or she expects the hearers will need to resolve the reference being made. Thus, while 'cataphoric reference' is originally a text-based term – indicating the simple notion of a reference forward in the text from a referring expression (the cataphor) to a subsequent explicit mention of the referent, it is possible to characterize the processing involved in an instance of cataphoric reference. The speaker produces the referring expression, thereby making a commitment to see to it that the hearers are able to resolve the intended reference. Resolution may already be possible, based on material in the preceding discourse and/or through the use of inference, or the speaker may need to provide elaboration or modification to facilitate resolution.

In ex.15 we have a case in which the speaker actually seems to change his mind about what he wants the *that* he has just used to refer to. The talk has been about different phone companies and different types of phone service in general, and WATS lines, or '800 service', in particular. J is a representative of US Sprint, and is explaining how '800 service' works. Several minutes earlier, the US Sprint 'phone card' had been discussed. J uses *that*, apparently to refer to using the phone card for '800 service'. But then he seems to realize that the phone card really isn't all that relevant to using '800 service' and redefines what *that* should be understood to refer to – the fact that a call goes over

¹⁴ Fox 1987a refers to this as 'replacement' of a sequence (pp. 72 ff.)

a fiber optic line. Thus we see the shared, interactive nature of the process of reference; the reference of a linguistic expression can be restructured, or altered, depending on the immediate and changing demands of the conversational situation.

- (15) J: *You hit you know 8 7 7 46 46 it's a Sprint line .. you know and it's uh you know 800 // 5 5 5 46 46 it's a .. AT&T line or something.*
 R: *Yeah.*
 R: *Must be any company could have any number.*
 J: *Mhm,*
 |
 R: *People always want something // like 1 800 call us or whatever.*
 L: *(??/?)*
 J: *Mhm, Mhm,*
 ⇒ J: *Well that's what makes the phone card so fun .. is that's a US Sprint .. 800 line so it's guar- it's on a fiber optic line.*

3.7. SEMANTIC CONSTRAINTS AND INTERFERENCE. One of the factors that may affect the use of an anaphoric expression at a given point in a discourse is the degree of 'interference' from other potential referents that intervene between one mention of a referent and the previous mention. In personal pronoun anaphora, this situation may arise when multiple same-gender referents are present together in a discourse. In Fox's 1987a account of personal pronoun anaphora in conversational discourse it is argued that when structural factors (such as boundaries between different discourse units) cannot disambiguate between the potential referents of a pronoun, 'semantic plausibility' may come into play. Based on the patterns found in the dinner text, it appears that semantic constraints play a somewhat larger role in non-personal pronoun anaphora. Each instance of a *that*-anaphor in the text was coded for whether a semantically compatible (potentially interfering) referent intervened between the occurrence of that and its antecedent (SEQUENTIAL INTERFERENCE) or occurred in the same discourse unit as the *that*-anaphor (SAME-UNIT INTERFERENCE); see Table 3. A potentially interfering referent was found to intervene sequentially between antecedent and anaphor in 62 (44%) of the occurrences of anaphoric *that*, and to occur within the same discourse unit in 96 (68%) of the occurrences. However, in 31 (22%) of the sequential cases, and 47 (33%) of the same-unit cases, semantic constraints (incompatibilities between the true referent and its 'competitors') were found to contribute to disambiguating the reference. Thus it appears, at least in the case of *that*-anaphora, that semantic constraints commonly play at least a partial role in disambiguating anaphoric reference.

	SEQUENTIAL INTERFERENCE		SAME-UNIT INTERFERENCE	
	N	%	N	%
+	31	22	49	35
-	79	56	45	32
+ with semantic disambiguation	<u>31</u>	<u>22</u>	<u>47</u>	<u>33</u>
Total	141	100	141	100

TABLE 3. Potential referential interference from other referents.

4. CONCLUSIONS. The goal of this study was to examine a naturally produced conversational English discourse in order to contribute to an account of the anaphoric use of the demonstrative pronoun *that* which takes into account the processes involved in producing and understanding acts of reference. It was found that anaphoric *that* patterns differently from personal pronoun anaphora in several interesting ways:

(a) Most of the referents of anaphoric *that* are non-physical objects (in terms of their animacy). This is in contrast to *it*, which, though similar in its lexical semantic properties, is used to refer to NPOs much less frequently. This is likely due to the fact that *that* functions as a deictic. In addition, anaphoric *that* serves less of a 'participant tracking' function than anaphoric personal pronouns, which may also be due to the differences in animacy between typical referents of *that* and those of the personal pronouns.

(b) Many of the referents of anaphoric *that* must be inferred; they are either not present, or are only partially present in the text. Only an account of anaphora that takes into consideration the process of inference – the process by which the hearer constructs an understanding (representation) of the referent – can provide an explanation for these instances of *that*-anaphora.

(c) Some of the inferred referents are partially cataphoric. That is, the referent is a complex proposition, part of which is present in, or can be inferred from, the preceding discourse, and part of which follows the use of the *that* anaphor.

(d) Some of the cases of partially cataphoric reference involve a process of reconstruction, or redefinition, of the

reference of the anaphor, in which the speaker revises the identity of the referent after the use of the anaphor. A full account of this phenomenon must include a description of the processes by which a speaker decides that the reference of an anaphor is somehow inadequate and should be revised, how this revision is accomplished, and how hearers reconstruct their understanding of the reference.

These differences suggest that a purely text-oriented analysis of anaphora is inadequate. Rather, the processes the discourse participants engage in to produce and comprehend anaphora in the discourse must be taken into consideration in a complete account, an account that explicates both the product (the text), and the processes by which it was produced.

REFERENCES

- BOSCH, PETER. 1988. Representing and accessing focused referents. *Language and Cognitive Processes* 3.165-274.
- CHAFE, WALLACE. 1987. Cognitive constraints on information flow. In Tomlin 1987:21-51.
- CLARK, HERBERT, & C. J. SENGUL. 1979. In search of referents for nouns and pronouns. *Memory & Cognition* 7.35-41.
- DELL, G.; G. MCKOON, G.; & R. RATCLIFF. 1983. The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior* 22.121-32.
- DU BOIS, JOHN. 1987. The discourse basis of ergativity. *Language* 63.805-55.
- FLETCHER, C. 1981. Short-term memory processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior* 20.564-74.
- FOX, BARBARA. 1987a. Discourse structure and anaphora: Written and conversational English. Cambridge: University Press.
- . 1987b. Anaphora in popular written English narratives. In Tomlin 1987:157-174.
- . 1987c. Interactional reconstruction in real-time language processing. *Cognitive Science* 11.365-87.
- GARROD, S., and A. SANFORD. 1983. Topic dependent effects in language processing. The process of language understanding, ed. by Giovanni B. Flores d'Arcais & Robert J. Jarvella (eds.), 271-295. New York: Wiley.
- GIVÓN, TALMY. 1979 (ed.) *Discourse and syntax* (Syntax and semantics, 12). New York: Academic Press.
- . 1983. Introduction. Topic continuity in discourse: A quantitative cross-language study, ed. by Talmy Givón, 1-42. Amsterdam: Benjamins.
- . 1988. The grammar of referential coherence as mental processing instructions (MS).
- . 1989. *Mind, code and context: Essays in pragmatics*. Hillsdale, NJ: Erlbaum.
- GROSZ, BARBARA. 1977. The representation and use of focus in dialogue understanding. Berkeley: University of California dissertation.
- , and CANDICE SIDNER. 1985. The structures of discourse structure. (CSLI Technical Report CSLI-85-39). Stanford, CA: Stanford University.
- LEVINSON, STEPHEN. 1983. *Pragmatics*. Cambridge: University Press.
- LINDE, CHARLOTTE. 1979. Focus of attention and the choice of pronouns in discourse. In Givón 1979:337-54.
- MARSLÉN-WILSON, WILLIAM D., & LORRAINE K. TYLER. 1980. The temporal structure of spoken language understanding. *Cognition* 8.1-71.
- MARSLÉN-WILSON, WILLIAM D.; E. LEVY; and LORRAINE K. TYLER. 1982. Producing interpretable discourse: the establishment and maintenance of reference. *Speech, place and action*, ed. by Robert J. Jarvella & Wolfgang Klein, 339-378. New York: Wiley.
- MCKOON, G., & R. RATCLIFF. 1980. The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior* 19.668-82.
- PASSONNEAU (SCHIFFMAN), REBECCA. 1984. The two nominal anaphors it and that. *Chicago Linguistic Society* 20.344-57.
- . 1989. Getting at discourse referents. *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 51-59.
- PRINCE, ELLEN. 1981. Toward a taxonomy of given/new information. *Radical pragmatics*, ed. by Peter Cole, 223-55. New York: Academic Press.
- SCHEGLOFF, EMANUEL. 1979. The relevance of repair to syntax-for-conversation. In Givón 1979:261-86.
- . 1981. Discourse as an interactional achievement. *Analyzing discourse: Text and talk* (GURT 1981), ed. by Deborah Tannen, 71-93. Washington, DC: Georgetown University Press.
- TOMLIN, RUSSELL. 1987 (ed.) *Coherence and grounding in discourse*. Amsterdam: Benjamins.
- TYLER, LORRAINE K., and WILLIAM D. MARSLÉN-WILSON. 1982. The resolution of discourse anaphors: Some on-line studies. *Text* 2.263-91.
- VAN DIJK, TEUN, and WALTER KINTSCH. 1983. *Strategies of discourse comprehension*. New York: Academic Press.
- WALKER, C. H., and F. R. YEKOVICH. 1987. Activation and use of script-based antecedents in anaphoric reference. *Journal of Memory and Language* 26.673-91.