Anaphora and Discourse Structure

Bonnie Webber[*]
Edinburgh University

Matthew Stone[†]
Rutgers University

Aravind Joshi[‡]
University of Pennsylvania

Alistair Knott[§]
University of Otago

We argue in this article that many common adverbial phrases generally taken to signal a discourse relation between syntactically connected units within discourse structure instead work anaphorically to contribute relational meaning, with only indirect dependence on discourse structure. This allows a simpler discourse structure to provide scaffolding for compositional semantics and reveals multiple ways in which the relational meaning conveyed by adverbial connectives can interact with that associated with discourse structure. We conclude by sketching out a lexicalized grammar for discourse that facilitates discourse interpretation as a product of compositional rules, anaphor resolution, and inference.

1. Introduction

It is a truism that a text means more than the sum of its component sentences. One source of additional meaning are **relations** taken to hold between adjacent sentences "syntactically" connected within a larger discourse structure. It has been very difficult, however, to say what discourse relations there are, either theoretically (Mann and Thompson 1988; Kehler 2002; Asher and Lascarides 2003) or empirically (Knott 1996).

Knott's empirical attempt to identify and characterize **cue phrases** as evidence for **discourse relations** illustrates some of the difficulties. Knott used the following theory-neutral test to identify cue phrases: For a potential cue phrase ϕ in naturally occurring text, consider in isolation the clause in which it appears. If the clause appears incomplete without an adjacent left context, whereas it appears complete if ϕ is removed, then ϕ is a cue phrase. Knott's test produced a nonexhaustive list of about two hundred different phrases from 226 pages of text. He then attempted to characterize the discourse relation(s) conveyed by each phrase by identifying when (always, sometimes, never) one phrase could substitute for another in a way that preserved meaning. He showed how these substitution patterns could be a consequence of a set of semantic features and their values. Roughly speaking, one cue phrase could *always* substitute for another if it had the same set of features and values, *sometimes* do so if it was less specific than the other in terms of its feature values, and *never* do so if their values conflicted for one or more features.

---

[*] School of Informatics, University of Edinburgh, 2 Buccleuch Place, Edinburgh, EH8 9LW, UK. E-mail: bonnie@inf.ed.ac.uk.
[†] Department of Computer Science, Rutgers Universtiy, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019. E-mail: mdstone@cs.rutgers.edu.
[‡] Department of Computer & Information Science, University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104-6389. E-mail: joshi@linc.cis.upenn.edu.
[§] Department of Computer Science, University of Otago, P.O. Box 56, DUNEDIN 9015, New Zealand. E-mail: alik@cs.otago.ac.nz.

By assuming that cue phrases contribute meaning in a uniform way, Knott was led to a set of surprisingly complex directed acyclic graphs relating cue phrases in terms of features and their values, each graph loosely corresponding to some family of discourse relations. But what if the relational meaning conveyed by cue phrases could in fact interact with discourse meaning in multiple ways? Then Knott's substitution patterns among cue phrases may have reflected these complex interactions, as well as the meanings of individual cue phrases themselves.

This article argues that cue phrases *do* depend on another mechanism for conveying extrasentential meaning—specifically, **anaphora**. One early hint that adverbial cue phrases (called here **discourse connectives**) might be anaphoric can be found in an ACL workshop paper in which Janyce Wiebe (1993) used the following example to question the adequacy of tree structures for discourse:

(1) a. The car was finally coming toward him.
     b. He [Chee] finished his diagnostic tests,
     c. feeling relief.
     d. *But then* the car started to turn right.

The problem Wiebe noted was that the discourse connectives *but* and *then* appear to link clause (1d) to two different things: *then* to clause (1b) in a **sequence** relation (i.e., the car's starting to turn right being the next relevant event after Chee's finishing his tests) and *but* to a grouping of clauses (1a) and (1c) (i.e., reporting a contrast between, on the one hand, Chee's attitude toward the car coming toward him and his feeling of relief and, on the other hand, his seeing the car turning right). (Wiebe doesn't give a name to the relation she posits between (1d) and the grouping of (1a) and (1c), but it appears to be some form of **contrast**.)

If these relations are taken to be the basis for discourse structure, some possible discourse structures for this example are given in Figure 1. Such structures might seem advantageous in allowing the semantics of the example to be computed directly by compositional rules and defeasible inference. However, both structures are directed acyclic graphs (DAGs), with acyclicity the only constraint on what nodes can be connected. Viewed syntactically, arbitrary DAGs are completely unconstrained systems. They substantially complicate interpretive rules for discourse, in order for those rules to account for the relative scope of unrelated operators and the contribution of syntactic nodes with arbitrarily many parents.[1]

We are not committed to trees as the limiting case of discourse structure. For example, we agree, by and large, with the analysis that Bateman (1999) gives of

(2) (vi) The first to do that were the German jewellers, (vii) in particular Klaus Burie. (viii) And Morris followed very quickly after, (ix) using a lacquetry technique to make the brooch, (x) and using acrylics, (xi) and exploring the use of colour, (xii) and colour is another thing that was new at that time.

---

1 A reviewer has suggested an alternative analysis of (1) in which clause (1a) is **elaborated** by clause (1b), which is in turn elaborated by (1c), and clause (1d) stands in both a sequence relation and a contrast relation to the segment as a whole. Although this might address Wiebe's problem, the result is still a DAG, and such a fix will not address the additional examples we present in section 2, in which a purely structural account still requires DAGs with crossing arcs.
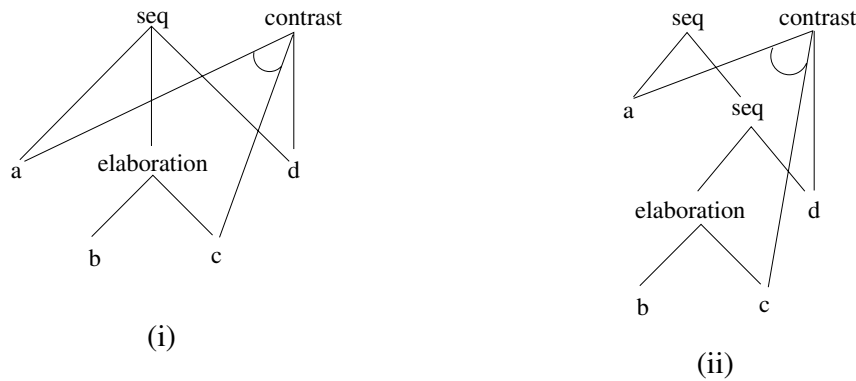
(i)

(ii)

Figure 1
Possible discourse structure for example (1). Each root and internal node is labeled by the type of relation that Wiebe takes to hold between the daughters of that node. (i) uses an *n*-ary branching sequence relation, whereas in (ii), sequence is binary branching.
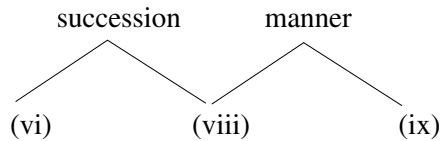

Figure 2
Simple multiparent structure.

in which clause (ix) stands in a **manner** relation with clause (viii), which in turn stands in a **succession** (i.e., sequence) relation with clause (vi). This is illustrated in Figure 2, which shows a DAG (rather than a tree), but without crossing dependencies.

So it is the cost of moving to *arbitrary* DAGs for discourse structure that we feel is too great to be taken lightly. This is what has led us to look for another explanation for these and other examples of apparent complex and crossing dependencies in discourse.

The position we argue for in this article, is that whereas adjacency and explicit conjunction (coordinating conjunctions such as *and, or, so*, and *but*; subordinating conjunctions such as *although, whereas*, and *when)* imply discourse relations between (the interpretation of) adjacent or conjoined discourse units, **discourse adverbials** such as *then, otherwise, nevertheless*, and *instead* are **anaphors**, signaling a relation between the interpretation of their matrix clause and an entity in or derived from the discourse context. This position has four advantages:

1. Understanding discourse adverbials as anaphors recognizes their **behavioral similarity** to the pronouns and definite noun phrases (NPs) that are the bread and butter of previous work on anaphora. This is discussed in section 2.

2. By understanding and exploring the full range of phenomena for which an anaphoric account is appropriate, we can better characterize anaphors and devise more accurate algorithms for resolving them. This is explored in section 3.

3. Any theory of discourse must still provide an account of how a sequence of adjacent discourse units (clauses, sentences, and the larger units that they can comprise) means more than just the sum of its component

units. This is a goal that researchers have been pursuing for some time, using both **compositional** rules and defeasible inference to determine these additional aspects of meaning (Asher and Lascarides 1999; Gardent 1997; Hobbs et al. 1993; Kehler 2002; Polanyi and van den Berg 1996; Scha and Polanyi 1988; Schilder 1997a, 1997b; van den Berg 1996) If that portion of discourse semantics that can be handled by mechanisms already needed for resolving other forms of anaphora and deixis is factored out, there is less need to stretch and possibly distort compositional rules and defeasible inference to handle everything.[2] Moreover, recognizing the possibility of two separate relations (one derived anaphorically and one associated with adjacency and/or a structural connective) admits additional richness to discourse semantics. Both points are discussed further in section 4.

4.   Understanding discourse adverbials as anaphors allows us to see more clearly how a lexicalized approach to the computation of clausal syntax and semantics extends naturally to the computation of discourse syntax and semantics, providing a single syntactic and semantic matrix with which to associate speaker intentions and other aspects of pragmatics (section 5.)

The account we provide here is meant to be compatible with current approaches to discourse semantics such as DRT (Kamp and Reyle 1993; van Eijck and Kamp 1997), dynamic semantics (Stokhof and Groenendijk 1999), and even SDRT (Asher 1993; Asher and Lascarides 2003), understood as a representational scheme rather than an interpretive mechanism. It is also meant to be compatible with more detailed analyses of the meaning and use of individual discourse adverbials, such as Jayes and Rossari (1998a, 1998b) and Traugott, (1995, 1997). It provides what we believe to be a more coherent account of *how* discourse meaning is computed, rather than an alternative account of *what* that meaning is or *what speaker intentions* it is being used to achieve.

2. Discourse Adverbials as Anaphors

2.1 Discourse Adverbials Do Not Behave like Structural Connectives
We take the building blocks of the most basic level of discourse structure to be explicit **structural connectives** between adjacent discourse units (i.e., coordinating and subordinating conjunctions and "paired" conjunctions such as *not only . . . but also*, and *on the one hand . . . on the other (hand)* and **inferred relations** between adjacent discourse units (in the absense of an explicit structural connective). Here, adjacency is what triggers the inference. Consider the following example:

(3) You shouldn't trust John. He never returns what he borrows.

Adjacency leads the hearer to hypothesize that a discourse relation of something like **explanation** holds between the two clauses. Placing the subordinate conjunction (structural connective) *because* between the two clauses provides more evidence for this rela-

---

2 There is an analogous situation at the sentence level, where the relationship between syntactic structure and compositional semantics is simplified by factoring away intersentential anaphoric relations. Here the factorization is so obvious that one does not even think about any other possibility.

tion. Our goal in this section is to convince the reader that many **discourse adverbials**, including *then, also, otherwise, nevertheless*, and *instead*, do not behave in this way.

Structural connectives and discourse adverbials do have one thing in common: Like verbs, they can both be seen as heading a predicate-argument construction; unlike verbs, their arguments are independent clauses. For example, both the subordinate conjunction *after* and the adverbial *then* (in its temporal sense) can be seen as binary predicates (e.g., sequence) whose arguments are clausally derived events, with the earlier event in first position and the succeeding event in second.

But that is the only thing that discourse adverbials and structural connectives have in common. As we have pointed out in earlier papers (Webber, Knott, and Joshi 2001; Webber et al., 1999a, 1999b), structural connectives have two relevant properties: (1) they admit **stretching** of predicate-argument dependencies; and (2) they do not admit **crossing** of those dependencies. This is most obvious in the case of preposed subordinate conjunctions (example (4)) or "paired" coordinate conjunctions (example (5)). With such connectives, the initial predicate signals that its two arguments will follow.

(4) *Although* John is generous, he is hard to find.

(5) *On the one hand*, Fred likes beans. *On the other hand*, he's allergic to them.

Like verbs, structural connectives allow the distance between the predicate and its arguments to be "stretched" over embedded material, without loss of the dependency between them. For the verb *like* and an object argument *apples*, such stretching without loss of dependency is illustrated in example (6b).

(6) a. Apples John likes.

     b. Apples Bill thinks he heard Fred say John likes.

That this also happens with structural connectives and their arguments is illustrated in example (7) (in which the first clause of example (4) is elaborated by another preposed subordinate-main clause construction embedded within it) and in example (8) (in which the first conjunct of example (5) is elaborated by another paired-conjunction construction embedded within it). Possible discourse structures for these examples are given in Figure 3.

(7) a. Although John is very generous—
    b. if you need some money,
    c. you only have to ask him for it—
    d. he's very hard to find.

(8) a. On the one hand, Fred likes beans.
    b. Not only does he eat them for dinner.
    c. But he also eats them for breakfast and snacks.
    d. On the other hand, he's allergic to them.

But, as already noted, structural connectives do not admit crossing of predicate-argument dependencies. If we admit crossing dependencies in examples (7) and (8), we get

(9) a. Although John is very generous—
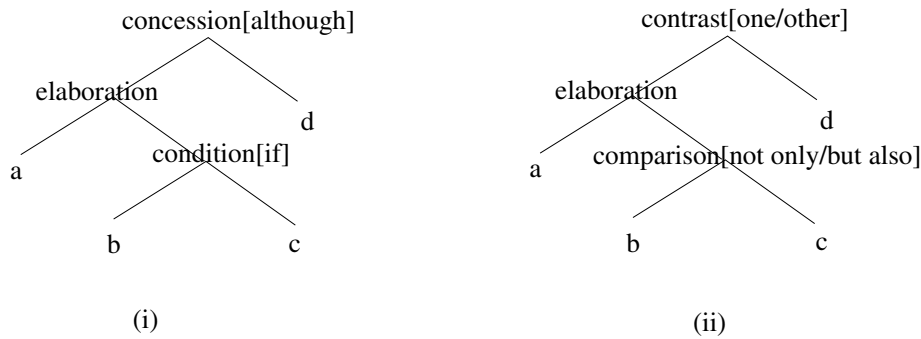   b. if you need some money—

Figure 3
Discourse structures associated with (i) example (7) and (ii) Example (8).
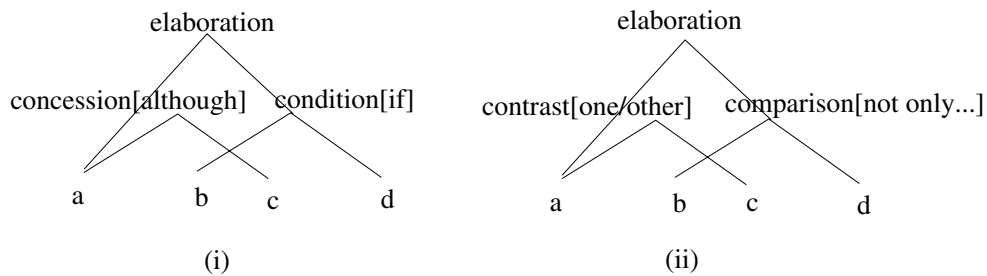

Figure 4
(Impossible) discourse structures that would have to be associated with (i) Example (9) and
(ii) example (10).

       c. he's very hard to find—
       d. you only have to ask him for it.

(10) a. On the one hand, Fred likes beans.
       b. Not only does he eat them for dinner.
       c. On the other hand, he's allergic to them.
       d. But he also eats them for breakfast and snacks.

Possible discourse structures for these (impossible) discourses are given in Figure 4.
Even if the reader finds no problem with these crossed versions, they clearly do not
mean the same thing as their uncrossed counterparts: In (10), *but* now appears to link
(10d) with (10c), conveying that despite being allergic to beans, Fred eats them for
breakfast and snacks. And although this might be inferred from (8), it is certainly not
conveyed directly. As a consequence, we stipulate that structural connectives do not
admit crossing of their predicate-argument dependencies.[3]

    That is not all. Since we take the basic level of discourse structure to be a conse-
quence of (1) relations associated with explicit structural connectives and (2) relations

---

  3 A reviewer has asked how much stretching is possible in discourse without losing its thread or having
    to rephrase later material in light of the intervening material. One could ask a similar question about
    the apparently unbounded dependencies of sentence-level syntax, which inattentive speakers are prone
    to lose track of and "fracture." Neither question seems answerable on theoretical grounds alone, with
    both demanding substantial amounts of empirical data from both written and spoken discourse. The
    point we are trying to make is simply that there is a difference in discourse between any amount of
    stretching and even the smallest amount of crossing.

whose defeasible inference is triggered by adjacency, we stipulate that *discourse structure itself does not admit crossing structural dependencies*. (In this sense, discourse structure may be truly simpler than sentence structure. To verify this, one might examine the discourse structure of languages such as Dutch that allow crossing dependencies in sentence-level syntax. Initial cursory examination does not give any evidence of crossing dependencies in Dutch discourse.)

If we now consider the corresponding properties of discourse adverbials, we see that they do admit crossing of predicate-argument dependencies, as shown in examples (11)–(13).

(11) a. John loves Barolo.
      b. So he ordered three cases of the '97.
      c. But he had to cancel the order
      d. because *then* he discovered he was broke.

(12) a. High heels are fine for going to the theater.
      b. But wear comfortable shoes
      c. if *instead* you plan to go to the zoo.

(13) a. Because Fred is ill
      b. you will have to stay home
      c. whereas *otherwise* the two of you could have gone to the zoo.

Consider first the discourse adverbial *then* in clause (11d). For it to get its first argument from (11b) (i.e., the event that the discovery in (d) is "after"), it must cross the structural connection between clauses (c) and (d) associated with *because)*. This crossing dependency is illustrated in Figure 5(i). Now consider the discourse adverbial *instead)* in clause (12c). For it to get its first argument from (12a) (i.e., going to the zoo is an alternative to going to the theater), it must cross the structural connection between clauses (12b) and (12c) associated with *if*. This crossing dependency is illustrated in Figure 5(ii). Example (13) is its mirror image: For the discourse adverbial *otherwise* in (13c) to get its first argument from (13a) (i.e., alternatives to the state/condition of Fred being ill), it must cross the structural connection associated with *because*. This is illustrated in Figure 5(iii).

Crossing dependencies are not unusual in discourse when one considers anaphora (e.g., pronouns and definite NPs), as for example in
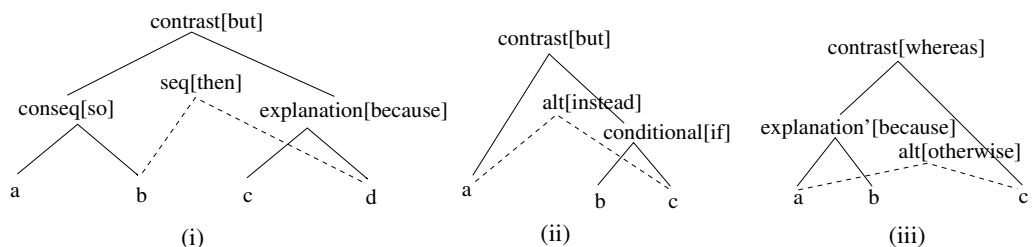

Figure 5
Discourse structures for examples (11)–(13). Structural dependencies are indicated by solid lines and dependencies associate with discourse adverbials are indicated by broken lines. (*explanation'* is the inverse of *explanation*—i.e., with its arguments in reverse order. Such relations are used to maintain the given linear order of clauses.)

(14) Every man$_i$ tells every woman$_j$ he$_i$ meets that she$_j$ reminds him$_i$ of his$_i$ mother.

(15) Sue$_i$ drives an Alfa Romeo. She$_i$ drives too fast. Mary$_j$ races her$_i$ on weekends. She$_j$ often beats her$_i$. (Strube 1998)

This suggests that in examples (11)–(13), the relationship between the discourse adverbial and its (initial) argument from the previous discourse might usefully be taken to be anaphoric as well.[4]

2.2 Discourse Adverbials Do Behave like Anaphors
There is additional evidence to suggest that *otherwise, then*, and other discourse adverbials are anaphors. First, anaphors in the form of definite and demonstrative NPs can take implicit material as their referents. For example, in

(16) Stack five blocks on top of one another. Now close your eyes and try knocking {*the tower*, *this tower*} over with your nose.

both NPs refer to the structure which is the implicit result of the block stacking. (Further discussion of such examples can be found in Isard [1975]; Dale [1992]; and Webber and Baldwin [1992].) The same is true of discourse adverbials. In

(17) Do you want an apple? *Otherwise* you can have a pear.

the situation in which you can have a pear is one in which you don't want an apple—that is, one in which your answer to the question is "no." But this answer isn't there structurally: It is only inferred. Although it appears natural to resolve an anaphor to an inferred entity, it would be much more difficult to establish such links through purely structural connections: To do so would require complex transformations that introduce invisible elements into discourse syntax with no deeper motivation. For example, in (17), we would need a rule that takes a discourse unit consisting solely of a yes/no question *P*? and replaces it with a complex segment consisting of *P*? and the clause *it is possible that P*, with the two related by something like elaboration. Then and only then could we account for the interpretation of the subsequent *otherwise* structurally, by a syntactic link to the covert material (i.e., to the possibility that *P* holds, which *otherwise* introduces an alterative to).

Second, discourse adverbials have a wider range of options with respect to their initial argument than do structural connectives (i.e., coordinating and subordinating conjunctions). The latter are constrained to linking a discourse unit on the **right frontier** of the evolving discourse (i.e., the clause, sentence and larger discourse units to its immediate left). Discourse adverbials are not so constrained. To see this, consider the following example:

---

4 We are aware that crossing examples such as (11)—(13) are rare in naturally occurring discourse. We believe that this is because they are only possible when, as here, strong constraints from the discourse adverbial and from context prevent the adverbial from relating to the closest (leftmost) eventuality or an eventuality coerced from that one. But rarity doesn't necessarily mean ill-formedness or marginality, as readers can see for themselves if they use Google to search the Web for strings such as *because then, if instead*, and *whereas otherwise*, and consider (1) whether the hundreds, even thousands, of texts in which these strings occur are ill-formed, and (2) what *then, instead,* and *otherwise* are relating in these texts. One must look at rare events if one is studying complex linguistic phenomena in detail. Thus it is not the case that only common things in language are real or worth further study.

(18) If the light is red, stop. *Otherwise* you'll get a ticket.
(*If you do something other than stop, you'll get a ticket.*)

This can be paraphrased using the conjunction *or*:

If the light is red, stop, or you'll get a ticket.

Here *or* links its right argument to a unit on the right frontier of the evolving discourse—in this case, the clause *stop* on its immediate left. Now consider the related example

(19) If the light is red, stop. *Otherwise* go straight on.
(*If the light is not red, go straight on.*)

This cannot be paraphrased with *or*, as in

(20) If the light is red, stop, or go straight on.

even though both *stop* and *If the light is red, stop* are on the right frontier of the evolving discourse structure. This is because *otherwise* is accessing something else, so that (20) means something quite different from either (18) or (19). What *otherwise* is accessing, which *or* cannot, is the interpretation of the condition alone.[5] Thus discourse adverbials, like other anaphors, have access to material that is not available to structural connectives.

Finally, discourse adverbials, like other anaphors, may require semantic representations in which their arguments are **bound variables** ranging over discourse entities. That is, whereas it might be possible to represent *Although P, Q* using a binary modal operator

(21) *although(p, q)*

where formulas *p* and *q* translate the sentences *P* and *Q* that *although* combines, we cannot represent *P . . . Nevertheless, Q* this way. We need something more like

(22) $p \wedge nevertheless(e, q)$

The motivation for the variable *e* in this representation is that discourse adverbials, like pronouns, can appear intrasententially in an analog of donkey sentences. Donkey sentences such as example (23) are a special kind of bound-variable reading:

(23) Every farmer who owns a donkey feeds *it* rutabagas.

In donkey sentences, anaphors are interpreted as covarying with their antecedents: The *it* that is being fed in (23) varies with the farmer who feeds it. These anaphors, however, appear in a structural and interpretive environment in which a direct syntactic relationship between anaphor and antecedent is normally impossible, so they cannot be a reflex of true binding in the syntax-semantics interface. Rather, donkey sentences show that *discourse* semantics has to provide variables to translate pronouns,

---

5 This was independently pointed out by several people when this work was presented at ESSLLI'01 in Helsinki in August 2001. The authors would like to thank Natalia Modjeska, Lauri Karttunen, Mark Steedman, Robin Cooper, and David Traum for bringing it to their attention.

and that *discourse* mechanisms must interpret these variables as bound—even though the pronouns appear "free" by syntactic criteria.

Thus, it is significant that discourse adverbials can appear in their own version of donkey sentences, as in

(24) a. Anyone who has developed innovative new software has *then* had to hire a lawyer to protect his/her interests. (i.e., *after developing innovative new software*)

   b. Several people who have developed innovative new software have *nevertheless* failed to profit from it. (i.e., *despite having developed innovative new software*)

   c. Every person selling "The Big Issue" might *otherwise* be asking for spare change. (i.e., *if he/she weren't selling "The Big Issue"*)

The examples in (24) involve binding in the interpretation of discourse adverbials. In (24a), the temporal use of *then* locates each hiring event after the corresponding software development. Likewise in (24b), the adversative use of *nevertheless* signals each developer's unexpected failure to turn the corresponding profit. And in (24c), *otherwise* envisions each person's begging if that person weren't selling "The Big Issue".

Such bound interpretations require variables in the semantic representations and alternative values for them in some model—hence the representation given in (22). Indeed, it is clear that the binding here has to be the discourse kind, not the syntactic kind, for the same reason as in (23), although we cannot imagine anyone arguing otherwise, since discourse adverbials have always been treated as elements of discourse interpretation. So the variables must be the discourse variables usually used to translate other kinds of discourse anaphors.[6]

These arguments have been directed at the behavioral similarity between discourse adverbials and what we normally take to be discourse anaphors. But this isn't the only reason to recognize discourse adverbials as anaphors: In the next section, we suggest a framework for anaphora that is broad enough to include discourse adverbials as well as definite and demonstrative pronouns and NPs, along with other discourse phenomena that have been argued to be anaphoric, such as VP ellipsis (Hardt 1992, 1999; Kehler 2002), tense (Partee 1984; Webber 1988) and modality (Kibble 1995; Frank and Kamp 1997; Stone and Hardt 1999).

3. A Framework for Anaphora

Here we show how only a single extension to a general framework for discourse anaphora is needed to cover discourse adverbials. The general framework is presented in Section 3.1, and the extension in Section 3.2.

3.1 Discourse Referents and Anaphor Interpretation
The simplest discourse anaphors are **coreferential**: definite pronouns and definite NPs that denote one (or more) discourse referents in focus within the current discourse

---

6 Although rhetorical structure theory (RST) (Mann and Thompson 1998) was developed as an account of the relation between adjacent units within a text, Marcu's guide to RST annotation (Marcu 1999) has added an "embedded" version of each RST relation in order to handle examples such as (24). Although this importantly recognizes that material in an embedded clause can bear a semantic relation to its matrix clause, it does not contribute to understanding the nature of the phenomenon.

context. (Under *coreference* we include **split reference**, in which a plural anaphor such as *the companies* denotes all the separately mentioned companies in focus within the discourse context.) Much has been written about the factors affecting what discourse referents are taken to be in focus. For a recent review by Andrew Kehler, see chapter 18 of Jurafsky and Martin (2000). For the effect of different types of quantifiers on discourse referents and focus, see Kibble (1995).

Somewhat more complex than coreference is **indirect anaphora** (Hellman and Fraurud 1996) (also called **partial anaphora** [Luperfoy 1992], **textual ellipsis** [Hahn, Markert, and Strube 1996], **associative anaphora** [Cosse 1996] **bridging anaphora** [Clark 1975; Clark and Marshall 1981; Not, Tovena, and Zancanaro 1999], and **inferrables** [Prince 1992]), in which the anaphor (usually a definite NP) denotes a discourse referent *associated with* one (or more) discourse referents in the current discourse context; for example,

(25) Myra darted to *a phone* and picked up *the receiver*.

Here *the receiver* denotes the receiver associated with (by virtue of being *part of*) the already-mentioned phone Myra darted to.

Coreference and indirect anaphora can be uniformly modeled by saying that the discourse referent e_α denoted by an anaphoric expression α is either equal to or associated with an existing discourse referent e_r, that is, $e_\alpha = e_r$ or $e_\alpha \in \mathrm{assoc}(e_r)$. But coreference and associative anaphora do not exhaust the space of constructs that derive all or part of their sense from the discourse context and are thus anaphoric. Consider "other NPs" (Bierner 2001a; Bierner and Webber 2000; Modjeska 2001, 2002), as in:

(26) Sue grabbed one phone, as Tom darted to *the other phone*.

Although "other NPs" are clearly anaphoric, should the referent of *the other phone* (e_α)—the phone other than the one Sue grabbed (e_r)—simply be considered a case of $e_\alpha \in \mathrm{assoc}(e_r)$? Here are two reasons why they should not.

First, in all cases of associative anaphora discussed in the literature, possible associations have depended only on the antecedent e_r and not on the anaphor. For example, only antecedents that have parts participate in whole-part associations (e.g., phone $\rightarrow$ receiver). Only antecedents with functional schemata participate in schema-based associations (e.g., lock $\rightarrow$ key). In (26), the relationship between e_α, the referent of *the other phone*, and its antecedent, e_r, depends in part on the anaphor, and not just on the antecedent—in particular, on the presence of the word *other*.

Second, we also have examples such as

(27) Sue lifted the receiver as Tom darted to *the other phone*.[7]

in which the referent of *the other phone* (e_α) is the phone *other than* the phone *associated with* the receiver that Sue lifted. Together, these two points argue for a third possibility, in which an anaphoric element can convey a specific function f_α that is idiosyncratic to the anaphor, which may be applied to either e_r or an associate of e_r. The result of that application is e_α. For want of a better name, we will call these **lexically specified** anaphors.

Other lexically specified anaphors include noun phrases headed by *other* (example (28)), NPs with *such* but no postmodifying *as* phrase (example (29)), comparative

---

7 Modjeska (2001) discovered such examples in the British National Corpus.

NPs with no postmodifying *than* phrase (example (30)), and the pronoun *elsewhere* (example (31)) (Bierner 2001b)

(28) Some dogs are constantly on the move. *Others* lie around until you call them.

(29) I saw a 2kg lobster in the fish store yesterday. The fishmonger said it takes about five years to grow to *such a size*.

(30) Terriers are very nervous. *Larger dogs* tend to have calmer dispositions.

(31) I don't like sitting in this room. Can we move *elsewhere*?

To summarize the situation with anaphors so far, we have coreference when $e_\alpha = e_r$, indirect anaphora when $e_\alpha \in \text{assoc}(e_r)$, and lexically specified anaphora when $e_\alpha = f_\alpha(e_i)$ where $e_i = e_r$ or $e_i \in \text{assoc}(e_r)$.

3.2 Discourse Adverbials as Lexical Anaphors
There is nothing in this generalized approach to discourse anaphora that requires that the source of e_r be an NP, or that the anaphor be a pronoun or NP. For example, the antecedent e_r of a singular demonstrative pronoun (in English, *this* or *that*) is often an **eventuality** that derives from a clause, a sentence, or a larger unit in the recent discourse (Asher 1993; Byron 2002; Eckert and Strube 2000; Webber 1991). We will show that this is the case with discourse adverbials as well.

The extension we make to the general framework presented above in order to include discourse adverbials as discourse anaphors is to allow more general functions f_α to be associated with lexically specified anaphors. In particular, for the discourse adverbials considered in this article, the function associated with an adverbial maps its anaphoric argument—an eventuality derived from the current discourse context—to a function that applies to the interpretation of the adverbial's matrix clause (itself an eventuality). The result is a **binary relation** that holds between the two eventualities and is added to the discourse context. For example, in

(32) John loves Barolo. So he ordered three cases of the '97. But he had to cancel the order because he *then* discovered he was broke.

then, roughly speaking, contributes the fact that its matrix clause event (John's finding he was broke) is *after* the anaphorically derived event of his ordering the wine.[8] Similarly, in

(33) John didn't have enough money to buy a mango. *Instead*, he bought a guava.

instead contributes the fact that its matrix clause event (buying a guava) is an alternative to the anaphorically derived event of buying a mango. The relation between the two sentences is something like **result**, as in *So instead, he bought a guava*.

---

8 Words and phrases that function as discourse adverbials usually have other roles as well: For example, *otherwise* also serves as an adjectival modifier, as in *I was* otherwise *occupied with grading exams*. This overloading of closed-class lexico-syntactic items is not unusual in English, and any ambiguities that arise must be handled as part of the normal ambiguity resolution process.

Note that our only concern here is with the compositional and anaphoric **mechanisms** by which adverbials contribute meaning. For detailed analysis of the lexical semantics of adverbials (but no attention to mechanism), the reader is referred to Jayes and Rossari (1998a, 1998b, Lagerwerf (1998), Traugott (1995, 1997), and others.

Formally, we represent the function that a discourse adverbial α contributes as a λ-expression involving a binary relation R_α that is idiosyncratic to α, one of whose arguments (represented here by the variable EV) is resolved anaphorically:

$$\lambda x \; . \; R_\alpha(x, EV)$$

R_α gets its other argument compositionally, when this λ-expression is applied to α's matrix clause S interpreted as an eventuality σ, that is,

$$[\lambda x \; . \; R_\alpha(x, EV)]\sigma \equiv R_\alpha(\sigma, EV)$$

The result of both function application and resolving EV to some eventuality e_i derived from the discourse context either directly or by association is the proposition $R_\alpha(\sigma, e_i)$, one of whose arguments (e_i) has been supplied by the discourse context and the other (σ) compositionally from syntax.

Note that this is a formal model, meant to have no implications for how processing takes place. We have not tried at this stage to instantiate our view of how discourse adverbials are resolved in the context of (simultaneous) sentence-level and discourse-level processing. Our basic view is that resolution is initiated when the discourse adverbial (α) is encountered. As α's matrix clause S is incrementally parsed and interpreted, producing eventuality σ, the resolution process polls the discourse context and either finds an appropriate eventuality e_i (or creates one by a bridging inference, as illustrated in the next section) such that $R_\alpha(\sigma, e_i)$ makes sense with respect to the discourse so far. As is the case with resolving a **discourse deictic** (Asher 1993; Byron 2002; Eckert and Strube 2000; Webber 1991) this resolution process would use syntactic and semantic constraints that it accumulates as the incremental sentence-level parser/interpreter processes S. As with discourse deixis, this is best seen as a constraint satisfaction problem that involves finding or deriving an eventuality from the current discourse context that meets the constraints of the adverbial with respect to the eventuality interpretation of the matrix clause. (Examples of this are given throughout the rest of the article.)

3.3 A Logical Form for Eventualities

Before using this generalized view of anaphora to show what discourse adverbials contribute to discourse and how they interact with discourse relations that arise from adjacency or explicit discourse connectives, we briefly describe how we represent clausal interpretations in logical form (LF).

Essentially, we follow Hobbs (1985) in using a rich ontology and a representation scheme that makes explicit all the individuals and **abstract objects** (i.e., propositions, facts/beliefs, and eventualities) (Asher 1993) involved in the LF interpretation of an utterance. We do so because we want to make intuitions about individuals, eventualities, lexical meaning, and anaphora as clear as possible. But certainly, other forms of representation are possible.

In this LF representation scheme, each clause and each relation between clauses is indexed by the label of its associated abstract object. So, for example, the LF interpretation of the sentence

(34) John left because Mary left.

would be written

e_1:left(j) $\wedge$ john(j) $\wedge$ e_2:left(m) $\wedge$ mary(m) $\wedge$ e_3:because(e_1,e_2)

where the first argument of the asymmetric binary predicate *because* is the consequent and the second is the eventuality leading to this consequent. Thus when *because* occurs sentence-medially, as in the above example, the eventuality arguments are in the same order as their corresponding clauses occur in the text. When *because* occurs sentence-initially (as in *Because Mary left, John did*), the interpretation of the second clause (*John* [*left*]) will appear as the first argument and the interpretation of the first clause (*Mary left*) will appear as the second.[9]

   The set of available discourse referents includes both individuals like *j* and *m*, and also abstract objects like e_1 and e_2. We then represent resolved anaphors by reusing these discourse referents. So, for example, the LF interpretation of the follow-on sentence

   (35) This upset Sue.

would be written

e_4:upset(DPRO, s) $\wedge$ sue(s)

where DPRO is the anaphoric variable contributed by the demonstrative pronoun *this*. Since the subject of *upset* could be either the eventuality of John's leaving or the fact that he left because Mary left, DPRO could be resolved to either e_1 or e_3, that is,

   a. e_4:upset(e_1, s) $\wedge$ sue(s)
   b. e_4:upset(e_3, s) $\wedge$ sue(s)

depending on whether one took Sue to have been upset by (1) John's leaving or (2) that he left because Mary left.

3.4 The Contribution of Discourse Adverbials to Discourse Semantics
Here we step through some examples of discourse adverbials and demonstrate how they make their semantic contribution to the discourse context. We start with example (32), repeated here as (36):

   (36) a. John loves Barolo.
         b. So he ordered three cases of the '97.
         c. But he had to cancel the order
         d. because he *then* discovered he was broke.

---

9 We are not claiming to give a detailed semantics of discourse connectives except insofar as they may affect how discourse adverbials are resolved. Thus, for example, we are not bothering to distinguish among different senses of *because* (epistemic vs. nonepistemic), *while* (temporal vs. concessive), *since* (temporal vs. causal), etc. Of course, these distinctions are important to discourse interpretation, but they are independent of and orthogonal to the points made in this article. Similarly, Asher (1993) argues that a simple ontology of eventualities is too coarse-grained, and that discourse representations need to distinguish different kinds of abstract objects, including actions, propositions, and facts as well as eventualities. Different discourse connectives will require different kinds of abstract objects as arguments. This distinction is also orthogonal to the points made in this article, because we can understand these abstract referents to be associates of the corresponding Hobbsian eventualities and leave the appropriate choice to the lexical semantics of discourse connectives. Byron (2002) advocates a similar approach to resolving discourse anaphora.

Using the above LF representation scheme and our notation from Section 3.2, namely,

- α = the anaphoric expression (here, the discourse adverbial)
- R_α = the relation name linked with α
- S = the matrix clause/sentence containing α
- σ = the interpretation of S as an abstract object

and ignoring, for now, the conjunction *because* (discussed in section 4), the relevant elements of (36d) can be represented as:

α = *then*
R_α = *after*
S = he [John] discovered he was broke
σ = e_4:find(j,e_5), where e_5:broke(j)

This means that the unresolved interpretation of (36d) is

$$[\lambda x \, . \, R_\alpha(x,EV)]\sigma \; \equiv \; [\lambda x \, . \, after(x,EV)]e_4 \equiv after(e_4, EV)$$

The anaphoric argument EV is resolved to the eventuality e_2, derived from (36b)—e_2:order(j, c_1).

$$after(e_4,EV) \rightarrow after(e_4,e_2)$$

That is, the eventuality of John's finding he was broke is after that of John's ordering three cases of the '97 Barolo. The resulting proposition $after(e_4,e_2)$ would be given its own index, e_6, and added to the discourse context.

When *then* is understood temporally, as it is above, as opposed to logically, it requires a culminated eventuality from the discourse context as its first argument (which Vendler (1967) calls an **achievement** or an **accomplishment**). The ordering event in (36b) is such a Vendlerian accomplishment. In example (37), though, there is no culminated eventuality in the discourse context for *then*), to take as its first argument.

(37) a. Go west on Lancaster Avenue.
      b. *Then* turn right on County Line.

How does (37b) get its interpretation?
    As with (36d), the relevant elements of (37b) can be represented as

α = *then*
R_α = *after*
S = turn right on County Line
σ = e_3:turn-right(*you*, *county_line*)

and the unresolved interpretation of (37b) is thus

$$[\lambda x \, . \, after(x, EV)]e_3 \equiv after(e_3, EV)$$

As for resolving *EV*, in a well-known article, Moens and Steedman (1988) discuss several ways in which an eventuality of one type (e.g., a process) can be coerced into an eventuality of another type (e.g., an accomplishment, which Moens and Steedman call a **culminated process**). In this case, the matrix argument of *then* (the eventuality of *turning right on County Line*) can be used to coerce the process eventuality in (37b) into a culminated process of *going west on Lancaster Avenue until County Line*. We treat this coercion as a type of associative or bridging inference, as in the examples discussed in section 3.1. That is,

$$e_2 = \text{culmination}(e_1) \in \text{assoc}(e_1), \text{ where } e_1\text{:go-west}(you, \text{lancaster\_ave})$$

Taking this e_2 as the anaphoric argument *EV* of *then* yields the proposition

$$after(e_3, e_2)$$

That is, the eventuality of turning right onto County Line is after that of going west on Lancaster Avenue to County Line. This proposition would be indexed and added to the discourse context.

It is important to stress here that the level of representation we are concerned with is essentially an LF for discourse. Any reasoning that might then have to be done on the content of LFs might then require making explicit the different modal and temporal contexts involved, their accessibility relations, the status of abstract objects as facts, propositions or eventualities, and so on. But as our goal here is primarily to capture the mechanism by means of which discourse adverbials are involved in discourse structure and discourse semantics, we will continue to assume for as long as possible that an LF representation will suffice.

Now it may appear as if there is no difference between treating adverbials as anaphors and treating them as structural connectives, especially in cases like (37) in which the antecedent comes from the immediately left-adjacent context, and in which the only obvious semantic relation between the adjacent sentences appears to be the one expressed by the discourse adverbial. (Of course, there may also be a separate **intentional** relation between the two sentences [Moore and Pollack 1992], independent of the relation conveyed by the discourse adverbial.)

One must distinguish, however, between whether a theory allows a distinction to be made and whether that distinction needs to be made in a particular case. It is clear that there are many examples in which the two approaches (i.e., a purely structural treatment of all connectives, versus one that treats adverbials as linking into the discourse context anaphorically) appear to make the same prediction. We have already, however, demonstrated cases in which a purely structural account makes the wrong prediction, and in the next section, we will demonstrate the additional power of an account that allows for two relations between an adverbial's matrix clause or sentence and the previous discourse: one arising from the anaphoric connection and the other inferred from adjacency or conveyed explicitly by a structural connective.

Before closing this section, we want to step through examples (19)–(20), repeated here as examples (38)–(39).

(38) If the light is red, stop. *Otherwise* you'll get a ticket.

(39) If the light is red, stop. *Otherwise* go straight on.

Roughly speaking, *otherwise* conveys that the **complement** of its anaphorically derived argument serves as the **condition** under which the interpretation of its structural argument holds. (This complement must be with respect to some contextually relevant set.)[10]

If we represent a conditional relation between two eventualities with the asymmetric relation *if(e_1,e_2)*, where e_1 is derived from the **antecedent** and e_2 from the **consequent**, and we approximate a single contextually relevant alternative e_2 to an eventuality e_1 using a symmetric *complement* relation, *complement(e_1, e_2)*, then we can represent the interpretation of *otherwise* as

$$\lambda x \; . \; \textit{if}(VE, x), \text{ where } \textit{complement}(VE, EV)$$

where variable *EV* is resolved anaphorically to an eventuality in the current discourse context that admits a complement. That is, *otherwise* requires a contextually relevant complement to its antecedent and asserts that if that complement holds, the argument to the λ-expression will as well. The resulting λ-expression applies to the interpretation of the matrix clause of *otherwise*, resulting in the conditional's being added to the discourse context:

$$[\lambda x \; . \; \textit{if}(VE,x)] \; \sigma \equiv \textit{if}(VE,\sigma), \text{ where } \textit{complement}(VE,EV)$$

Here the relevant elements of (38b) and (39b) can be represented as

α = otherwise
R_α = *if*
S_{38} = you get a ticket
σ_{38} = e_3, where e_3:get_ticket(you)
S_{39} = go straight on
σ_{39} = $e_{3'}$, where $e_{3'}$:go_straight(you)

The unresolved interpretations of (38b) and (39b) are thus:

$$[\lambda x \; . \; \textit{if}(VE_{38},x)] \; e_3 \; \equiv \; \textit{if}(VE_{38},e_3), \text{ where } \textit{complement}(VE_{38},EV_{38})$$
$$[\lambda x \; . \; \textit{if}(VE_{39},x)] \; e_{3'} \; \equiv \; \textit{if}(VE_{39},e_{3'}), \text{ where } \textit{complement}(VE_{39},EV_{39})$$

As we showed in section 2.2, different ways of resolving the anaphoric argument lead to different interpretations. In (38), the anaphoric argument is resolved to e_2:stop(you),

---

10 Kruijff-Korbayová and Webber (2001a) demonstrate that the **information structure** of sentences in the previous discourse (theme-rheme partitioning, as well as focus within theme and within rheme [Steedman 2000a]) can influence what eventualities e_r are available for resolving the anaphorically derived argument of *otherwise*. This then correctly predicts different interpretations for "otherwise" in (i) and (ii):

  (i) Q. How should I transport the dog?
     A. You should CARRY the dog. Otherwise you might get HURT.

  (ii) Q. What should I carry?
     A. You should carry THE DOG. Otherwise you might get HURT.

In both (i) and (ii), the questions constrain the theme/rheme partition of the answer. Small capitals represent focus within the rheme. In (i), the *otherwise* clause will be interpreted as warning the hearer (H) that H might get hurt if he/she transports the dog in some way other than carrying it (e.g., H might get tangled up in its lead). In (ii), the *otherwise* clause warns H that he/she might get hurt if what she is carrying is not the dog (e.g., H might be walking past fanatical members of the Royal Kennel Club).

whereas in (39), it is resolved to e_1:red(light1). Thus the resulting interpretations of (38b) and (39b) are, respectively,

> *if(e_4,e_3)*, where *complement(e_2,e_4)* and e_2:stop(you)
> (*If you do something other than stop, you'll get a ticket.*)

> *if($e_{4'}$, $e_{3'}$)*, where *complement(e_1,$e_{4'}$)* and e_1:red(light)
> (*If the light is not red, go straight on.*)

We have not been specific about how the anaphoric argument of *otherwise* (or of any other discourse adverbial) is resolved, other than having it treated as a constraint satisfaction problem. This is the subject of current and future work, exploring the empirical properties of resolution algorithms with data drawn from appropriately annotated corpora and from psycholinguistic studies of human discourse interpretation. To this end, Creswell et al. (2002) report on a preliminary annotation study of discourse adverbials and the location and type of their antecedents. This initial effort involves nine discourse adverbials—three each from the classes of **concessive**, **result**, and **reinforcing** (additive) conjuncts given in Quirk et al. (1972). Meanwhile, Venditti et al. (2002) present a preliminary report on the use of a constraint satisfaction model of interpretation, crucially combining anaphoric and structural reasoning about discourse relations, to predict subjects' on-line interpretation of discourses involving stressed pronouns. In addition, two proposals have recently been submitted to construct a larger and more extensively annotated corpus, covering more adverbials, based on what we have learned from this initial effort. This more extensive study would be an adequate basis for developing resolution algorithms.[11]

3.5 Summary
In this section, we have presented a general framework for anaphora with the following features:

- Anaphors can access one or more discourse referents or entities associated with them through bridging inferences. These are sufficient for interpreting anaphoric pronouns, definite NPs and demonstrative NPs, allowing entities to be evoked by NPs or by clauses. In the case of clauses, this may be on an as-needed basis, as in Eckert and Strube (2000).

- A type of anaphor α that we call *lexically specified* can also contribute additional meaning through a function f_α that is idiosyncratic to α, that can be applied to either an existing discourse referent or an entity associated with it through a bridging inference. In the case of the premodifier *other*, f_α applied to its argument produces contextually

---

11 With respect to how many discourse adverbials there are, Quirk et al. (1972) discuss 60 conjunctions and discourse adverbials under the overall heading *time relations* and 123 under the overall heading *conjuncts*. Some entries appear under several headings, so that the total number of conjunctions and discourse adverbials they present is closer to 160. In another enumeration of discourse adverbials, Forbes and Webber (2002) start with all annotations of sentence-level adverbials in the Penn Treebank, then filter them systematically to determine which draw part of their meaning from the preceding discourse and how they do so. What we understand from both of these studies is that there are fewer than 200 adverbials to be considered, many of which are minor variations of one another (*in contrast, by contrast, by way of contrast, in comparison, by comparison, by way of comparison* that are unlikely to differ in their anaphoric properties, and some of which, such as *contrariwise, hitherto*, and *to cap it all*, will occur only rarely in a corpus of modern English.

relevant alternatives to that argument. In the case of the premodifier
such, it yields a set of entities that are similar to its argument in a
contextually relevant way.

- Discourse adverbials are lexically specified anaphors whose meaning
  function f_α is a λ-expression involving a binary relation R_α that is
  idiosyncratic to α, one of whose arguments is resolved anaphorically and
  the other is provided compositionally, when the λ-expression is applied
  to α's matrix clause interpreted as an eventuality σ.

In the next section, we move on to consider how the presence of both a semantic rela-
tion associated with a discourse adverbial and a semantic relation associated with the
adjacency of two clauses or a structural connective between them allows for interesting
interactions between the two.

4. Patterns of Anaphoric Relations and Structural/Inferred Relations

Prior to the current work, researchers have treated both explicit structural connec-
tives (coordinating and subordinating conjunctions, and "paired" conjunctions) and
discourse adverbials simply as evidence for a particular structural relation holding
between adjacent units. For example, Kehler (2002) takes *but* as evidence of a **contrast**
relation between adjacent units, *in general* as evidence of a **generalization** relation,
in other words as evidence of an **elaboration** relation, *therefore* as evidence of a **result**
relation, *because* as evidence of an **explanation** relation, and *even though* as evidence
of a **denial of preventer** relation (Kehler 2002, Section 2.1). Here Kehler has probably
correctly identified the *type* of relation that holds between elements, but not which
elements it holds between.

 In one respect, we follow previous researchers, in that we accept that when clauses,
sentences, or larger discourse units are placed adjacent to one another, listeners infer a
relation between the two, and that the structural connective (coordinate or subordinate
conjunction) gives evidence for the relation that is intended to hold between them.

 Because we take discourse adverbials to contribute meaning through an anaphoric
connection with the previous discourse, however, this means that there may be two
relations on offer and opens up the possibility that the relation contributed by the
discourse adverbial can interact in more than one way with the relation conveyed
by a structural connective or inferred through adjacency. Below we show that this
prediction is correct.

 We start from the idea that, in the absence of an explicit structural connective, de-
feasible inference correlates with structural attachment of adjacent discourse segments
in discourse structure, relating their interpretations. The most basic relation is that the
following segment in some way describes the same object or eventuality as the one it
abuts (*elaboration*). But evidence in the segments can lead (via defeasible inference) to
a more specific relation, such as one of the *resemblance* relations (e.g., *parallel*, *contrast*,
exemplification, *generalisation*), or *cause-effect* relations (*result*, *explanation*, *violated expecta-
tion*), or *contiguity* relations (*narration*) described in Hobbs (1990) and Kehler (2002). If
nothing more specific can be inferred, the relation will remain simply *elaboration*. What
explicit structural connectives can do is convey relations that are not easy to convey
by defeasible inference (e.g., *if*, conveying *condition*, and *or*, conveying *disjunction*) or
provide nondefeasible evidence for an inferrable relation (e.g., *yet, so,* and *because*).

 Discourse adverbials can **interact** with structural connectives, with adjacency-
triggered defeasible inference, and with each other. To describe the ways in which we

have so far observed discourse adverbials to interact with relations conveyed structurally, we extend the notation used in the previous section:

- α = discourse adverbial

- R_α = the name of the relation associated with α

- S = the matrix clause/sentence of α

- σ = the logical form (LF) interpretation of S

adding the following:

- D = the discourse unit that is left-adjacent to S, to which a relationship holds by either inference or a structural connective

- δ = the LF interpretation of D

- R = the name of the relation that holds with δ

Although δ is one argument of R, we show below that the other argument of R may be one of at least two different abstract objects.

Case 1: σ separately serves as an argument to both R_α and R. This is the case that holds in example (36) (repeated here):

> (36) a. John loves Barolo.
>   b. So he ordered three cases of the '97.
>   c. But he had to cancel the order
>   d. because he *then* discovered he was broke.

We have already seen that the interpretation of the clause in (36d) following *because* involves

> $R_\alpha = $ *after*
> $\sigma = e_4$:discover(j,e_5), where e_5:broke(j)
> $[\lambda x \,.\, after(x,EV)]e_4 \equiv after(e_4, EV)$

where EV is resolved to e_2:order(j, c_1), and the proposition $after(e_4, e_2)$ is added to the discourse context—that is, John's discovering he was broke is after his ordering the wine.

    Now consider the *explanation* relation R associated with *because* in (36d). It relates e_4, John's finding he was broke, to the intepretation of (36c), e_3:cancel(j,o_1)—that is, *explanation*(e_4,e_3). Clause 36d thus adds both *explanation*(e_4,e_3) and $after(e_4,e_2)$ to the discourse. Although these two propositions share an argument (e_4), they are nevertheless distinct.[12]

---

12 Because eventuality e_4, *John's finding he was broke*, both *explains* the canceling and *follows* the ordering, it follows that the canceling is after the ordering.

Case 2: $R_\alpha(\sigma, e_i)$ is an argument of R. In case 1, it is the interpretation of the adverbial's matrix clause σ that serves as one argument to the discourse relation R. In contrast, in case 2, that argument is filled by the relation contributed by the discourse adverbial (itself an abstract object available for subsequent reference). In both cases, the other argument to R is δ.

One configuration in which case 2 holds is with the discourse adverbial *otherwise*. Recall from section 3.4 that the interpretation of *otherwise* involves a conditional relation between the complement of its anaphoric argument and the interpretation σ of its matrix clause:

$$[\lambda x \,.\, if(VE,x)] \; \sigma \equiv if(VE,\sigma), \text{ where } complement(VE,EV)$$

With variable *EV* resolved to an eventuality in the discourse context, it is the resulting relation (viewed as an abstract object) that serves as one argument to R, with δ serving as the other. We can see this most clearly by considering variants of examples (38) and (39) that contain an explicit connective between the clauses. In (38), the conjunction *because* is made explicit (example (40)), and in (39), the connective is simply *and* or *but* (example (41)).

(40) If the light is red, stop, because *otherwise* you'll get a ticket.
$R_\alpha = if$
$\sigma_{38} = e_3$:get_ticket(you)

(41) If the light is red, stop, and/but *otherwise* go straight on.
$R_\alpha = if$
$\sigma_{39} = e_{3'}$:go_straight(you)

In the case of (40), resolving *otherwise* contributes the relation

e_6: $if(e_4,e_3)$, where *complement*(e_4,e_2) and e_2:stop(you)
(*If you do something other than stop, you'll get a ticket.*)

At the level of LF, the abstract object e_6 that is associated with the conditional relation serves as one argument to the *explanation* relation contributed by *because*, with e_2 being the other. That is, *because* and *otherwise* together end up contributing *explanation*(e_2,e_6) (i.e., your needing to stop is explained by the fact that if you do something other than stop, you'll get a ticket).

In the case of (41), resolving *otherwise* contributes the relation

$e_{6'}$:$if(e_{4'}, e_{3'})$, where *complement*$(e_{4'},e_1)$ and e_1:red(light)
(*If the light is not red, go straight on.*)

What is the discourse relation to which *otherwise* contributes this abstract object $e_{6'}$? Whether the connective is *and* or *but*, both its conjuncts describe (elaborate) alternative specializations of the same situation e_0 introduced earlier in the discourse (e.g., e_0 could be associated with the first sentence of *Go another mile and you'll get to a bridge. If the light is red, stop. Otherwise go straight on.*) If the connective is *and*, what is added to context might simply be *elaboration*$(e_{6'},e_0)$. (Note that without *otherwise*, the relation *elaboration*(e_5,e_0) would have been added to context, where e_5 is the abstract object associated with the interpretation of *If the light is red, stop.*) If the connective is *but*, then one might also possibly add *contrast*$(e_{6'},e_5)$—that is, the situation that (if the light is

red) you should stop is in contrast to the situation that if the light is not red, you should go straight on.[13]

As is clear from the original pair of examples (38) and (39), interpretations can arise through adjacency-triggered inference that are similar to those that arise with an explicit connective. In either case, the above treatment demonstrates that there is no need for a separate *otherwise* relation, as proposed in rhetorical structure theory (Mann and Thompson 1988). We are not, however, entirely clear at this point when case 1 holds and when case 2 does. A more careful analysis is clearly required.

Case 3: R_α is parasitic on R. Case 3 appears to apply with discourse adverbials such as *for example* and *for instance*. The interpretation of such adverbials appears to be parasitic on the relation associated with a structural connective or discourse adverbial to their left, or on an inferred relation triggered by adjacency. The way to understand this is to first consider intraclausal *for example*, where it follows the verb, as in

(42) Q. What does this box contain?
     A. It contains, *for example*, some hematite.

The interpretation of *for example* here involves abstracting the meaning of its matrix structure with respect to the material to its right, then making an assertion with respect to this abstraction. That is, if the LF contributed by the matrix clause of (42A) is, roughly,

i. contain(box1,hematite1)

then the LF resulting from the addition of *for example* can be written either with set notation (as in (ii)), taking an entity to exemplify a set, or with λ-notation (as in (iii)), taking an entity to exemplify a property:

ii. *exemplify*(hematite1, {X | contain(box1,X)})
iii. *exemplify*(hematite1, λX . contain(box1,X))

Both express the fact that *hematite* is an example of what is contained in the box.[14] Since one can derive (i) logically from either (ii) or (iii), one might choose to retain only (ii) or (iii) and derive (i) if and when it is needed. In the remainder of the article, we use the λ notation given in (iii). Note that from the perspective of compositional semantics, *for example* resembles a quantifier, in that the scope of its interpretation is not isomorphic to its syntactic position. Thus producing an interpretation for *for example* will require techniques similar to those that have long been used in interpreting quantifiers (Woods, 1978; Barwise and Cooper 1981). We take this up again in section 5.

If we look at the comparable situation in discourse, such as (43)–(44), where *for example* occurs to the right of a discourse connective, it can also be seen as abstracting

---

13 A much finer-grained treatment of the semantics of *otherwise* in terms of context-update potential is given in Kruijff-Korbayová and Webber (2001b). Here we are just concerned with its interaction with structural connectives and adjacency-triggered relations.

14 The material to the right of *for example* can be any kind of constituent, including such strange ones as

John gave, for example, a flower to a nurse.

Here, *a flower to a nurse* would be an example of the set of object-recipient pairs within John's givings. Such nonstandard constituents are also found with coordination, which was one motivation for combinatory categorial grammar (Steedman 1996). This just illustrates another case in which such nonstandard constituents are needed.

the interpretation of its discourse-level matrix structure, with respect to the material to its right:

(43) John just broke his arm. *So, for example*, he can't cycle to work now.

(44) You shouldn't trust John *because, for example*, he never returns what he borrows.

In (43), the connective *so* leads to

$result(\sigma, \delta)$

being added to the discourse, where σ is the interpretation of *John can't cycle to work now*, and δ is the interpretation of *John just broke his arm*. *For example* then abstracts this relation with respect to the material to its right (i.e., σ), thereby contributing

$exemplify(\sigma, \lambda X . result(X, \delta))$

That is, *John can't cycle to work* is an example of what results from *John's breaking his arm*. Similarly, *because* in (44) leads to

$explanation(\sigma, \delta)$

being added to the discourse, where σ is the interpretation of *he never returns what he borrows*, δ is the interpretation of *you shouldn't trust John*, and *for example* adds

$exemplify(\sigma, \lambda X . explanation(X, \delta))$

that is, that σ is an example of the reasons for not trusting John.
  For example interacts with discourse adverbials in the same way:

(45) Shall we go to the Lincoln Memorial? *Then, for example*, we can go to the White House.

(46) As a money manager and a grass-roots environmentalist, I was very disappointed to read in the premiere issue of Garbage that The Wall Street Journal uses 220,000 metric tons of newsprint each year, but that only 1.4% of it comes from recycled paper. *By contrast*, the Los Angeles Times, *for example*, uses 83% recycled paper. [WSJ, from Penn Treebank /02/wsj-0269]

In example (45), the resolved discourse adverbial *then* leads to $after(\sigma, \delta)$ being added to the discourse context, where σ is the interpretation of *we can go to the White House*, δ is the interpretation of *we shall go to the Lincoln Memorial*, and *for example* adds

$exemplify(\sigma, \lambda X . after(X, \delta))$

that is, that σ is an example of the events that [can] follow going to the Lincoln Memorial. (As already noted, we are being fairly fast and loose regarding tense and modality, in the interests of focusing on the types of interactions.)

In example (46), the resolved discourse anaphor *by contrast* contributes *contrast*(σ,δ), where σ is the interpretation of *the Los Angeles Times's using 83% recycled paper* and δ is the intepretation of *only 1.4% of it [newsprint used by the WSJ] comes from recycled paper*. *For example* then contributes

exemplify(σ, λX . *contrast*(X,δ))

that is, that σ is one example of contrasts with the WSJ's minimal use of recycled paper.

What occurs with discourse connectives and adverbials can also occur with relations added through adjacency-triggered defeasible inference, as in

(47) You shouldn't trust John. *For example*, he never returns what he borrows.
explanation(δ,σ)
exemplify(σ, λX . *explanation*(δ,X))

Here, as in (44), the relation provided by adjacency-triggered inference is $R = $ *explanation*, which is then used by *for example*.

But what about the many cases in which only *exemplify* seems present, as in

(48) In some respects they [hypertext books] are clearly superior to normal books, for example they have database cross-referencing facilities ordinary volumes lack. (British National Corpus, CBX 1087)

(49) He [James Bellows] and his successor, Mary Anne Dolan, restored respect for the editorial product, and though in recent years the paper had been limping along on limited resources, its accomplishments were notable. *For example*, the Herald consistently beat its much-larger rival on disclosures about Los Angeles Mayor Tom Bradley's financial dealings.

There are at least two explanations: One is that *for example* simply provides direct nondefeasible evidence for *exemplify*, which is the only relation that holds. The other explanation follows the same pattern as the examples given above, but with no further relation than *elaboration*(σ,δ). That is, we understand in (48) that *having database cross-referencing facilities* elaborates the respects in which hypertext books are superior to normal books, whereas in (49), we understand that *the Herald's [newspaper] consistently beating its much-larger rival* elaborates the claim that *its accomplishments were notable*. This *elaboration* relation is then abstracted (in response to *for example*) to produce:

exemplify(σ, λX . *elaboration*(X, δ))

that is, that this is one example of many possible elaborations. Because this is more specific than *elaboration* and seems to mean the same as *exemplify*(σ,δ), one might simply take it to be the only relation that holds. Given that so many naturally occuring instances of *for example* occur with *elaboration*, it is probably useful to persist with the above shorthand. But it shouldn't obscure the regular pattern that appears to hold.

Before going on to case 4, we should comment on an ambiguity associated with *for example*. When *for example* occurs after an NP, a PP, or a clause that can be interpreted as a general concept or a set, it can contribute a relation between the general concept or set and an instance, rather than being parasitic on another relation. For example, in:

(50) In the case of the managed funds they will be denominated in a leading
       currency, for example US dollar, ... (BNC CBX 1590)

for example relates the general concept denoted by *a leading currency* to a specific in-
stance, U.S. dollars. (In British English, the BNC shows that most such examples occur
with *such as*—i.e., in the construction *such as for example*. This paraphrase does not work
with the predicate-abstracting *for example* that is of primary concern here, as in exam-
ple (42).)

     But *for example* occurring after an NP, a PP, or a clause can, alternatively, contribute
a more subtle parasitic relationship to the previous clause, as in

(51) All the children are ill, so Andrew, *for example*, can't help out in the shop.

This differs from both (43) and (50). That is, one cannot paraphrase (51) as (52) as in
(43) where *for example* follows *so*:

(52) All the children are ill, so *for example* Andrew can't help out in the shop.

Example (52) simply specifies an example consequence of all the children being ill, as
does

(53) All the children are ill, so *for example* one of us has to be at home at all
       times.

In contrast, (51) specifies an example consequence for Andrew, *as one of the children*.
Support for this comes from the fact that in (52), Andrew doesn't have to be one of
the children: He could be their nanny or child minder, now stuck with dealing with a
lot of sick kids. But (51) is not felicitous if Andrew is not one of the children.

     We suspect here the involvement of **information structure** (Steedman 2000a):
Whereas the interpretation conveyed by *for example* is parasitic on the adjacency rela-
tion (*result* in example (51)), its position after the NP *Andrew* in (51) may indicate a
contrastive theme with respect to the previous clause, according to which Andrew *in
contrast to the other children* suffers this particular consequence. But more work needs
to be done on this to gain a full understanding of what is going on.

Case 4: R_α is a defeasible rule that incorporates R. Case 4 occurs with discourse
adverbials that carry the same presupposition as the discourse connectives *although*
and the concessive sense of *while* (Lagerwerf 1998). Case 4 shares one feature with
case 1, in that the discourse relation *R* conveyed by a structural connective or inferred
from adjacency holds between σ (the interpretation of the adverbial's matrix clause)
and δ (the interpretation of the left-adjacent discourse unit). Where it differs is that
the result is then incorporated into the **presupposition** of the discourse adverbial. This
presupposition, according to Lagerwerf (1998), has the nature of a presupposed (or
conventionally implicated) **defeasible rule** that fails to hold in the current situation.
He gives as an example

(54) Although Greta Garbo was called the yardstick of beauty, she never mar-
       ried.

This asserts both that Greta Garbo was called the yardstick of beauty and that she
never married. The first implies that Greta Garbo was beautiful. The example also

presupposes that, in general, if a woman is beautiful, she will marry. If such a pre-supposition can be accommodated, it will simply be added to the discourse context. If not, the hearer will find the utterance confusing or possibly even insulting.

We argue here that the same thing happens with the discourse adverbials *nevertheless* and *though*. The difference is that, with discourse adverbials, the antecedent to the rule derives anaphorically from the previous discourse, whereas the consequent derives from the adverbial's matrix clause. (With the conjunctions *although* and con-cessive *while*, both arguments are provided structurally.)

We first illustrate case 4 with two examples in which *nevertheless* occurs in the main clause of a sentence containing a preposed subordinate clause. The subordinate conjunction helps clarify the relation between the clauses that forms the basis for the presupposed defeasible rule. After these, we give a further example in which the relation between the adjacent clauses comes through inference.

(55) *While* John is discussing politics, he is *nevertheless* thinking about his fish.

In (55), the conjunction *while* conveys a temporal relation R between the two clauses it connects:

$$during(e_2, e_1), \text{ where } e_1:\text{discuss(john,politics) and } e_2:\text{think\_about(john,fish)}$$

What *nevertheless* contributes to (55) is a defeasible rule based on this relation, which we will write informally as

$during(X,E) \land E:\text{discuss}(Y,\text{politics})) > \neg X:\text{think\_about}(Y,\text{fish}))$
Normally, whatever one does during the time one is discussing politics, it is not thinking about one's fish.

This rule uses Asher and Morreau's (1991) defeasible implication operator ($>$) and abstracts over the individual (John), which seems appropriate for the general statement conveyed by the present tense of the utterance.

Similarly, in

(56) Even *after* John has had three glasses of wine, he is *nevertheless* able to solve difficult math problems.

the conjunction *after* contributes a relation between the two clauses it connects:

$$after(e_2, e_1), \text{ where } e_1:\text{drink(john,wine) and } e_2:\text{solve(john,hard\_problems)}$$

What *nevertheless* contributes to this example is a defeasible rule that we will again write informally as

$after(X,E) \land E:\text{drink}(Y,\text{wine})) > \neg X:\text{solve}(Y,\text{hard\_problems}))$
Normally, whatever one is able to do after one has had three glasses of wine, it is not solving difficult algebra problems.

Again, we have abstracted over the individual, as the presupposed defeasible rule associated with the present-tense sentence appears to be more general than a statement about a particular individual.[15]

On the other hand, in the following example illustrating a presupposed defeasible rule and a discourse relation associated with adjacency, it seems possible for the presupposed defeasible rule to be about John himself:

(57) John is discussing politics. *Nevertheless*, he is thinking about his fish.

Here the discourse relation between the two clauses, each of which denotes a specific event, is

$during(e_2, e_1)$, where e_1:discuss(john,politics) and e_2:think_about(john,fish)

(Note that our LF representation isn't sufficiently rich to express the difference between (55) and (57).) What *nevertheless* contributes here is the presupposed defeasible rule

$during(X,e_1) > \neg X = e_2$
Normally what occurs during John's discussing politics is not John's thinking about his fish.

Lagerwerf (1998) does not discuss how specific or general will be the presupposed defeasible rule that is accommodated or what factors affect the choice. Kruijff-Korbayová and Webber (2001a) also punt on the question, when considering the effect of information structure on what presupposed defeasible rule is associated with *although*. Again, this seems to be a topic for future work.

Summary

We have indicated four ways in which we have found the relation associated with a discourse adverbial to interact with a relation R triggered by adjacency or conveyed by structural connectives or, in some cases, by another relational anaphor:

1. σ separately serves as an argument to both R_α and R.

2. $R_\alpha(\sigma, e_i)$ is an argument of R.

3. R_α is parasitic on R.

4. R_α is a defeasible rule that incorporates R.

We do not know whether this list is exhaustive or whether a discourse adverbial always behaves the same way vis-à-vis other relations. Moreover, in the process of setting down the four cases we discuss, we have identified several problems that we have not addressed, on which further work is needed. Still, we hope that we have convinced the reader of our main thesis: that by recognizing discourse adverbials as doing something different from simply signaling the discourse relation between adjacent discourse units and by considering their contribution as relations in their own right, one can begin to characterize different ways in which anaphoric and structural relations may themselves interact.

---

15 We speculate that the reason examples such as (55) and (56) sound more natural with the focus particle *even* applied to the subordinate clause is that *even* conveys an even greater likelihood that the defeasible rule holds, so *nevertheless* emphasizes its failure to do so.

5. Lexicalized Grammar for Discourse Syntax and Semantics

The question we consider in this section is how the treatment we have presented of discourse adverbials and structural connectives can be incorporated into a general approach to discourse interpretation. There are three possible ways.

The first possibility is simply to incorporate our treatment of adverbials and connectives into a sentence-level grammar, since such grammars already cover the syntax of sentence-level conjunction (both coordinate and subordinate) and the syntax of adverbials of all types. The problem with this approach is that sentence-level grammars, whether phrasal or lexicalized, stop at explicit sentence-level conjunction and do not provide any mechanism for forming the meaning of multiclausal units that cross sentence-level punctuation. Moreover, as we have already shown in section 3, the interpretation of discourse adverbials can interact with the implicit relation between adjacent sentences, as well as with an explicitly signaled relation, so that a syntax and compositional semantics that stops at the sentence will not provide all the structures and associated semantics needed to build the structures and interpretations of interest.

The second possibility is to have a completely different approach to discourse-level syntax and semantics than to sentence-level syntax and semantics, combining (for example) a definite clause grammar with rhetorical structure theory. But as we and others have already noted, this requires discourse semantics reaching further and further into sentence-level syntax and semantics to handle relations between main and embedded clauses, and between embedded clauses themselves, as in example (58).

(58) If they're drunk and they're meant to be on parade and you go to their room and they're lying in a pool of piss, then you lock them up for a day. (*The Independent*, June 17, 1997)

Thus it becomes harder and harder to distinguish the scope of discourse-level syntax and semantics from that at the sentence-level.

The third possibility is to recognize the overlapping scope and similar mechanisms and simply extend a sentence-level grammar and its associated semantic mechanisms to discourse. The additional responsibilities of the grammer would be to account for the formation of larger units of discourse from smaller units; the projection of the interpretation of smaller discourse units onto the interpretation of the larger discourse units they participate in; and the effect of discourse unit interpretation on the evolving discourse model. There are two styles of grammar one could use for this: (1) a phrase structure grammar (PSG) extended to discourse, as in Figure 6, or (2) a lexicalized grammar that extends to discourse, a sentence-level lexicalized grammar such as tree-adjoining grammar (Joshi, 1987; XTAG-Group 2001) or combinatory categorial grammar (CCG) (Steedman 1996, 2000b).

Whereas Polanyi and van den Berg (1996) extend a PSG to discourse, we argue for extending a lexicalized grammar, even though TAG and CCG are weakly context-sensitive (CS) and the power needed for a discourse grammar with no crossing dependencies is only context-free (section 2.1). Our argument is based on our desire to use a discourse grammar in natural language generation (NLG). It is well-known that context-free PSGs (CF PSGs) set up a complex search space for NLG. A discourse grammar specified in terms of phrase structure rules such as those shown in Figure 6 doesn't provide sufficient guidance when reversed for use in generating discourse. For example, one might end up having to guess randomly how many sentences and connectives one had, in what order, before being able to fill in the sentences and connectives with any content. More generally, trying to generate exactly a given semantics

Seg := SPunct Seg | Seg SPunct | SPunct |
        on the one hand Seg *on the other hand* Seg |
        not only Seg *but also* Seg
SPunct := S Punctuation
Punctuation := . | ; | : | ? | !
S := S Coord S | S Subord S | Subord S S | Sadv S |
        NP Sadv VP | S Sadv | . . .
Coord := *and* | *or* | *but* | *so*
Subord := *although* | *after* | *because* | *before* | ...
Sadv := DAdv | SimpleAdv
DAdv := *instead* | *otherwise* | *for example* | *meanwhile* | ...
SimpleAdv := *yesterday* | *today* | *surprisingly* | *hopefully* | ...

Figure 6
PS rules for a discourse grammar.

when semantics underspecifies syntactic dependency (as discourse semantics must, on our account) is known to be intractable (Koller and Striegnitz 2002). An effective solution is to generate semantics and syntax simultaneously, which is straightforward with a lexicalized grammar (Stone et al. 2001).

Given the importance of various types of inference in discourse understanding, there is a second argument for using a lexicalized discourse grammar that derives from the role of implicature in discourse. Gricean reasoning about implicatures requires a hearer be able to infer the meaningful alternatives that a speaker had in composing a sentence. With lexicalization, these alternatives can be given by a grammar, allowing the hearer, for example, to ask sensible questions like "Why did the speaker say 'instead' here instead of nothing at all?" and draw implicatures from this. A CF PSG, on the other hand, might suggest questions like "Why did the speaker say two sentences rather than one here?" which seem empirically not to lead to any real implicatures. (On the contrast between choices, which seem to lead to implicatures, and mere alternative linguistic formulations, which do not seem to, see, for example, Dale and Reiter [1995] and Levison [2000].)

In several previous papers (Webber, Knott, and Joshi, 2001; Webber et al., 1999a, 1999b), we described how our approach fits into the framework of tree-adjoining grammar. This led to the initial version of a discourse parser (Forbes et al. 2001) in which the same parser that builds trees for individual clauses using clause-level LTAG trees then combines them using discourse-level LTAG trees. Here we simply outline the grammar, called DLTAG (section 5.1), then show how it supports the approach to structural and anaphoric discourse connectives presented earlier (section 5.2).

(Of course, one still needs to account for how speakers realize their intentions through text and how what is achieved through a single unit of text contributes to what a speaker hopes to achieve through any larger unit in which it is embedded. Preliminary accounts are given in Grosz and Sidner [1990] and Moser and Moore [1996]. Given the complex relation between individual sentences and speaker intentions, however, it is unlikely that the relation between multisentence discourse and speaker intentions can be modeled in a straightforward way similar to the basically monotonic compositional process that we have discussed in this article for discourse semantics.)

α: *subconj_mid* α: *subconj_pre*
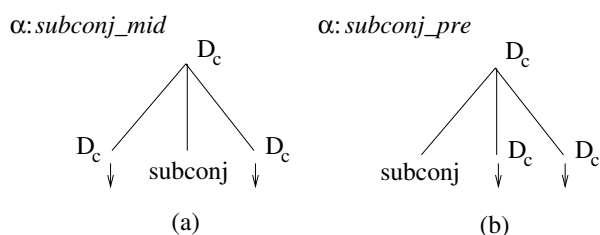
Figure 7
Initial trees for a subordinate conjunction: (a) postposed; (b) preposed. D_c stands for *discourse clause*, ↓ indicates a substitution site, and *subconj* stands for the particular subordinate conjunction that anchors the tree.

5.1 DLTAG and Discourse Syntax

A lexicalized TAG begins with the notion of a lexical **anchor**, which can have one or more associated tree structures. For example, the verb *likes* anchors one tree corresponding to *John likes apples*, another corresponding to the topicalized *Apples John likes*, a third corresponding to the passive *Apples are liked by John*, and others as well. That is, there is a tree for each minimal syntactic construction in which *likes* can appear, all sharing the same predicate-argument structure. This syntactic/semantic encapsulation is possible because of the extended domain of locality of LTAG.

A lexicalized TAG contains two kinds of **elementary** trees: **initial** trees that reflect basic functor-argument dependencies and **auxiliary** trees that introduce recursion and allow elementary trees to be modified and/or elaborated. Unlike the wide variety of trees needed at the clause level, we have found that extending a lexicalized TAG to discourse requires only a few elementary tree structures, possibly because clause-level syntax exploits structural variation in ways that discourse doesn't.

5.1.1 Initial Trees. DLTAG has initial trees associated with subordinate conjunctions, with parallel constructions, and with some coordinate conjunctions. We describe each in turn.

In the large LTAG developed by the XTAG project (XTAG-Group 2001) subordinate clauses are seen as **adjuncts** to sentences or verb phrases (i.e., as auxiliary trees) because they are outside the domain of locality of the verb. In DLTAG, however, it is predicates on clausal arguments (such as coordinate and subordinate conjunctions) that define the domain of locality. Thus, at this level, these predicates anchor initial trees into which clauses substitute as arguments. Figure 7 shows the initial trees for (a) postposed subordinate clauses and (b) preposed subordinate clauses.[16] At both leaves and root is a **discourse clause** (D_c): a clause or a structure composed of discourse clauses.

One reason for taking something to be an initial tree is that its local dependencies can be stretched long distance. At the sentence level, the dependency between *apples* and *likes* in *Apples John likes* is localized in all the trees for *likes*. This dependency can be stretched long distance, as in *Apples, Bill thinks John may like*. In discourse, as we noted in section 2, local dependencies can be stretched long distance as well, as in

(59) a. Although John is generous, he's hard to find.

---

16 Although in an earlier paper (Webber and Joshi 1998), we discuss reasons for taking the lexical anchors of the initial trees in Figures 7 and 8 to be feature structures, following the analysis in Knott (1996) and Knott and Mellish (1996), here we just take them to be specific lexical items.
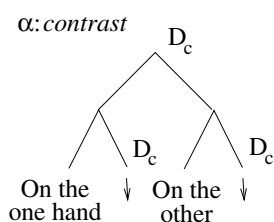
α: *contrast*

Figure 8
An initial tree for parallel constructions. This particular tree is for a contrastive construction anchored by *on the one hand* and *on the other hand*.

b. Although John is generous—for example, he gives money to anyone who asks him for it—he's hard to find.

(60) a. On the one hand, John is generous. On the other hand, he's hard to find.

b. On the one hand, John is generous. For example, suppose you needed some money: You'd only have to ask him for it. On the other hand, he's hard to find.

Thus DLTAG also contains initial trees for parallel constructions as in (60). Such an initial tree is shown in Figure 8. Like some initial trees in XTAG (XTAG-Group 2001), such trees can have a pair of anchors. Since there are different ways in which discourse units can be parallel, we assume a different initial tree for *contrast* (*on the one hand. . . on the other (hand). . .*), *disjunction* (*either. . . or. . .*), *addition* (*not only. . . but also. . .*), and *concession* (*admittedly. . . but. . .*).

Finally, there are initial trees for structural connectives between adjacent sentences or clauses that convey a particular relation between the connected units. One clear example is *so*, conveying *result*. Its initial tree is shown in Figure 9. We will have a better sense of what other connectives to treat as structural as a result of annotation efforts of the sort described in Creswell et al. (2002).[17]

5.1.2 Auxiliary Trees. DLTAG uses auxiliary trees in two ways: (1) for discourse units that continue a description in some way, and (2) for discourse adverbials. Again we describe each in turn.

---

17 For example, one might also have initial trees for **marked** uses of *and* and *or* that have a specific meaning beyond simple conjunction or disjunction, as in

(61) a. Throw another spit ball *and* you'll regret it.

      b. Eat your spinach *or* you won't get dessert.

These differ from the more frequent, simple coordinate uses of *and* and *or* in that the second conjunct in these marked cases bears a discourse relation to the first conjunct (*result* in both (61a) and (61b)). With simple coordinate uses of *and* and *or*, all conjuncts (disjuncts) bear the same relation to the same immediately left-adjacent discourse unit. For example, in (62), each conjunct is a separate *explanation* for not trusting John, wheras in (63), each disjunct conveys an alternative result of John's good fortune:

(62) You shouldn't trust John. He never returns what he borrows, *and* he bad-mouths his associates behind their backs.

(63) John just won the lottery. So he will quit his job, *or* he will at least stop working overtime.

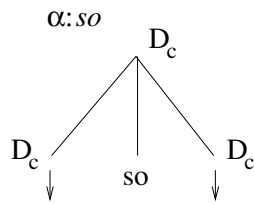For simple coordinate uses of *and* and *or*, we have auxiliary trees (section 5.1.2).

575

α: *so*

Figure 9
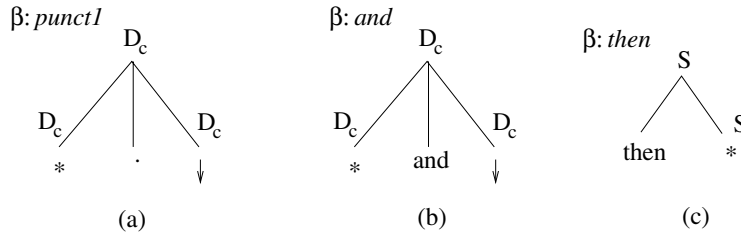Initial tree for coordinate conjunction *so*.

Figure 10
Auxiliary trees for basic elaboration. These particular trees are anchored by (a) the punctuation mark "period" and (b) *and*. The symbol ∗ indicates the **foot node** of the auxiliary tree, which has the same label as its root. (c) Auxiliary tree for the discourse adverbial *then*.
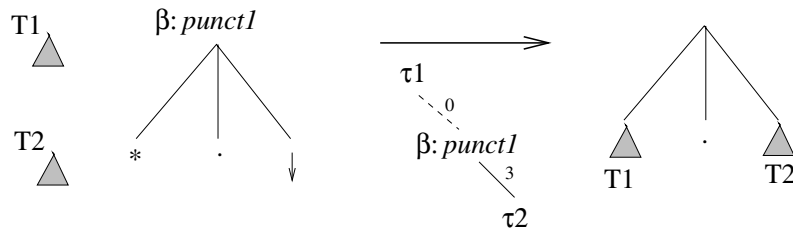
Figure 11
TAG derivation of example (64).

First, auxiliary trees anchored by punctuation (e.g., *period*, *comma*, *semicolon*.) (Figure 10a) or by **simple coordination** (Figure 10b) are used to provide further description of a situation or of one or more entities (objects, events, situations, states, etc.) within the situation.[18] The additional information is conveyed by the discourse clause that fills its substitution site. Such auxiliary trees are used in the derivation of simple discourses such as

(64) a. John went to the zoo.
      b. He took his cell phone with him.

Figure 11 shows the DLTAG derivation of example (64), starting from LTAG derivations of the individual sentences.[19] To the left of the horizontal arrow are the elementary trees to be combined: T1 stands for the LTAG tree for clause (64a), T2 for clause

---

18 The latter use of an auxiliary tree is related to **dominant topic chaining** in Scha and Polanyi (1988) and **entity chains** in Knott et al. (2001).
19 We comment on left-to-right incremental construction of DLTAG structures in parallel with sentence-level LTAG structures at the end of Section 5.2.

(64b), and *β:punct1* for the auxiliary tree assocated with the period after (64a). In the derivation, the foot node of *β:punct1* is adjoined to the root of T1 and its substitution site filled by T2, resulting in the tree to the right of the horizontal arrow. (A standard way of indicating TAG **derivations** is shown under the horizontal arrow, where broken lines indicate adjunction and solid lines, substitution. Each line is labeled with the address of the argument at which the operation occurs. $\tau 1$ is the derivation tree for T1 and $\tau 2$, the derivation tree for T2.)

The other auxiliary trees used in the lexicalized discourse grammar are those for discourse adverbials, which are simply auxiliary trees in a sentence-level LTAG (XTAG-Group 2001), but with an interpretation that projects up to the discourse level. An example is shown in Figure 10c. Adjoining such an adverbial to a clausal/sentential structure contributes to how information conveyed by that structure relates to the previous discourse.

There is some lexical ambiguity in this grammar, but no more than serious consideration of adverbials and conjunctions demands. First, as already noted, discourse adverbials have other uses that may not be anaphoric (65a–b) and may not be clausal (65a–c):

(65) a. John ate an apple *instead* of a pear.
      b. *In contrast* with Sue, Fred was tired.
      c. Mary was *otherwise* occupied.

Second, many of the adverbials found in second position in parallel constructions (e.g., *on the other hand, at the same time, nevertheless*) can also serve as simple adverbial discourse connectives on their own. In the first case, they will be one of the two anchors of an initial tree (Figure 8), and in the second, they will anchor a simple auxiliary tree (Figure 10c). These lexical ambiguities correlate with structural ambiguity.

5.2 Example Derivations

It should be clear by now that our approach aims to explain discourse semantics in terms of a product of the same three interpretive mechanisms that operate within clause-level semantics:

- compositional rules on syntactic structure (here, discourse structure)
- anaphor resolution
- inference triggered by adjacency and structural connection

For the compositional part of semantics in DLTAG (in particular, computing interpretations on derivation trees), we follow Joshi and Vijay-Shanker (2001). Roughly, they compute interpretations on the derivation tree using a bottom-up procedure. At each level, function application is used to assemble the interpretation of the tree from the interpretation of its root node and its subtrees. Where multiple subtrees have function types, the interpretation procedure is potentially nondeterministic: The resulting ambiguities in interpretation may be admitted as genuine, or they may be eliminated by a lexical specification. Multicomponent TAG tree sets are used to provide an appropriate compositional treatment for quantifiers, which we borrow for interpreting *for example* (examples (66c–d)).

In showing how DLTAG and an interpretative process on its derivations operate, we must, of necessity, gloss over how inference triggered by adjacency or associated with a structural connective provides the intended relation between adjacent discourse

units: It may be a matter simply of statistical inference, as in Marcu and Echihabi (2002), or of more complex inference, as in Hobbs et al. (1993). As we noted, our view is that there are three mechanisms at work in discourse semantics, just as there are in clause-level semantics: Inference isn't the only process involved. Thus the focus of our presentation here is on how compositional rules and anaphor resolution (which itself often appears to require inference) operate together with inference to yield discourse semantics.

   We start with previous examples (44) (here (66c)) and (47) (here (66d)) and two somewhat simpler variants (66a–b):

(66) a. You shouldn't trust John because he never returns what he borrows.

   b. You shouldn't trust John. He never returns what he borrows.

   c. You shouldn't trust John because, for example, he never returns what he borrows.

   d. You shouldn't trust John. For example, he never returns what he borrows.

This allows us to show how (66a–b) and (66c–d) receive similar interpretations, despite having somewhat different derivations, and how the discourse adverbial *for example* contributes both syntactically and semantically to those interpretations.

   We let *T*1 stand for the LTAG parse tree for *you shouldn't trust John*, τ1, for its derivation tree, and interp(*T*1), for the eventuality associated with its interpretation. Similarly, we let *T*2 stand for the LTAG parse tree for *he never returns what he borrows*, τ2, for its derivation tree, and interp(*T*2), for the eventuality associated with its interpretation.

   Example (66a) involves an initial tree (α:*because-mid*) anchored by *because* (Figure 12). Its derived tree comes from *T*1 substituting at the left-hand substitution site of α:*because-mid* (index 1) and *T*2 at its right-hand substitution site (index 3). Compositional interpretation of the resulting derivation tree yields *explanation*(interp(*T*2),interp(*T*1)). (A more precise interpretation would distinguish between the direct and epistemic causality senses of *because*, but the derivation would proceed in the same way.)

   In contrast with (66a), example (66b) employs an auxiliary tree (β:*punct1*) anchored by a period (Figure 13). Its derived tree comes from *T*2 substituting at the right-hand substitution site (index 3) of β:*punct1*, and β:*punct1* adjoining at the root of *T*1 (index 0). Compositional interpretation of the derivation tree yields merely that *T*2 continues the description of the situation associated with *T*1, that is, *elaboration*(interp(*T*2),interp(*T*1)). Further inference triggered by adjacency and structural connection leads to a con-
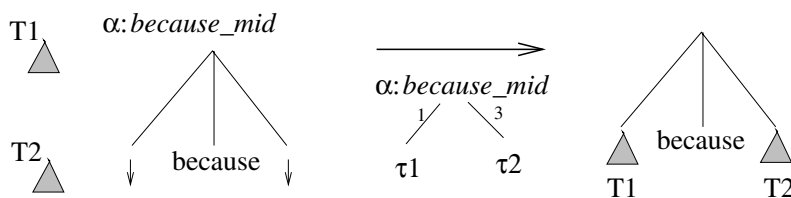

Figure 12
Derivation of example (66a). The derivation tree is shown below the arrow, and the derived tree, to its right. (Node labels D$_c$ have been omitted for simplicity.)
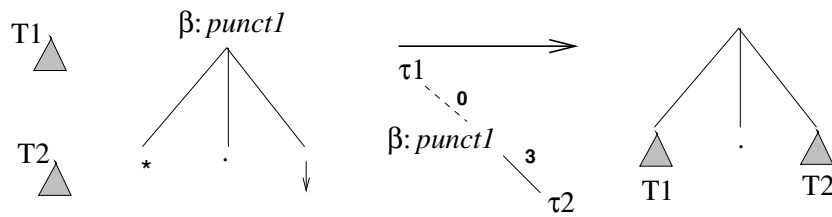
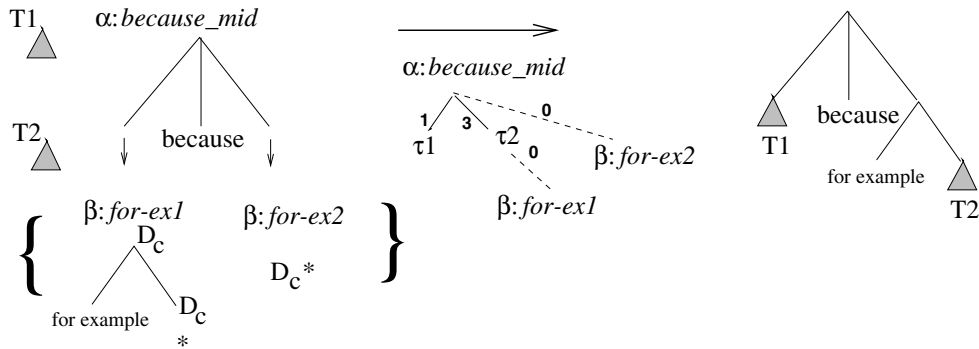**Figure 13**
Derivation of example (66b).

Figure 14
Derivation of example (66c).

clusion of causality between them, that is, *explanation*(interp(*T2*),interp(*T1*)), but this conclusion is defeasible because it can be denied without a contradiction: for example,

(67) You shouldn't trust John. He never returns what he borrows. But that's not why you shouldn't trust him.

Example (66c) differs from (66a) in containing *for example* in its second clause. As noted earlier, *for example* resembles a quantifier with respect to its semantics, as its interpretation takes wider scope than would be explained by its syntactic position. We handle this in the same way that quantifiers are handled in Joshi and Vijay-Shanker (2001) by associating with *for example* a two-element TAG tree set (Figure 14). Both trees in the tree set participate in the derivation: The auxiliary tree β:*for_ex1* adjoins at the root of *T2*, whereas the auxiliary tree β:*for_ex2* adjoins at the root of the higher discourse unit. Since we saw from example (66a) that the interpretation of this higher discourse unit is *explanation*(interp(*T2*),interp(*T1*)), the interpretation associated with the adjoined β:*for_ex2* node both embeds and abstracts this interpretation, yielding

exemplification(interp(*T2*), λX . *explanation*(X,interp(*T1*)))

That is, John's never returning what he borrows is one instance of a set of explanations.

Similarly, example (66d) differs from (66b) in containing *for example* in its second sentence (Figure 15). As in example (66b), an inferred relation is triggered between the interpretations of *T2* and *T1*, namely, *explanation*(interp(*T2*),interp(*T1*)). Then, as a result of β:*for_ex1* adjoining at *T2* and β:*for_ex2* adjoining at the root of the higher
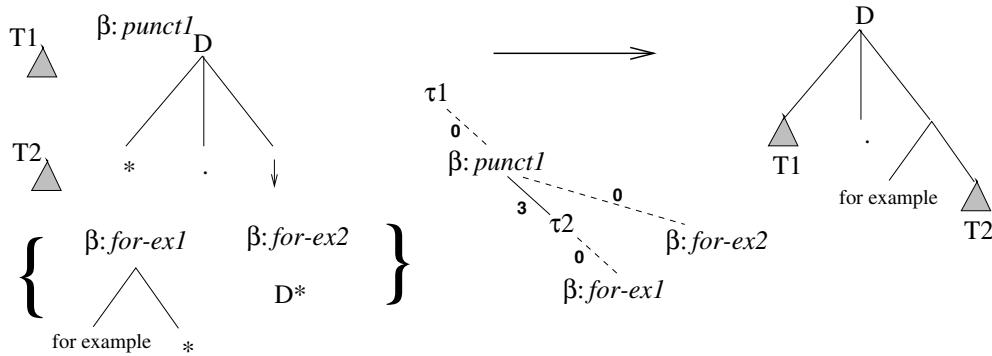
Figure 15
Derivation of example (66d).

discourse unit, *for example* again contributes the interpretation

$$\textit{exemplification}(\text{interp}(T2),\ \lambda X\ .\ \textit{explanation}(X, \text{interp}(T1))$$

Thus (66c) and (66d) differ only in the derivation of the interpretation that *for example* then abstracts over.

The next example we will walk through is example (11) (repeated here as example (68)):

(68) John loves Barolo. So he ordered three cases of the '97. But he had to cancel the order because *then* he discovered he was broke.

As shown in Figure 16, this example involves two initial trees (α:*so*, α:*because_mid*) for the structural connectives *so* and *because*; an auxiliary tree for the structural connective *but* (β:*but*), since *but* functions as a simple conjunction to continue the description of the situation under discussion; an auxiliary tree (β:*then*) for the discourse adverbial *then*; and initial trees for the four individual clauses $T1$–$T4$. As can be seen from the derivation tree, $T1$ and $T2$ substitute into α:*so* as its first and third arguments, and β:*but* root-adjoins to the result. The substitution argument of β:*but* is filled by α:*because_mid*, with $T3$ and $T4$ substituted in as its first and third arguments, and β:*then* is root-adjoined to $T4$. The interpretation contributed by *then*, after its anaphoric argument is resolved to interp($T2$), is

$$\iota 4:\ \textit{after}(\text{interp}(T4),\ \text{interp}(T2))$$

The interpretations derived compositionally from the structural connectives *so*, *because*, and *but* are

$$\iota 1:\ \textit{result}(\text{interp}(T2),\ \text{interp}(T1))$$
$$\iota 2:\ \textit{explanation}(\text{interp}(T4),\ \text{interp}(S3))$$
$$\iota 3:\ \textit{elaboration}(\iota 2, \iota 1)$$

Further inference may then refine *elaboration* to *contrast*, based on how *but* is being used.

Finally, we want to point out one more way in which texts that seem to be close paraphrases get their interpretations in different ways. Consider the two texts in example (69):
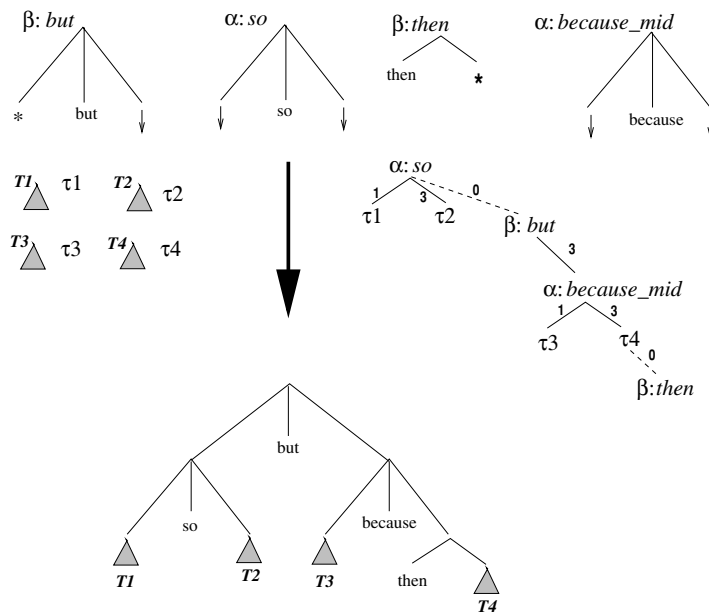
Figure 16
Derivation of example (68).

(69) a. You should eliminate part 2 before part 3 because part 2 is more sus-
          ceptible to damage.

      b. You should eliminate part 2 before part 3. This is because part 2 is
          more susceptible to damage.

Example (69b) is a simpler version of an example in Moser and Moore (1995), in which
This is because is treated as an unanalyzed cue phrase, no different from *because* in (69a).
We show here that this isn't necessary: One can analyze (69b) using compositional
semantics and anaphor resolution and achieve the same results.

     First consider (69a). Given the interpretations of its two component clauses, its
overall interpretation follows in the same way as (66a), shown in Figure 12. Now
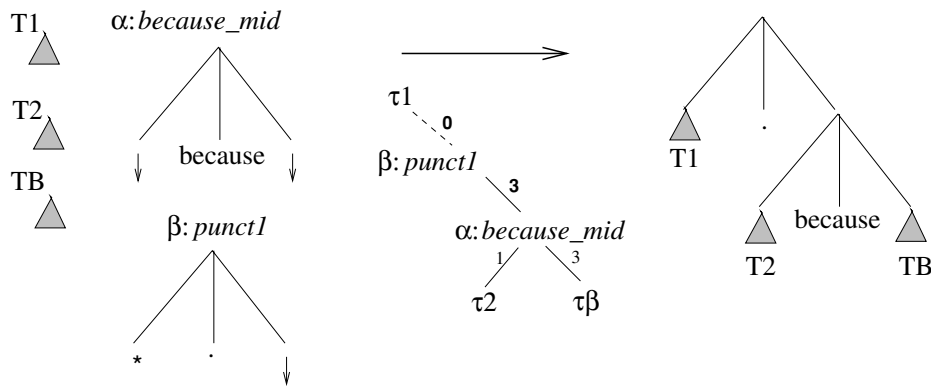consider (69b) and the derivation shown in Figure 17. Here the initial tree α:*because_mid*

Figure 17
Derivation of example (69b).

has its two arguments filled by *T2*, the TAG analysis of *this is* and *TB*, the TAG analysis of *part 2 is more susceptible to damage*. The overall derived tree for (69b) comes from *β:punct1* root-adjoining to *T1* (the TAG analysis of *You should eliminate part 2 before part 3*), with the subsitution site of *β:punct1* filled by the *α:because_mid* derivation. The compositional interpretation of the derivation tree yields the interpretation of the *α:because_mid* tree (*i1*) as an *elaboration* of the interpretation of *T1*:

> *i1*: *explanation*(interp(*TB*),interp(*T2*))
> *i2*: *elaboration*(*i1*,interp(*T1*))

But this is not all. The pronoun *this* in *T2* is resolved anaphorically to the nearest consistent eventuality (Eckert and Strube 2000; Byron 2002) which in this case is interp(*T1*). Taking this as the interpretation of *T2* and substituting, we get

> *i1*: *explanation*(interp(*TB*),interp(*T1*))
> *i2*: *elaboration*(*i1*,interp(*T1*))

Notice that *i1* is also the interpretation of (69a). To this, *i2* adds the somewhat redundant information that *i1* serves to elaborate the advice in *T1*. Thus (69a) and (69b) receive similar interpretations but by different means. This treatment has the added advantage that one does not have to treat *This is not because* as a separate cue phrase. Rather, negation simply produces

> *i1*: ¬*explanation*(interp(*TB*),interp(*T1*))
> *i2*: *elaboration*(*i1*,interp(*T1*))

That is, *T1* is elaborated by a denial of a (possible) explanation. Presumably, the text would go on to provide the actual explanation.

Finally, we want to comment on the Holy Grail of discourse parsing: a realistic model that is computed in parallel with incremental sentence-level parsing. Neither the analyses given in this section nor the discourse parsing described in Forbes et al. (2001) is done in a left-to-right incremental fashion, in parallel with incremental left-to-right sentence-level parsing.

What would an integrated incremental method of sentence-discourse processing require? At minimum, we believe it would involve:

- A left-to-right parser that would simultaneously compute increments to sentence-level syntactic structure, sentence-level semantics, discourse-level syntactic structure, and discourse-level semantics. Increments to the latter two would occur only at clause boundaries and with discourse adverbials and structural connectives.

- An incremental anaphor resolution mechanism, similar to that in Strube (1998), but extended both to deictic pronouns, as in Eckert and Strube (2000) and Byron (2002) and to the anaphoric argument of discourse adverbials.

- Incremental computation of discourse structure in terms of *elaboration* relations and further nondefeasible reasoning to more specific relations, where possible.

A left-to-right parser that simultaneously produces sentence-level syntactic and semantic analyses already exists for combinatory categorial grammar (Steedman 1996,

2000b; Hockenmaier, Bierner, and Baldridge, forthcoming), and it would seem straight-forward to extend such a parser to computing discourse-level syntax and semantics as well. Similarly, it seems straightforward to produce an incremental version of any of the current generation of anaphor resolution mechanisms, extended to deictic pronouns, although current approaches attempt to resolve *this* and *that* only with the interpretation of a single clause, not with that of any larger discourse unit. As these approaches are also not very accurate as yet, incremental anaphor resolution awaits improvements to anaphor resolution in general. Moreover, as we better understand the specific anaphoric properties of discourse adverbials through empirical analysis such as Creswell et al. (2000), such anaphor resolution mechanisms can be extended to include them as well.

As for building discourse structure incrementally in parallel with syntactic structure, there is no working prototype yet that will do what is needed. But we have no doubt that as psycholinguistics and computation together develop a better understanding of incremental semantic processing, researchers' desire for a working prototype will eventually result in the development of one.

6. Conclusion

In this article, we have argued that discourse adverbials make an anaphoric, rather than a structural, connection with the previous discourse (section 2), and we have provided a general view of anaphora in which it makes sense to talk of discourse adverbials as being anaphoric (section 3). We have then shown that this view of discourse adverbials allows us to characterize a range of ways in which the relation contributed by a discourse adverbial can interact with the relation conveyed by a structural connective or inferred through adjacency (section 4), and then illustrated how discourse syntax and semantics can be treated as an extension of sentence-level syntax and semantics, using a lexicalized discourse grammar (section 5).

We are clearly not the first to have proposed a grammatical treatment of low-level aspects of discourse semantics (Asher and Lascarides 1999; Gardent 1997; Polanyi and van den Berg 1996; Scha and Polanyi 1988; Schilder 1997a, 1997b; van den Berg 1996), but we are the first to have recognized that the key to avoiding problems of maintaining a compositional semantics for discourse lies in recognizing discourse adverbials as anaphors and not trying to shoehorn everything into a single class of discourse connectives. Although we are not yet able to propose a solution to the problem of correctly resolving discourse adverbials or a way of achieving the Holy Grail of computing discourse syntax and semantics in parallel with incremental sentence processing, the proposed approach does simplify issues of discourse structure and discourse semantics in ways that have not before been possible.

References
Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*. Kluwer, Boston.
Asher, Nicholas and Alex Lascarides. 1999. The semantics and pragmatics of

presupposition. *Journal of Semantics*, 15(3):239–300.

Asher, Nicholas and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge, England.

Asher, Nicholas and Michael Morreau. 1991. Commonsense entailment. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence IJCAI'91*, pages 387–392, Sydney, Australia.

Barwise, Jon and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.

Bateman, John. 1999. The dynamics of "surfacing": An initial exploration. In *Proceedings of International Workshop on Levels of Representation in Discourse (LORID'99)*, pages 127–133, Edinburgh.

Bierner, Gann. 2001a. Alternative phrases and natural language information retrieval. In *Proceedings of the 39*th *Annual Conference of the Association for Computational Linguistics*, Toulouse, France, July.

Bierner, Gann. 2001b. *Alternative Phrases: Theoretical Analysis and Practical Application*. Ph.D. thesis, University of Edinburgh.

Bierner, Gann and Bonnie Webber. 2000. Inference through alternative set semantics. *Journal of Language and Computation*, 1(2):259–274.

Byron, Donna. 2002. Resolving pronominal reference to abstract entities. In *Proceedings of the 40*th *Annual Meeting, Association for Computational Linguistics*, pages 80–87, University of Pennsylvania.

Clark, Herbert. 1975. Bridging. In *Proceedings of Theoretical Issues in Natural Language Processing (TINLAP-1)*, pages 169–174, Cambridge, MA.

Clark, Herbert and Catherine Marshall. 1981. Definite reference and mutual knowledge. In Aravind Joshi, Bonnie Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, Cambridge, England, pages 10–63.

Cosse, Michel. 1996. Indefinite associative anaphora in French. In *Proceedings of the IndiAna Workshop on Indirect Anaphora*, University of Lancaster, Lancaster, England.

Creswell, Cassandre, Kate Forbes, Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2002. The discourse anaphoric properties of connectives. In *Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium*, Lisbon, Portugal.

Dale, Robert. 1992. *Generating Referring Expressions*. MIT Press, Cambridge, MA.

Dale, Robert and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 18:233–263.

Eckert, Miriam and Michael Strube. 2000. Synchronising units and anaphora resolution. *Journal of Semantics*, 17:51–89.

Forbes, Katherine, Eleni Miltsakaki, Rashmi Prasad, Anoop Sarkar, Aravind Joshi, and Bonnie Webber. 2001. D-LTAG system: Discourse parsing with a lexicalized tree-adjoining grammar. In *ESSLLI'2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics*, Helsinki, Finland.

Forbes, Kate and Bonnie Webber. 2002. A semantic account of adverbials as discourse connectives. In *Proceedings of Third SIGDial Workshop*, pages 27–36, Philadelphia, PA.

Frank, Anette and Hans Kamp. 1997. On context dependence in modal constructions. In *SALT-97*, Stanford, CA.

Gardent, Claire. 1997. Discourse tree adjoining grammars. Claus Report no. 89, University of the Saarland, Saarbrücken, Germany.

Grosz, Barbara and Candace Sidner. 1990. Plans for discourse. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, pages 417–444.

Hahn, Udo, Katja Markert, and Michael Strube. 1996. A conceptual reasoning approach to textual ellipsis. In *Proceedings of the 12*th *European Conference on Artificial Intelligence*, pages 572–576, Budapest, Hungary.

Hardt, Dan. 1992. VP ellipsis and contextual interpretation. In *Proceedings of International Conference on Computational Linguistics(COLING-92)*, pages 303–309, Nantes.

Hardt, Dan. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, 22:187–221.

Hellman, Christina and Kari Fraurud. 1996. *Proceedings of the IndiAna Workshop on Indirect Anaphora*. University of Lancaster, Lancaster, England.

Hobbs, Jerry. 1985. Ontological promiscuity. In *Proceedings of the 23*rd *Annual Meeting of the Association for Computational Linguistics*, pages 61–69, Palo Alto, CA. Morgan Kaufmann.

Hobbs, Jerry. 1990. *Literature and Cognition*. Volume 21 of CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, CA.

Hobbs, Jerry, Mark Stickel, Paul Martin, and Douglas Edwards. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142.

Hockenmaier, Julia, Gann Bierner, and Jason Baldridge. Forthcoming. Providing robustness for a CCG system. *Journal of Language and Computation*.

Isard, Stephen. 1975. Changing the context. In Edward Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge, England, pages 287–296.

Jayez, Jacques and Corinne Rossari. 1998a. Pragmatic connectives as predicates. In Patrick Saint-Dizier, editor, *Predicative Structures in Natural Language and Lexical Knowledge Bases*. Kluwer Academic, Dordrecht, the Netherlands, pages 306–340.

Jayez, Jacques and Corinne Rossari. 1998b. The semantics of pragmatic connectives in TAG: The French *donc* example. In Anne Abeillé and Owen Rambow, editors, *Proceedings of the TAG+4 Conference*. CSLI Publications, Stanford, CA.

Joshi, Aravind. 1987. An introduction to tree adjoining grammar. In Alexis Manaster-Ramer, editor, *Mathematics of Language*. John Benjamins, Amsterdam, pages 87–114.

Joshi, Aravind and K. Vijay-Shanker. 2001. Compositional semantics with lexicalized tree-adjoining grammar (LTAG): How much underspecification is necessary? In Harry Bunt, Reinhard Muskens, and Elias Thijsse, editors, *Computing Meaning*, Volume 2, Kluwer, Dordrecht, the Netherlands, pages 147–163.

Jurafsky, Dan and James Martin. 2000. *Speech and Language Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer, Dordrecht, the Netherlands.

Kehler, Andrew. 2002. *Coherence, Reference and the Theory of Grammar*. CSLI Publications, Stanford, CA.

Kibble, Rodger. 1995. Modal insubordination. In *Empirical Issues in Formal Syntax and Semantics, Selected Papers from the Colloque de Syntaxe et de Sémantique de Paris*, pages 317–332.

Knott, Alistair. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Ph.D. thesis, Department of Artificial Intelligence, University of Edinburgh.

Knott, Alistair and Chris Mellish. 1996. A feature-based account of the relations signalled by sentence and clause connectives. *Language and Speech*, 39(2–3):143–183.

Knott, Alistair, Jon Oberlander, Mick O'Donnell, and Chris Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In T. Sanders, J. Schilperoord, and W. Spooren, editors, *Text Representation: Linguistic and Psycholinguistic Aspects*. John Benjamins, Amsterdam, pages 181–196.

Koller, Alexander and Kristina Striegnitz. 2002. Generation as dependency parsing. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 17–24, Philadelphia, PA.

Kruijff-Korbayová, Ivana and Bonnie Webber. 2001a. Concession, implicature and alternative sets. In *Fourth International Workshop on Computational Semantics*, Tilburg, the Netherlands.

Kruijff-Korbayová, Ivana and Bonnie Webber. 2001b. Information structure and the semantics of "otherwise." In *ESSLLI'2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics*, pages 61–78, Helsinki, Finland.

Lagerwerf, Luuk. 1998. *Causal Connectives Have Presuppositions*. Holland Academic Graphics, The Hague, the Netherlands.

Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge, MA.

Luperfoy, Susann. 1992. The representation of multimodal user interface dialogues using discourse pegs. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 22–31, University of Delaware, Newark.

Mann, William and Sandra Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Marcu, Daniel. 1999. Instructions for manually annotating the discourse structure of texts. Available from http://www.isi.edu/~marcu.

Marcu, Daniel and Abdessamad Echihabi. 2002. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th Annual Meeting, Association for Computational Linguistics*, pages 368–375, University of Pennsylvania, Philadelphia.

Modjeska, Natalia Nygren. 2001. Towards a resolution of comparative anaphora: A corpus study of "other." In *PAPACOL*, Italy.

Modjeska, Natalia Nygren. 2002. Lexical and grammatical role constraints in resolving *other*-anaphora. In *Proceedings of*

the Discourse Anaphora and Anaphor Resolution Colloquium, Lisbon, Portugal.

Moens, Marc and Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14(1):15–28.

Moore, Johanna and Martha Pollack. 1992. A problem for RST: The need for multi-level discouse analysis. *Computational Linguistics*, 18(4):537–544.

Moser, Megan and Johanna Moore. 1995. Investigating cue selection and placement in tutorial discourse. In *Proceedings of the 33rd Annual Meeting, Association for Computational Linguistics*, pages 130–135, MIT, Cambridge, MA.

Moser, Megan and Johanna Moore. 1996. Toward a synthesis of two accounts of discourse structure. *Computational Linguistics*, 22(3):409–419.

Not, Elena, Lucia Tovena, and Massimo Zancanaro. 1999. Positing and resolving bridging anaphora in deverbal NPs. In *ACL'99 Workshop on the Relationship between Discourse/Dialogue Structure and Reference*, College Park, MD.

Partee, Barbara. 1984. Nominal and temporal anaphora. *Linguistics and Philosophy*, 7(3):287–324.

Polanyi, Livia and Martin H. van den Berg. 1996. Discourse structure and discourse interpretation. In P. Dekker and M. Stokhof, editors, *Proceedings of the Tenth Amsterdam Colloquium*, pages 113–131, University of Amsterdam.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness and information-status. In Susan Thompson and William Mann, editors, *Discourse Description: Diverse Analyses of a Fundraising Text*. John Benjamins, Amsterdam, pages 295–325.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartik. 1972. *A Grammar of Contemporary English*. Longman, Harlow, England.

Scha, Remko and Livia Polanyi. 1988. An augmented context free grammar for discourse. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*, pages 573–577, Budapest, Hungary, August.

Schilder, Frank. 1997a. Towards a theory of discourse processing: Flashback sequences described by D-trees. In *Proceedings of the Formal Grammar Conference (ESSLLI'97)*, Aix-en-Provence, France, August.

Schilder, Frank. 1997b. Tree discourse grammar, or how to get attached to a discourse. In *Proceedings of the Second International Workshop on Computational Semantics*, Tilburg, the Netherlands,

January.

Steedman, Mark. 1996. *Surface Structure and Interpretation*. Volume 30 of *Linguistic Inquiry Monograph*, 5, MIT Press, Cambridge, MA.

Steedman, Mark. 2000a. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 34:649–689.

Steedman, Mark. 2000b. *The Syntactic Process*. MIT Press, Cambridge, MA.

Stokhof, Martin and Jeroen Groenendijk. 1999. Dynamic semantics. In Robert Wilson and Frank Keil, editors, *MIT Encyclopedia of Cognitive Science*. MIT Press. Cambridge, MA, pages 247–249

Stone, Matthew, Christine Doran, Bonnie Webber, Tonia Bleam, and Martha Palmer. 2001. Microplanning from communicative intentions: Sentence planning using descriptions (SPUD). Technical Report no. RUCCS TR68, Department of Cognitive Science, Rutgers University, New Brunswick, NJ.

Stone, Matthew and Daniel Hardt. 1999. Dynamic discourse referents for tense and modals. In Harry Bunt, editor, *Computational Semantics*. Kluwer, Dordrecht, the Netherlands, pages 287–299.

Strube, Michael. 1998. Never look back: An alternative to centering. In *Proceedings, COLING/ACL'98*, pages 1251–1257, Montreal, Quebec, Canada.

Traugott, Elizabeth. 1995. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at ICHL XII, Manchester, England. Revised version of (1997) available at http://www.stanford. edu/traugott/ect-papersonline.html.

Traugott, Elizabeth. 1997. The discourse connective under all: A historical pragmatic account. Paper presented at ICL, Paris. Available at http://www. stanford.edu/traugott/ect-papersonline.html.

van den Berg, Martin H. 1996. Discourse grammar and dynamic logic. In P. Dekker and M. Stokhof, editors, *Proceedings of the Tenth Amsterdam Colloquium*, pages 93–111, ILLC/Department of Philosophy, University of Amsterdam.

van Eijck, Jan and Hans Kamp. 1997. Representing discourse in context. In Jan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*. Elsevier Science B.V., Amsterdam, pages 181–237.

Venditti, Jennifer J., Matthew Stone, Preetham Nanda, and Paul Tepper. 2002. Discourse constraints on the

interpretation of nuclear-accented
pronouns. In *Proceedings of Symposium on
Speech Prosody*, Aix-en-Provence, France.
Available at http://www.lpl.
univ-aix.fr/sp2002/papers.htm.

Vendler, Zeno. 1967. *Linguistics in Philosophy*.
Cornell University Press, Ithaca, NY.

Webber, Bonnie. 1988. Tense as discourse
anaphor. *Computational Linguistics*,
14(2):61–73.

Webber, Bonnie. 1991. Structure and
ostension in the interpretation of
discourse deixis. *Language and Cognitive
Processes*, 6(2):107–135.

Webber, Bonnie and Breck Baldwin. 1992.
Accommodating context change. In
*Proceedings of the 30th Annual Meeting of the
Association for Computational Linguistics
(ACL)*, pages 96–103, University of
Delaware, Newark.

Webber, Bonnie and Aravind Joshi. 1998.
Anchoring a lexicalized tree-adjoining
grammar for discourse. In *COLING/ACL
Workshop on Discourse Relations and
Discourse Markers*, pages 86–92, Montreal,
Quebec, Canada.

Webber, Bonnie, Alistair Knott, and Aravind
Joshi. 2001. Multiple discourse
connectives in a lexicalized grammar for
discourse. In Harry Bunt, Reinhard
Muskens, and Elias Thijsse, editors,
Computing Meaning, volume 2. Kluwer,
Dordrecht, the Netherlands, pages
229–249.

Webber, Bonnie, Alistair Knott, Matthew
Stone, and Aravind Joshi. 1999a.
Discourse relations: A structural and
presuppositional account using lexicalised
TAG. In *Proceedings of the 36th Annual
Meeting of the Association for Computational
Linguistics*, pages 41–48, College Park,
MD.

Webber, Bonnie, Alistair Knott, Matthew
Stone, and Aravind Joshi. 1999b. What
are little trees made of: A structural and
presuppositional account using lexicalised
TAG. In *Proceedings of International
Workshop on Levels of Representation in
Discourse (LORID'99)*, pages 151–156,
Edinburgh.

Wiebe, Janyce. 1993. Issues in linguistic
segmentation. In *Workshop on Intentionality
and Structure in Discourse Relations,
Association for Computational Linguistics*,
pages 148–151, Ohio State University.

Woods, William. 1978. Semantics and
quantification in natural language
question answering. In Marshall C. Yovits,
editor, *Advances in Computers*, volume 17.
Academic Press, New York, pages 1–87.

XTAG-Group. 2001. A lexicalized tree
adjoining grammar for English. Technical
Report no. IRCS 01-03, University
of Pennsylvania, Philadelphia. Available at
ftp://ftp.cis.upenn.edu/pub/ircs/technical-
reports/01-03.