

Anaphora and Discourse Semantics

Bonnie Webber*
Edinburgh University

Aravind Joshi
University of Pennsylvania

Matthew Stone
Rutgers University

Alistair Knott
University of Otago

We argue in this paper that many common adverbial phrases generally taken to be discourse connectives signalling discourse relations between adjacent discourse units are instead anaphors. We do this by (i) demonstrating their behavioral similarity with more common anaphors (pronouns and definite NPs); (ii) presenting a general framework for understanding anaphora into which they nicely fit; (iii) showing the interpretational benefits of understanding discourse adverbials as anaphors; and (iv) sketching out a lexicalised grammar that facilitates discourse interpretation as a product of compositional rules, anaphor resolution and inference.

Introduction

Several years ago, in an ACL workshop paper, Janyce Wiebe (1993) cited Example 1 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.

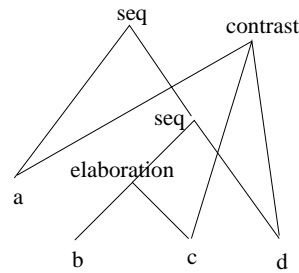
The problem she noted was that the discourse connectives *but* and *then* appear to link clause (1d) to two different things: “then” to clause (1b) – i.e., the car starting to turn right being the next relevant event after Chee’s finishing his tests – and “but” to a combination of clauses (1a) and (1c) – i.e., the car turning right failing the expectation of its continuing in the same direction and being the car that Chee is awaiting. (The former link is commonly called a *sequence* relation, and the latter, a form of *contrast*.)

These relations are usually taken to be the basis for low-level discourse structure, leading to something like Figure 1 for Example 1. This structure might seem advantageous in allowing the semantics of the example to be computed directly by compositional rules and defeasible inference. However, this structure is in fact a *DAG* – a directed acyclic graph.¹ 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.

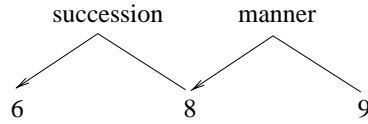
While we are not committed to discourse structure being a tree (e.g. Figure 2 from (Bateman, 1999)), we feel that the cost to discourse theory of moving to arbitrary DAGS for discourse structure is too great to be taken lightly. So we want to suggest another explanation for these and other examples of apparent complex and crossing dependencies

* Division of Informatics, University of Edinburgh, 2 Buccleuch Place, Edinburgh UK EH8 9LW.
E-mail: bonnie.webber@ed.ac.uk

¹ The structure in Figure 1 is labelled with the type of relation taken to hold and its “support” from either a connective (“but”) or adverbial (“then”). There are other possible structures for Example 1, but all of them are DAGS, so our point still holds.

**Figure 1**

Possible discourse structure for Example 1.

**Figure 2**

Simple multi-parent structure

in discourse: while *structural connectives* such as coordinating (e.g., “but”) and subordinating (e.g., “although”) conjunctions do indeed signal discourse relations between (the interpretation of) their conjuncts, *discourse adverbials* such as “then”, “otherwise”, and “nevertheless” are instead simply *anaphors*, signalling a relation between the interpretation of their matrix clause and the discourse context. We argue that understanding discourse adverbials as anaphors accomplishes four important goals:

1. It recognises their *behavioral similarity* with the pronouns and definite noun phrases (NPs) that are the “bread and butter” of previous work on anaphora (Section 1).
2. It contributes substance to the view, expressed for example by Carter (1987) that anaphora comprises more than just pronouns and definite NPs:

Anaphora is the special case of cohesion where the meaning (sense and/or reference) of one item in a cohesive relationship (the *anaphor*) is, in isolation, somehow vague and incomplete, and can only be properly interpreted by considering the meanings of the other item(s) in the relationship (the *antecedent(s)*). (Carter, 1987, page 33)

This is explored in Section 2.

3. It supports the direct computation of discourse semantics through *compositional* rules and defeasible inference. This is a goal that researchers have been struggling after for some time (Asher and Lascarides, 1999; Gardent, 1997; Kehler, 1995; Polanyi and van den Berg, 1996; Scha and Polanyi, 1988; Schilder, 1997a; Schilder, 1997b; van den Berg, 1996), and that Wiebe essentially recognises in her concern about the consequences for discourse structure of examples such as (1). Enabling anaphor resolution to contribute to meaning simplifies the process of compositional semantics² and directs

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

attention to how the meaning of discourse adverbials supports and complements other aspects of discourse semantics (Section 3).

4. It allows us to see more clearly how a lexicalised approach to the computation of clausal syntax and semantics extends naturally to the computation of discourse syntax and semantics, providing a single semantic matrix with which to associate speaker intentions and other aspects of pragmatics. (Section 4)

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) and Dynamic Semantics (Stokhof and Groenendijk, 1999) and with more detailed analyses of the meaning and use of individual discourse adverbials, such as (Jayez and Rossari, 1998a; Traugott, 1997): it provides what we believe to be a simpler and 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.

1 Discourse Adverbials as Anaphors

1.1 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”, “on the one hand ... on the other (hand)”, etc.) and *inferred relations* between adjacent discourse units (in the absence of an explicit structural connective). Here, adjacency is what triggers the inference. Consider the following example:

- (2) You shouldn’t trust John. He never returns what he borrows.

Adjacency leads the hearer to hypothesize that the second clause is related to its left-adjacent neighbor – and more specifically, that a form of rhetorical relation holds between the two. (We discuss this more in Section 4.) Our goal in this section is to convince the reader that many *discourse adverbials* – including “then”, “also”, “otherwise”, “nevertheless”, “instead” – behave not like structural connectives, but instead like anaphors.

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 (*sequence* or *after*) whose arguments are clausally-derived events.

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, 1999; Webber et al., 1999a; Webber et al., 1999b), structural connectives have two relevant properties: while they admit *stretching* of predicate-argument dependencies, they do not tolerate their *crossing*. This is most obvious in the case of preposed subordinate conjunctions (Example 3) or “paired” coordinate conjunctions (Example 4). With such connectives, the initial predicate signals that its two arguments will follow.

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

- (4) *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 5b.

- (5) a. Apples John likes.

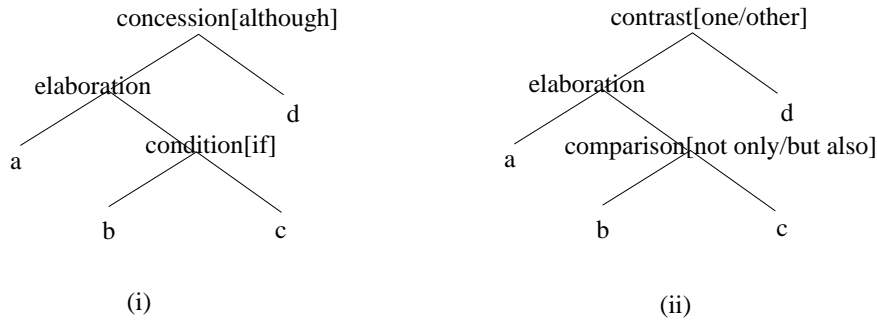


Figure 3
Discourse structures associated with (i) Example 6 and (ii) Example 7.

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

That this also happens with structural connectives and their arguments, is illustrated in Example 6 (in which the first clause of Example 3 is elaborated by another preposed subordinate-main clause construction embedded within it) and Example 7 (in which the first conjunct of Example 4 is elaborated by another paired-conjunction construction embedded within it). Possible discourse structures for these examples are given in Figure 3.

- (6) 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.
- (7) 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, another property of structural connectives is that they do not admit *crossing* of predicate-argument dependencies. If we do this with Examples 6 and 7, we get

- (8) a. Although John is very generous –
 b. if you need some money –
 c. he’s very hard to find –
 d. you only have to ask him for it.
- (9) 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 (9), “but” now appears to link (9d) with (9c), conveying that despite being allergic to beans, Fred eats them for breakfast and snacks. And while this might be inferred from (7), it is certainly not conveyed directly. As a consequence, we stipulate that structural connectives do not admit crossing of their predicate-argument dependencies.

That is not all. Since we take the basic level of discourse structure to be a consequence of (a) relations associated with explicit structural connectives and (b) relations

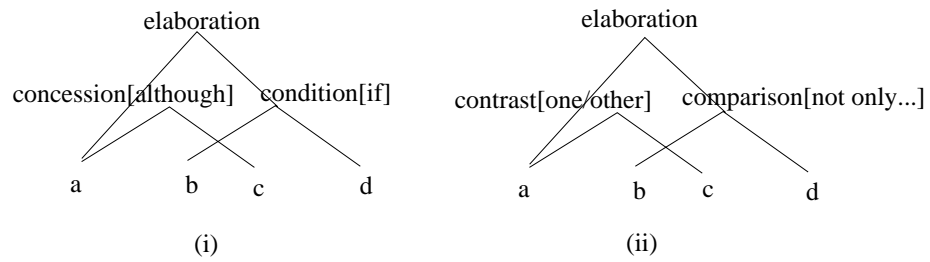
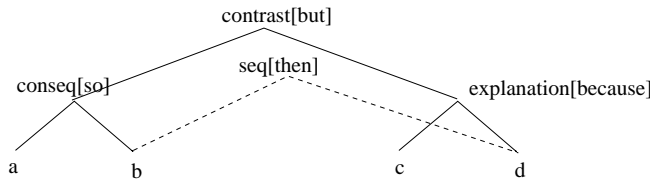


Figure 4
 (Impossible) discourse structures that would have to be associated with Example 8 (i) and with Example 9 (ii).

**Figure 5**

Example 10, with structural realisation of all dependencies

whose defeasible inference is triggered by adjacency, we stipulate that *low-level discourse structure itself does not admit crossing structural dependencies*. (In this sense, discourse structure may be truly simpler than sentence structure. To verify this, it might be useful to carefully 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. Example 10 shows this clearly. Clause 10(d) contains the discourse adverbial “then”. For it to get its first argument from (b) – i.e., the event that the discovery in (d) is “after”, it must cross the structural connection between clauses (c) and (d) associated “because”. This crossing dependency is illustrated in Figure 5.

- (10) 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.

But of course crossing dependencies are not unusual in discourse because anaphors (e.g., pronouns and definite NPs) do it all the time, for example:

- (11) Every man_i tells every woman_j he_i meets that she_j reminds him_i of his_i mother.

This suggests that in Example 10, the relationship between “then” and the previous discourse might usefully be taken to be anaphoric as well.

1.2 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 arguments. For example, in

- (12) 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; Webber and Baldwin, 1992).) The same is true of discourse adverbials. In

- (13) 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 – i.e., where your answer to the question is “no”. But this answer isn't there structurally: it is only inferred. While 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 involve a substantial commitment to covert constituents in discourse structure.

Secondly, attempts to paraphrase “otherwise” in terms of the structural connective “or” demonstrate that “otherwise” has a wider range of options.³ This is illustrated by the following pair of examples:

- (14) a. If the light is red, stop. *Otherwise* you’ll get a ticket.
 (If you do something other than stop, you’ll get a ticket.)
- b. If the light is red, stop. *Otherwise* go straight on.
 (If the light is not red, go straight on.)

Only one of these two ways of resolving “otherwise” in the context of a preceding if-construction can be paraphrased with “or” – that is, only the case where “otherwise” resolves to an alternative to the consequence clause, as in (14a) – cf.

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

Paraphrasing (14b) with “or”, as in

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

produces something whose meaning is quite different. Thus, “otherwise” has access to material that is not available to a structural connective. (Actually, in Section 4, we posit two separate lexico-syntactic entries for “or” as a structural connective – one for purely logical “or” and the other for “or” conveying an independent semantic relation between its arguments, as is the case here.)

Our final piece of evidence is that, like pronouns, these discourse adverbials can appear in an analogue of *donkey sentences*. Donkey sentences such as Example 15 have been used to argue the intrinsic discourse nature of pronominal anaphors: that pronouns are not merely a reflex of a syntactic binding operation.

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

In donkey sentences, anaphors appear in a structural and interpretive environment in which a direct syntactic relationship between anaphor and antecedent is normally impossible. Therefore, donkey sentences are evidence for interpreting an anaphor by accessing a discourse entity instead of by syntactic binding.

While no one has ever argued that discourse adverbials are a reflex of a syntactic binding operation – they have always been treated as elements of discourse interpretation, signalling relations between adjacent clauses – it is significant that they can appear in their own version of donkey sentences, as in

- (16) a. Anyone who has developed network software, has *then* had to hire a lawyer to protect his/her interests. (i.e., *after developing network software*)
- b. Many people who have developed network software, have *nevertheless* never gotten very rich. (i.e., *despite having developed network software*)
- c. Every person selling “The Big Issue” might *otherwise* be asking for spare change. (i.e., *if s/he weren’t selling “The Big Issue”*)

This suggests that discourse adverbials are accessing discourse entities (in particular, eventualities) rather than signalling a structural connection between adjacent clauses.⁴

³ This was pointed out independently by Natalia Modjeska, Lauri Karttunen, Mark Steedman, Robin Cooper and David Traum, on presentation of this work at ESSLLI’01 in Helsinki, August 2001.

⁴ While Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) 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 in (16c) and others, in which the material in an embedded clause (here, a relative clause) bears a semantic relation to its matrix clause. While this importantly recognises the phenomenon, it does not contribute to understanding its nature.

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 recognise them as anaphors: In the next section, we suggest a framework for anaphora in which discourse adverbials fit as neatly⁵ as pronouns and definite NPs.

2 A Framework for Anaphora

2.1 Discourse referents and anaphor interpretation

If we want to take discourse adverbials to be anaphors, we have to ask what kind, since on the surface, adverbials neither walk nor talk like the anaphors we are most familiar with – pronouns and definite NPs. All discourse anaphors involve, at the very least, an *anaphoric expression* α and one or more entities e_r from the discourse context or context of utterance⁶ that contribute in some way to the interpretation of α , e_α .

One thing we want to point out, although it is not critical to our discussion of discourse adverbials as anaphors, is that not all the material in the expression α may be anaphoric – i.e., interpreted with respect to e_r . For example, one type of expression that we take to be anaphoric is “other NPs”⁷, as in:

- (17) a. The new mayor of London has declared war on pigeons.
 b1. Other birds have not incurred his wrath.
 b2. Other birds that inhabit the city year round have not incurred his wrath.
 b3. Other birds with more sanitary habits have not incurred his wrath.
 b4. Other more sanitary birds have not incurred his wrath.

In (b1), one would obviously take the anaphoric expression to be the entire NP “other birds”. Its interpretation involves the entity e_r evoked by “pigeons” in (17a), which is excluded from the set of birds under consideration, which have not (we are told) incurred the mayor’s wrath. Similarly, in (b2), one would take the anaphoric expression to be the entire NP “other birds that inhabit the city year round”, with its interpretation involving the exclusion of e_r (pigeons) from that set. In (b3) and (b4), however, if we take the anaphoric expression to be the entire NP, then it is not the case that e_r (pigeons) is to be excluded from the set of birds with more sanitary habits (b3) or more sanitary birds (b4), since they don’t belong to either set: they are simply being excluded from the set of birds. So one may want to allow for an anaphoric expression to comprise only *part* of a constituent, though the interpretation of the entire constituent will, as a result, depend on how the anaphor is resolved.⁸

Now, besides e_r (the entity or entities from context_{d/u}) and e_α (the interpretation of the anaphoric expression α), we have been motivated to introduce a third entity e_i into the process of anaphor interpretation, which we call a *contextual parameter*: e_i is derived from e_r and supplied to the interpretation of α . The motivation relates to the familiar phenomenon variously called *textual ellipsis* (Hahn, Markert, and Strube, 1996), *partial anaphora* (Luperfoy, 1992), *indirect anaphora* (Hellman and Fraurud, 1996), *associative anaphora* (Cosse, 1996), and *bridging anaphora* (Not, Tovená, and Zancanaro, 1999), illustrated in Example 18.

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

⁵ to the extent that anything in human language can be considered “neat”

⁶ Since we refer to this disjunction so often, we abbreviate it simply context_{d/u}

⁷ There is more discussion of “other NPs” later in this section.

⁸ That this occurs even with definite NPs was observed over twenty years ago by one of the co-authors (Joshi, 1978), who considered the question of whether a definite NP could simultaneously co-refer and provide new information about its referent.

Here, *the receiver* is taken to be the one associated with the phone Myra darted to. Examples such as this can be modelled with the two entities we already have, by saying that e_α can be derived from e_r by *association*. However, related examples such as

(19) She lifted the receiver as Myra darted to *the other phone* ...

discovered by Modjeska (2001) in the British National Corpus, require more, since the anaphoric expression α (*the other phone*) is interpreted as the (contextually relevant) phone that is *not* the one associated with the receiver that has just be lifted. That is, getting from e_r to e_α in this case requires both *association* (as in Example 18) and *exclusion* from a contextually relevant set.

We can deal with this by introducing another entity – a contextual parameter – and computing e_α in two steps:

1. $e_r \rightarrow e_i$ (e.g., [receiver] \rightarrow [phone(e_r)])
2. $e_i \rightarrow e_\alpha$ (e.g., [phone(e_r)] \rightarrow [{contextually relevant phones} - {phone(e_r)}]).

That is, in Example 19, “the receiver” evokes a discourse entity e_r that is a receiver; from e_r , we derive e_i , the phone associated with it; and from e_i , we compute e_α , the interpretation of “the other phone” is the phone in context_{*d/u*} that is not e_i .

More generally, we distinguish possible relationships between e_r and e_i and between e_i and e_α as follows:

- e_i may be identical to e_r or some associate of e_r . This difference between COREFERENCE and MEDIATED REFERENCE is a property of *occurrences* of anaphors. That is, except for demonstrative NPs, which cannot be used for MEDIATED REFERENCE, the same type of anaphoric expression can be used for both.
- e_α may be identical to e_i or computed from e_i in some more complex way that is idiosyncratic to the particular anaphor. This difference between an interpretation that is SPECIFIED by the contextual parameter and one that is COMPUTED from it, is a property of the *type* of anaphoric expression.

These possible relations between e_r and e_i and between e_i and e_α give us the familiar case of *coreference* when $e_\alpha = e_i = e_r$. This is shown in Examples 20–23, using a variety of anaphoric expressions:

(20) **pronoun**: The terrier down the block bit me yesterday. *It’s* a vicious little beast.

(21) **definite NP**: John used to have both a terrier and a schnauzer. However, one day *the terrier* got loose and ran away.

(22) **demonstrative pronoun**: The terrier down the block bit me yesterday. My sister found *this* amusing.

(23) **demonstrative NP**: Some of the women in the ward are over 30. If any of *these primagravidae* requests a nurse, please attend to them right away.

Example 22 also illustrates identity with the discourse entity associated with an eventuality, while Example 23 illustrates our earlier point that not all the material in an anaphoric expression may be anaphoric, with “primagravidae” (applied to women over 30 who are giving birth for the first time) being new information provided about the entity that the demonstrative NP co-refers to.⁹

⁹ Accounting for how *coreference* is resolved is probably the most persistent topic in the literature on anaphora, both from a psycholinguistic and from an engineering perspective. However, it is not one we will address in this paper, though it is relevant to all forms of anaphora we discuss here.

Distinguishing the relation between e_r and e_i from that between e_i and e_α also gives the familiar case of *associative reference* (aka “textual ellipsis”, “partial anaphora”, “indirect anaphora”, “associative anaphora” and “bridging anaphora”) when $e_i = \text{assoc}(e_r)$, and either $e_\alpha = e_i$ or $e_\alpha \in e_i$. This is shown in Examples 24–26. The only constraint is that the association be licensed by the domain. (Notice that in the case of (24) and (26), $e_\alpha = e_i$ (i.e., the interpretation is SPECIFIED by the contextual parameter), while in (25), e_α is an element of e_i (i.e., the interpretation is COMPUTED from the contextual parameter).¹⁰

(24) **definite NP**: John forgot to put the picnic supplies in his cooler. So when he got to the picnic, *the beer* was too warm to drink.

$$e_r = [\text{picnic supplies}]$$

$$e_\alpha = e_i = [\text{beer}(e_i)]$$

(25) **indefinite NP**: The Number 26 bus had to detour to the Western General because *a passenger* fell unconscious.

$$e_r = [\text{the Number 26 bus}]$$

$$e_i = \text{passengers}(e_r)$$

$$e_\alpha \in e_i$$

(26) **demonstrative pronoun**: Multiply 14 times 51, and then divide *that* by 17.

$$e_r = [\text{act of multiplying } 14 * 51]$$

$$e_\alpha = e_i = [\text{result}(e_r)]$$

But e_α may also be computed from e_i in ways that are idiosyncratic to the particular anaphor, where either $e_i = e_r$ or $e_i = \text{assoc}(e_r)$. Anaphors of this type we call *lexical anaphors*. Some are referring expressions: The most common may be NPs of the form “(the) other X ”, which refer to the result of *excluding* e_i from a contextually-relevant set (Bierner, 2001a; Bierner, 2001b; Bierner and Webber, 2000; Modjeska, 2001). Here, e_i may be an individual (Example 27) or a set derived from a plural NP or through “split reference” (Example 28).

(27) Q: What’s the drinking age in Afganistan?

A: ...

Q: What’s it in *other countries*?

$$e_i = e_r = [\text{Afganistan}]$$

$$e_\alpha = \{\text{countries}\} - \{e_i\}$$

(28) Q: What’s the drinking age in Afganistan?

A: ...

Q: What’s it in Bolivia?

A: ...

Q: What’s it in *other countries*?

$$e_i = e_r = \{[\text{Afganistan}], [\text{Bolivia}]\}$$

$$e_\alpha = \{\text{countries}\} - e_i$$

As we have already seen in Example 19 (repeated here), the contextual parameter may be associated with e_r (MEDIATED REFERENCE), rather than identical with it:

(19) She lifted the receiver as Myra darted to *the other phone*

$$e_r = [\text{receiver}]$$

$$e_i = [\text{phone}(e_r)]$$

$$e_\alpha = [\{\text{contextually relevant phones}\} - \{e_i\}]$$

¹⁰ Again here, the main problem noted in the literature is that of specifying effective resolution procedures that take advantage of both the context_{d/u} and world knowledge – the latter to characterise licenced associations.

What is to be excluded can also come from the context of utterance, as in Example 29, asked by or of the first author,

- (29) Are there *other short people working on discourse*?
 $e_i = e_r = [\text{BLW}]$
 $e_\alpha = \{\text{short people working on discourse}\} - \{e_i\}$

or the excluded entity may be an eventuality associated with one or more clauses, as in the following from (Modjeska, 2001):

- (30) If the patient is very heavy or the carer cannot manage for *some other reason* ... [BNC]
 $e_i = e_r = \|\text{the patient is very heavy}\|$
 $e_\alpha = \{\text{reasons why the carer cannot manage}\} - \{e_i\}$

Note that “other *X*” also *presupposes* that the excluded entities belong to the set under consideration (Bierner and Webber, 2000; Bierner, 2001a; Bierner, 2001b). For example, in (31), “other rug-making countries” presupposes that China is a rug-making country, and “other artistic disciplines” presupposes that rug-making is an artistic discipline.

- (31) Unlike *other rug-making countries*, China mainly draws its design repertoire from *other artistic disciplines* (painting, etc), ...
 $e_{i1} = e_{r1} = [\text{China}]$
 $e_{\alpha1} = \{\text{rug-making countries}\} - \{e_{i1}\}$
presupposed: $e_{r1} \in \{\text{rug-making countries}\}$
 $e_{i2} = e_{r2} = [\text{rug-making}]$
 $e_{\alpha2} = \{\text{artistic disciplines}\} - \{e_{i2}\}$
presupposed: $e_{r2} \in \{\text{artistic disciplines}\}$

A hearer for whom this is new information must either *accommodate* it or refuse to do so. Certain relational anaphors such as “nevertheless” also have a presuppositional component to their meaning (Section 2.2). This will be relevant to our discussion in Section 3 of relations between the semantic contributions of anaphoric connectives and those of structural connectives and adjacency-triggered inferences.

Other expressions discussed in (Bierner, 2001b) that incorporate in an idiosyncratic way, an individual or set from the discourse context or context of utterance are noun phrases headed by “other” (Example 32), “such NPs” (Example 33), comparative NPs (Example 34), and the pronoun “elsewhere” (Example 35).

- (32) Some dogs are constantly on the move. *Others* lie around until you call them.
(33) I saw a 2kg lobster in the fish store yesterday. The fishmonger said it takes about 5 years to grow to *such a size*.
(34) Terriers are very nervous. *Larger dogs* tend to have calmer dispositions.
(35) I don’t like sitting in this room. Can we move *elsewhere*?

As Bierner (2001b) notes, these have similar presuppositions about membership of e_i in the set under consideration. In addition, all of them also appear in constructions that are not anaphoric – e.g., Xs other than Y, Xs such as Y, <comparative> Xs than Y, etc., where what is to be excluded or compared to is provided structurally. But ignoring these non-anaphoric versions, the same problem of how to resolve *lexical function* anaphors against the discourse context or context of utterance remains to be solved (Modjeska, 2001; Salmon-Alt, 2000).¹¹

¹¹ Our discussion here of relations between e_r , e_i and e_α focusses on referring expressions and

2.2 Discourse Adverbials as Lexical Anaphors

We want to claim that lexical anaphors do not have to be referring expressions. Instead, a lexical anaphor can express a binary relation between an anaphorically-derived argument (the contextual parameter e_i) and the interpretation of the anaphor’s matrix sentence or clause. The result is an additional proposition contributed to the discourse.

This is the way in which we claim that a variety of discourse adverbials are anaphoric – for example, “then” in Example 36 and “instead” in Example 37.¹²

- (36) 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. (i.e., after he ordered the wine, he discovered he was broke.)
- (37) John didn’t have enough money to buy a mango. *Instead*, he bought a guava. (i.e., he bought a guava as an alternative to buying a mango)

These paraphrases are only meant to convey rough approximations of the actual meaning of “then” and “instead”. Our concern here is only with the *mechanism* by which these adverbials get their meaning. For a detailed analysis of the meaning of discourse adverbials and connectives, see (Jayez and Rossari, 1998a; Jayez and Rossari, 1998b; Lagerwerf, 1998; Traugott, 1995; Traugott, 1997) and others.

Formally, we model such lexical anaphors – which we call *relational anaphors* – as a function that maps the contextual parameter e_i (either coreferential with e_r or associated with it) to an expression that is idiosyncratic to the anaphor

$$\alpha : e_i \rightarrow \lambda x. R_\alpha(x, e_i)$$

That function then applies to the eventuality σ that corresponds to the interpretation of the anaphor’s matrix clause S , yielding a proposition

$$\alpha S : [\lambda x. R_\alpha(x, e_i)]\sigma = R_\alpha(\sigma, e_i)$$

The two different function notations (\rightarrow , λ) and the explicit step of derivation here indicate that the contextual parameter e_i serves as an argument to R_α but that σ is supplied compositionally from syntax.

We will now use this to explain what is going on in a range of examples involving discourse adverbials. But to do so, we need to introduce and justify the representation we will use for clausal interpretations.

There are two common ways of representing predicate-argument relations in sentence-level interpretations. For example, the meaning of Example 38a can be roughly represented either as (b) or (c).

- (38) a. John likes some apples.
 b. likes’(e_j,a) \wedge john’(j) \wedge some-apples’(a)
 c. likes’(john’, some-apples’)

discourse adverbials. But other linguistic phenomena have been considered anaphoric, including VP ellipsis (Hardt, 1999; Kehler, 1995), “do so” anaphora (and, we suggest, “do otherwise”). Certain modifiers might also be considered anaphors (e.g., “different”, “differently”, “similar”, “similarly”, etc.). So it would be worthwhile to consider what, if anything, can be gained by analysing such anaphors in terms of these two stages of $e_r \rightarrow e_i$ and $e_i \rightarrow e_\alpha$.

¹² Words and phrases that function as discourse adverbials may have other functions as well – e.g., “otherwise” can be used as an adjectival modifier, as in “I was *otherwise* occupied with grading exams.” and “on the other hand” may serve as one half of the structural connective “On the one hand, . . . On the other (hand), . . . ” (cf. Section 4). Overloading closed-class lexico-syntactic items is not unusual in English, and must just be handled as part of the normal ambiguity resolution process.

(38b) makes explicit the role of individuals, eventualities and relationships among them in interpretation. It helps the reader understand the kind of meanings that lexical items can have, or the way lexical meanings contribute to utterance meaning. But representations like (38b) are less general than representations like (38c), as Example 39 illustrates.

- (39) a. The president opposes tax increases.
 b. $\text{oppose}'(e,p,i) \wedge \text{president}'(p) \wedge \text{taxincreases}'(i)$
 c. $\text{oppose}'(\text{president}', \text{taxincreases}')$

We can easily understand (39a) as a compact description of many different people (Reagan, Bush, Clinton, Bush) fulfilling a common role over time, and the attitudes they have taken to a number of purely hypothetical objects (planned increases that were never realized). Representations like (39b) accommodate this only by some rather extreme assumptions about what individuals and eventualities can count as values for e , p or i . Representations like (38c) and (39c) are more general, while affording less specific semantic intuitions.

We find the same tradeoff with predicate-argument structures in discourse. (For ease of reading later on, we will switch here to a minor syntactic variant of the “b” representation, in which the eventuality argument indexes the predicate – for example, $e:\text{likes}(j,a)$ rather than $\text{likes}(e,j,a)$.)

- (40) a. John left because Mary left.
 b. $e_1:\text{left}'(j) \wedge \text{john}'(j) \wedge e_2:\text{left}'(m) \wedge \text{mary}'(m) \wedge e_3:\text{because}(e_1,e_2)$
 c. $\text{because}'(\text{left}'(\text{john}'), \text{left}'(\text{mary}'))$

While (40b) gives good intuitions about the individuals and eventualities described in discourse and the compositionality involved in discourse interpretation, representing “because” this way requires understanding eventualities in a rich and potentially problematic way. On the other hand, while (40c) is unproblematic, its notation does not help us remember the many constraints on “because” – in particular, that John and Mary did in fact leave!

Anaphora and anaphor resolution complicate the picture: The b-representations represent resolved anaphors by reuse of a discourse referent, while the c-representations require the construction of E-type descriptions (Evans, 1980; Neale, 1990) that draw on material from previous discourse. We can see this by continuing Example 38 as follows:

- (41) a. ... [but] Bill hates them.
 b. ... $e_2:\text{hates}'(b,a) \wedge \text{bill}'(b)$
 c. $\text{hates}'(\text{bill}', \text{the}'(\text{apples}', \lambda x . \text{likes}'(\text{john}', x)))$

(41b) reuses the discourse referent a introduced by “some apples” in (38b). For (38c), there is no such discourse referent, so we have to construct a description: “the apples John likes”.

In this paper, we follow Hobbs (1985) in using b-style representations, because we want to make intuitions about individuals, eventualities, lexical meaning and anaphora as clear as possible. But this choice is not theoretically necessary. Using this representation, we treat our familiar Example 10, repeated here as (42), as follows:

- (42) 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.
 $\text{then}: e_i \rightarrow \lambda x . \text{after}(x, e_i)$
 $S = \text{he} [\text{John}] \text{discovered he was broke}$
 $\sigma = e_4$, where $e_4:\text{discover}(j, e_5)$ and $e_5:\text{broke}(j)$
 $e_i = e_r = e_2$, where $e_2:\text{order}(j, c_1)$
 $\text{then } S: [\lambda x . \text{after}(x, e_2)]e_4 \equiv \text{after}(e_4, e_2)$

That is, e_4 (the interpretation of S as an eventuality) is the event of John discovering that he was broke, while e_2 is the event of John ordering three cases of the '97 Barolo.

Resolving the anaphor “then” to e_2 leads to the proposition $after(e_4, e_2)$ being added to the discourse context.

Similarly, in Example 43, “then” picks up a culminated event from the discourse context and maps it to an expression that applies to the interpretation of its matrix sentence, adding a proposition to the discourse context.

(43) Go west on Lancaster Avenue. *Then* turn right on County Line.

then: $e_i \rightarrow \lambda x . after(x, e_i)$
 $S = \text{turn right on County Line}$
 $\sigma = e_3$, where $e_3:\text{turn-right}(you, cl)$
 $e_r = e_1$, where $e_1:\text{go-west}(you, la)$
 $e_i = e_2$, where $e_2:\text{culmination}(e_r)$
then S : $[\lambda x . after(x, e_2)]e_3 \equiv after(e_3, e_2)$

That is, e_3 is the event of the hearer (h) turning right on County Line (cl), which is the interpretation of S as an eventuality, while e_r resolves to e_1 , the event of the hearer going west on Lancaster Avenue. But since “then” requires a culminated eventuality as its second argument, e_i is its associated culmination, e_2 . But of course, the intended culmination is not the end of Lancaster Avenue (about 75 miles west of downtown Philadelphia), but its intersection with County Line (about 4 miles west of downtown). This must be derived through further inference that we do not discuss here. Finally, resolving “then” leads to the proposition $after(e_3, e_2)$ being added to the discourse context.

It is important to stress here that the level of representation we are concerned with is essentially a *logical form* (LF) for discourse – propositions of the form $after(e_3, e_2)$ and $if(e_4, e_5)$. Any reasoning that might then have to be done on their content might then require making explicit the different modal and temporal contexts involved, their accessibility relations, etc. But as our goal here is primarily to capture the mechanism in which discourse adverbials are involved in discourse structure and discourse semantics, we will continue to assume for as long as possible that a 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 – that is, as evidence for relations between adjacent discourse units. However, we claim that a relation conveyed by an anaphor can be distinct from any relation associated with structure. In fact, we will demonstrate in Section 3 a variety of ways in which discourse adverbials can *interact* with inferred relations and explicit structural relations.

One particular relational anaphor – “otherwise” – that we discussed previously in (Webber et al., 1999a) deserves more comment here. Roughly speaking, “otherwise” conveys that the *complement* of its anaphorically-derived contextual parameter e_i serves as a *condition* under which the interpretation of its structural matrix holds. (This complement must be with respect to some contextually relevant set.¹³)

13 Kruijff-Korbayová and Webber (2001) 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 , and thus what contextual parameters e_i , 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, while small capitals convey focus within the rheme. In (i), the “otherwise” clause will be interpreted as warning the hearer (H) that H might get hurt if s/he 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 s/he might

If we represent a conditional relational between two eventualities simply as $if(e_1, e_2)$, where e_1 is the *antecedent* and e_2 , the *consequent*, and we approximate the contextually relevant alternatives e_2 to an eventuality e_1 using a *complement* predicate – e.g., $complement(e_1, e_2)$ – then we can represent the interpretation of “otherwise” as

otherwise: $e_i \rightarrow \lambda x . if(e_{gi}, x)$, where $complement(e_i, e_{gi})$

and index gi is heretofore unused. That is, “otherwise” presupposes a contextually relevant complement to e_i and asserts that if any member of that complement holds, the argument to the λ -expression will. The resulting λ -expression applies to the interpretation of the matrix clause of “otherwise”, resulting in the both the complement and the conditional being added to the discourse context:

otherwise S: $[\lambda x . if(e_{gi}, x)] \sigma$, where $complement(e_i, e_{gi})$
 $\equiv if(e_{gi}, \sigma)$, where $complement(e_i, e_{gi})$

As we showed in Section 1.2, different ways of resolving the anaphoric argument lead to different interpretations:

- (44) If the light is red, stop. *Otherwise* you’ll get a ticket.
otherwise: $e_i \rightarrow \lambda x . if(e_{gi}, x)$, where $complement(e_i, e_{gi})$
 $S =$ you get a ticket
 $\sigma = e_3$, where $e_3: get_ticket(you)$
 $e_i = e_r = e_2$, where $e_2: stop(you)$
otherwise S: $if(e_{gi}, e_3)$, where $complement(e_2, e_{gi})$

i.e., If you do something other than stop, you’ll get a ticket.

- (45) If the light is red, stop. *Otherwise* go straight on.
otherwise: $e_i \rightarrow \lambda x . if(e_{gi}, x)$, where $complement(e_i, e_{gi})$
 $S =$ go straight on
 $\sigma = e_3$, where $e_3: go_straight(you)$
 $e_i = e_r = e_2$, where $e_2: red(light1)$
otherwise S: $if(e_{gi}, e_3)$, where $complement(e_2, e_{gi})$

i.e., if the light is not red, go straight on.

Like plural pronouns, definite NPs and “other NPs” (cf. Example 28), “otherwise” too can exploit “split antecedents”, which are then excluded from the context of interpretation of the matrix clause, as in:

- (46) If the light is red, you should stop.
 If it’s flashing yellow, you should slow down.
Otherwise you can continue on your way.

Here, the light being red and it being yellow are both excluded from the contextually relevant situations (i.e., ones related to the state of the light).

And as already noted (Section 1.2), limited forms of inference may also be required to resolve “otherwise” and other relational anaphors, as in:

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

Here the situations in which you can have a pear are ones alternative to those in which you want an apple – i.e., in which the answer to the question is “yes”.

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).

To close, we want to point to a final type of lexical anaphor which contributes to the discourse context neither an entity (by virtue of being a referring expression) nor a proposition (as above). Instead, its idiosyncratic contribution to the discourse context has the form of a *rule* – the same kind of presupposed defeasible rule (*PDR*) that Lagerwerf (1998) in his PhD dissertation attributes to the semantics of the structural connectives “although” and “but” – a rule whose applicability is denied in the current case.¹⁴ This is the contribution of the discourse adverbials “nevertheless” and “though”

Stretching our previous notation somewhat, we can represent the defeasible rule they contribute to the discourse as

$$\textit{nevertheless}: e_i \rightarrow \lambda x . e_i > \neg x$$

Applying this to the interpretation σ of the matrix clause of the adverbial yields

$$\textit{nevertheless } S: [\lambda x . e_i > \neg x]\sigma \equiv e_i > \neg\sigma$$

where $>$ is Asher & Morreau’s *commonsense entailment* operator (1991). (That is, normally if the anaphorically derived situation holds, σ doesn’t.) For example,

- (48) John graduated with honors. *Nevertheless* he was depressed.

$S =$ John was depressed

$\sigma = e_2$, where $e_2:\text{depressed}(\text{john})$

$e_i = e_r = e_1$, where $e_1:\text{graduate.cum.laude}(\text{john})$

nevertheless } S: $[\lambda x . e_1 > \neg x]e_2 \equiv e_1 > \neg e_2$

i.e., *Normally if John graduates with honors, he is not depressed.* This rule is then denied by the matrix clause – he was depressed. While we recognize that the defeasible rule that is accommodated (or conventionally implicated) here is more likely to involve generalisations of both e_i and σ – something like “normally if *someone* graduates with honors, they are not depressed”, rather than something particular to John – the process of abstracting to an appropriate level seems separable from that of resolving the anaphor and formulating the rule whose applicability is being denied.

2.3 Summary

The general framework we have presented for anaphora has two main features:

- It posits a third entity e_i – a *contextual parameter* – involved in the resolution process, in order to allow all types of anaphora to make use of the step from a discourse referent to one associated with it, and then a subsequent step that may simply be equivalence or something idiosyncratic to the type of anaphor.
- It allows anaphors to use e_i in idiosyncratic ways that may lead, not only to new referring expressions (and thus new entities), but also to additional propositions and presuppositions in the discourse context.

We have shown e_r and e_i being used systematically in a variety of ways in computing the interpretation e_α of anaphoric expressions α , and have thereby enlarged the range of expressions usefully thought of as anaphoric.

¹⁴ Earlier, both George Lakoff (1971a) and Robin Lakoff (1971b) called attention to such a presupposition.

3 Inferred, Structural and Anaphoric Relations

Prior to the current work, accounts have treated both explicit structural connectives (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 (1995) 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 a *elaboration* relation, “therefore” as evidence of a *result* relation, “because” as evidence of a *explanation* relation, and “even though” as evidence of a *denial of preventer* relation (Kehler, 1995, Chapter 2.1), and Marcu (1999), following Mann and Thompson (1988), appears to take “otherwise” as evidence for an *otherwise* relation.

Because we take discourse adverbials to contribute meaning in a different way than explicit structural connectives, we predict that they can interact in a variety of ways with relations conveyed structurally and inferred relations triggered by adjacency. Below we show that this prediction is correct.

We start from the idea that – in the absence of an explicit structural connective – defeasible 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 generalised 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 expectation*), or *contiguity* relations (*narration*) described in (Hobbs, 1990; Kehler, 1995). 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 non-defeasible evidence for an inferrable relation (e.g., “yet”, “so” and “because”).

This is not, we claim, what discourse adverbials do. Rather, they *interact* in a variety of ways with structural connectives, with adjacency-triggered defeasible inference and with each other. This section describes the kinds of interactions we have observed so far, using the same notation used in the previous section:

- α = discourse adverbial;
- S = the matrix clause/sentence of α ;
- σ = the logical form (LF) interpretation of S ;
- e_i = the contextual parameter supplied to the interpretation of α ;
- R_α = the name of the relation associated with α .

But because we will be considering how the relation between discourse-adjacent units can interact with the interpretation of a discourse adverbial, we need some additional notation as well.

- D = the immediately left-adjacent discourse unit that S relates to via an adjacency-triggered inference or an explicit structural connective;
- δ = the LF interpretation of D ;
- R = the name of the relation between σ and δ ;

Case 1: $R_\alpha(\sigma, e_i)$ is distinct from $R(\sigma, \delta)$ because $e_i \neq \delta$.

As before, we start with our familiar Example 10, repeated earlier as (42) and here as (49):

- (49) 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: $e_i \rightarrow \lambda x . \text{after}(x, e_i)$
 $S = \text{he} [\text{John}] \text{discovered he was broke}$
 $\sigma = e_4$, where $e_4:\text{discover}(j, e_5)$
 $D = \text{he} [\text{John}] \text{had to cancel the order}$
 $\delta = e_3$, where $e_3:\text{cancel}(j, o_1)$
 $R(\sigma, \delta) \equiv \text{explanation}(e_4, e_3)$
 $e_i = e_r = e_2$, where $e_2:\text{order}(j, c_1)$
 $R_\alpha(\sigma, e_i) \equiv [\lambda x . \text{after}(x, e_2)]e_4 \equiv \text{after}(e_4, e_2)$

That is, *after* relates the discovery (e_4) to the ordering (e_2), while *explanation* (conveyed by “because”) relates it to the cancelling (e_3).¹⁵

Case 2: R incorporates $R_\alpha(\sigma, e_i)$ as one argument. When the anaphor “otherwise” is resolved, the resulting conditional relation serves as one argument to R , and δ serves as the other. This holds whether R is conveyed structurally (cf. Example 50a with “because”, Example 51a with “but”) or a by adjacency-triggered inference, as shown in the parallel “b” examples, repeated here from (44) and (45).

- (50) a. If the light is red, stop, because *otherwise* you’ll get a ticket.

b. If the light is red, stop. *Otherwise* you’ll get a ticket.
otherwise: $e_i \rightarrow \lambda . \text{if}(e_{gi}, x)$, where $\text{complement}(e_i, e_{gi})$
 $S = \text{you get a ticket}$
 $\sigma = e_3$, where $e_3:\text{get_ticket}(\text{you})$
 $D = \text{stop}$
 $\delta = e_2$, where $e_2:\text{stop}(\text{you})$
 $e_i = e_r = e_2$, where $e_2:\text{stop}(\text{you})$
 $R_\alpha(\sigma, e_i) \equiv e_4:\text{if}(e_{gi}, e_3)$, where $\text{complement}(e_2, e_{gi})$
 $R(\sigma, \delta) \equiv \text{explanation}(e_4, e_2)$

i.e., If you do something other than stop, you’ll get a ticket.

- (51) a. If the light is red, stop, but *otherwise* go straight on.

b. If the light is red, stop. *Otherwise* go straight on.
otherwise: $e_i \rightarrow \lambda . \text{if}(e_{gi}, x)$, where $\text{complement}(e_i, e_{gi})$
 $S = \text{go straight on}$
 $\sigma = e_3$, where $e_3:\text{go_straight}(\text{you})$
 $D = \text{if the light is red, stop}$
 $\delta = e_4$, where $e_4:\text{if}(e_2, e_1)$, $e_1:\text{red}(\text{light1})$, $e_2:\text{stop}(\text{you})$
 $e_i = e_r = e_1$, where $e_1:\text{red}(\text{light1})$
 $R_\alpha(\sigma, e_i) \equiv e_5:\text{if}(e_{gi}, e_3)$, where $\text{complement}(e_2, e_{gi})$
 $R(\sigma, \delta) \equiv \text{contrast}(e_5, e_4)$

¹⁵ Because eventuality e_2 has both the properties of *explaining* the cancelling and of being *after* the ordering, it follows that what explains the cancelling is something that was after the ordering.

i.e., if the light is not red, go straight on.

Notice that the above treatment obviates the need for a separate *otherwise* relation. Mann and Thompson (1988) describe their proposed *otherwise* relation as having the *effect*:

R [the reader] recognizes the dependency relation of prevention between the realization of the situation presented in N [the nucleus] and the realization of the situation presented in S [the satellite] (Mann and Thompson, 1988, p.276)

and give as an example:

Anyone desiring to update their entry in this brochure should have their copy in by Dec. 1. Otherwise the existing entry will be used.

This is similar to Example 50. But given our overall approach, where lexico-syntactic material can contribute to both clause-level and discourse-level semantics, it is not difficult to see how resolving “otherwise” and inferring *explanation* (or having explicit evidence for it, in the case of Example 50a), is exactly what Mann and Thompson were after. And, as demonstrated, the above approach accounts for other instances of “otherwise” as well.

Case 3: The relation contributed by the adverbial is parasitic on *R*. Here we depart from our straight “discourse adverbial as anaphor” story because it appears that some adverbials – the clearest cases being “for example” and “for instance” – function in discourse neither as connectives nor as anaphors. Rather, they appear to derive their interpretation *parasitically* on the relation associated with a structural connective or discourse adverbial or on an inferred relation triggered by adjacency. The way to see this is to consider *intra-clausal* use of “for example”, where it follows the verb, as in

(52) The collection includes, *for example*, a piece of hematite.

Interpreting “for example” here involves abstracting the meaning of its matrix structure with respect to the material to its right, and then making an assertion with respect to this abstraction. That is, if the logical form (LF) contributed by the matrix clause of Example 52 is, roughly,

i. `include(collection1, hematite1)`

then the LF added by “for example” is

ii. `example_of(hematite1, {X | include(collection1, X)})`

That is, “hematite” is an example of things included in the collection.¹⁶ (Since with appropriate axioms, proposition (ii) implies proposition (i), one might choose to retain only (ii) and derive (i) when needed. But this is a matter of choice, not a claim about whether both or only one is the interpretation of (52).)

If we look at the comparable situation in discourse, where “for example” occurs to the right of an explicit structural connective such as “so” (Example 53a) or “because” (Example 53b) or a relational anaphor such as “then” (Example 53c), it can also be seen as abstracting the interpretation of its discourse-level matrix structure, with respect to the material to its right.

¹⁶ The material to the right of “for example” appears able to be any kind of CCG constituent (Steedman, 1996; Steedman, 2000b), including such strange ones as

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

Here, “a flower to a policeman” would be an example of the set of object-recipient pairs within John’s givings.

- (53) a. John just broke his arm. *So, for example*, he can't cycle to work now.
 b. You shouldn't trust John *because, for example*, he never returns what he borrows.
 c. Shall we go to the Lincoln Memorial? *Then, for example*, we can go to the White House.

In (53a), the interpretation of the discourse-level matrix structure headed by the interpretation of “so” is:

$$result(\sigma, \delta)$$

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 matrix interpretation with respect to the material to its right (i.e., σ), thereby contributing:

$$exemplification(\sigma, \{X \mid result(X, \delta)\})$$

That is, “John can't cycle to work” is an example of the results of “John breaking his arm”. Similarly, what is added by the matrix sentence of (53b) is

$$explanation(\sigma, \delta)$$

where σ is the interpretation of “he never returns what he borrows” and “for example” adds

$$exemplification(\sigma, \{X \mid explanation(X, \delta)\})$$

i.e., that this is an example of the reasons for not trusting John. And the proposition contributed by the resolved discourse adverbial “then” in (53c) is

$$after(\sigma, \delta)$$

where σ is the interpretation of “we can go to the White House”, and “for example” adds

$$exemplification(\sigma, \{X \mid after(X, \delta)\})$$

i.e., that this is an example of the events that follow going to the Lincoln Memorial. (**N.B.** We use the relation $R_\pi = exemplification$ here, rather than *example_of* used in the interpretation of (52), because it is what is commonly found in the literature on discourse relations. We are also being fairly fast and loose regarding tense and modality, in the interests of making a strong case for the basic scaffolding.)

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

- (54) You shouldn't trust John. *For example*, he never returns what he borrows.

$$explanation(\sigma, \delta)$$

$$exemplification(\sigma, \{X \mid explanation(X, \delta)\})$$

Here, as in Example 53b, the relation provided by adjacency-triggered inference is $R = explanation$, which is then used by “for example”.

But what about the many cases where only *exemplification* seems present, as in

- (55) a. 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]

- b. The shows that top the ratings are soft, for example we make Rupert the Bear and that gets a 60 per cent share of the kids’ audience. [BNC, K5C 909]

There are at least two explanations: One is that “for example” simply provides direct non-defeasible evidence for *exemplification*, 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 (55a) that “having database cross-referencing facilities” elaborates the respects in which hypertext books are superior to normal books, while in (55b), we understand that “Rupert the Bear getting a 60% share of the kids’ audience” elaborates the claim that “shows that top the ratings are soft”. This *elaboration* relation is then abstracted (in response to “for example”) to produce:

$$\textit{exemplification}(\sigma, \{X \mid \textit{elaboration}(X, \delta)\})$$

i.e., that this is one example of many possible elaborations. Because this is more specific than *elaboration* and seems to mean the same as *exemplification*(σ, δ), one might assume that it is the only relation that holds. Given that so many naturally-occurring 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 be the case.

Before going on to **Case 4**, we should comment on occurrences of “for example” elsewhere in a sentence or clause. Here it may simply contribute propositional meaning *intra-clausally* as a parenthetical, illustrating an abstraction introduced by an NP, PP or clause, rather than being parasitic on another relation, as in:

- (56) a. In the case of the managed funds they will be denominated in a leading currency, for example US dollar, . . . [BNC CBX 1590]
(i.e., *US dollar is an example of a leading currency*)
- b. And Kuhn himself argued that ideas that have been rejected by contemporary science – that heat, for example, is caused by phlogiston . . . [NY Times on the Web 21 July 2001, “Coming to Blows over how valid Science really is”]
(i.e., *that heat is caused by phlogiston is an example of ideas that have been rejected by contemporary science*)

(In “English” English – in contrast with “American” English, the BNC shows most such examples to 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, such as in Example 52.)

But there are also more subtle cases of clause-medial “for example”. Consider Example 57.

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

Here, as in Example 53a, “so” explicitly connects the two clauses. But (57) cannot be paraphrased as

All the children are ill, so *for example* Andrew can’t help out in the shop.

because it describes not just a example consequence of all the children being ill, as would

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

but a consequence with respect to an example instance from the set of children.

We suspect here the involvement of *Information Structure* (Steedman, 2000a): While the interpretation conveyed by “for example” is parasitic on the adjacency relation (*result* in Example 57), its position after the NP in (57) may indicate a contrastive theme with respect to the previous clause. 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 . Earlier (Section 2.2), we noted that the relation conveyed by certain discourse adverbials – notably, “nevertheless” and “though” – has the nature of a presupposed (or conventionally implicated) *defeasible rule* that fails to hold in the current situation. With discourse adverbials, the *antecedent* to the rule derives anaphorically from the previous discourse, while the *consequent* derives from the adverbial’s matrix clause.

Here we illustrate this possibility with examples in which “nevertheless” occurs in the main clause of a sentence containing a preposed subordinate clause. Where **Case 2** showed R_α incorporated into an argument of R , this possibility shows an abstraction of R incorporated into the defeasible rule that manifests R_α . For example,

- (59) *While* John showers, he *nevertheless* thinks about chess.
 $S =$ he [John] thinks ... $\sigma = e_2$, where e_2 :think_about(john, chess)
 $D =$ John showers $\delta = e_1$, where e_1 :shower(john)
 R : *during*(e_2, e_1) R_α : *during*(X, e_1) $> \neg (X = e_2)$

Paraphrase: Normally, whatever one does *during the time* one showers, it is not thinking about chess.

- (60) Even *after* John has had three glasses of wine, he is *nevertheless* able to solve difficult algebra problems.
 $S =$ he is able to solve ... $\sigma = e_2$, where e_2 :solve(john, hard-algebra-problems)
 $D =$ John has three glasses ... $\delta = e_1$, where e_1 :drink(john,wine)
 R : *after*(e_2, e_1) R_α : *after*(X, e_1) $> \neg (X = e_2)$

Paraphrase: Normally, whatever one is able to do *after* one has had three glasses of wine, it is not solving difficult algebra problems.

We speculate that the reason such examples sound more natural with the focus particle “even” applied to the subordinate clause, is that “even” conveys an even greater likelihood that the defeasible rules holds, so “nevertheless” emphasises its failure to do so.

Summary

We have indicated four ways in which we have found the relation associated with a discourse adverbial (R_α , in the case of a relational anaphor, and R_π , in the case of “for example”) to interact with a relation R triggered by adjacency or conveyed by structural connectives or, in some cases, by another relational anaphor:

1. $R_\alpha(\sigma, e_i)$ is distinct from $R(\sigma, \delta)$;
2. $R_\alpha(\sigma, e_i)$ serves as the first argument to R .
3. R_α contributed by the adverbial is parasitic on R ;
4. R_α is a defeasible rule that incorporates R ;

While we do not believe that this is an exhaustive list, and we do not know whether a discourse adverbial always behaves the same way vis-a-vis other relations, nevertheless we believe that acknowledging some discourse adverbials to be anaphors (and at least one to be neither anaphor nor connective) opens such issues up for exploration in ways they have not been before.

4 Lexicalised Grammar for Discourse Syntax and Semantics

As we noted in the Introduction, we do not believe that the relation between discourse and semantics is intrinsically different from that between a sentence and its semantics

– i.e., that both are, at least in part, a projection from the lexicon and syntax. The alternative – that discourse relates to semantics in a completely different way – seems strange when not only can a clause be part of a discourse, a discourse can be part of a clause:

- (61) Any farmer who has beaten a donkey and gone home regretting it and has then returned and apologised to the beast, deserves forgiveness.
- (62) 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, 17 June 1997]

That is, the successive conjuncts within the relative clause of (61) and within the conditional antecedent of (62) exhibit the cohesive and argumentative connections which are characteristic of the interpretation of discourse.

In the previous section, we showed how the semantics of discourse adverbials could be resolved within the clause and projected into discourse. Now we take another step back, to sketch out a coupling between discourse syntax and semantics that is a natural outgrowth of the coupling between clause-level syntax and semantics.

Because lexicalized grammars such as Lexicalized Tree-Adjoining Grammar (LTAG) (Joshi, 1987; XTAG-Group, 2001) and Combinatory Categorical Grammar (CCG) (Steedman, 1996; Steedman, 2000b) have been very successful in showing how clause-level syntax and semantics project from the lexicon, LTAG is our grammar of choice here.¹⁷ We have described this work in several conference papers (Webber, Knott, and Joshi, 1999; Webber et al., 1999a; Webber et al., 1999b), and this has 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, which we call DLTAG (Section 4.1), and then show how it supports the approach to structural and anaphoric discourse connectives presented earlier (Section 4.2).

4.1 DLTAG and Discourse Syntax

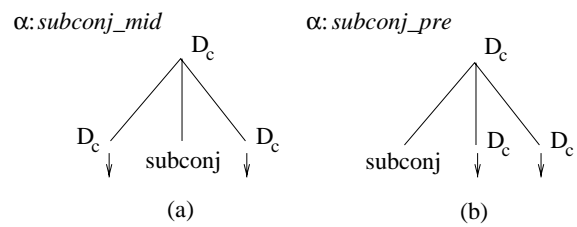
A lexicalized TAG begins with the notion of a lexical *anchor*, which can have one or more associated tree structures. For example, the verb *likes* anchors one tree corresponding to *John likes apples*, another corresponding to the topicalized *Apples John likes*, a third corresponding to the passive *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 only requires a few elementary tree structures, possibly because clause-level syntax exploits structural variation in ways that discourse doesn't.

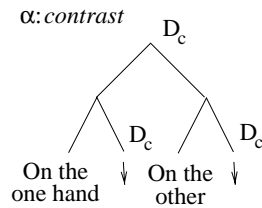
4.1.1 Initial Trees The grammar has initial trees for three types of construction: (a) subordinate-main clause constructions; (b) parallel constructions; and (c) what we call *relational coordination*. We describe each in turn.

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 –

¹⁷ CCG would have been an equally good choice.

**Figure 6**

Initial trees (a-b) for a subordinate conjunction. D_c stands for “discourse clause”, \downarrow indicates a substitution site, while “subconj” stands for the particular subordinate conjunction that anchors the tree.

**Figure 7**

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

because they are outside the domain of locality of the verb. From a discourse perspective, 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 6 shows the initial trees for postposed subordinate clauses (a) and preposed subordinate clauses (b).¹⁸ 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 1, local dependencies can be stretched long-distance as well – as in

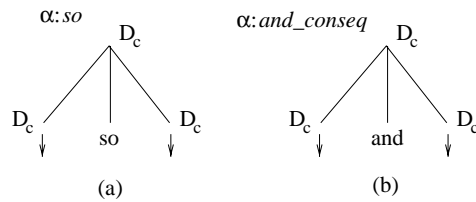
- (63) a. Although John is generous, he’s hard to find.
 b. Although John is generous – for example, he gives money to anyone who asks him for it – he’s hard to find.
- (64) a. On the one hand, John is generous. On the other hand, he’s hard to find.
 b. On the one hand, John is generous. For example, suppose you needed some money: You’d only have to ask him for it. On the other hand, he’s hard to find.

Thus our lexicalised discourse grammar also contains initial trees for parallel constructions as in (64) and Figure 7. 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” ...).

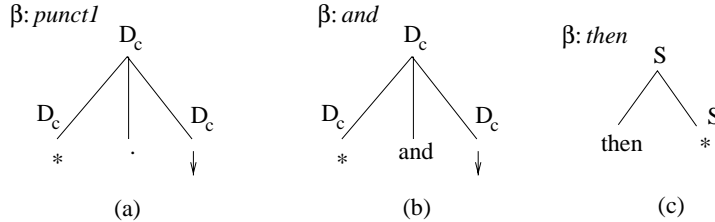
The third construction for which we have an initial tree is for structural connectives that convey a particular relation between the connected clauses. So, for example, there is an initial tree associated with “so” conveying *result* – cf. Figure 8a. Additionally, we posit initial trees for *relational coordination*, cases where “and” or “or” convey a particular relation between conjuncts (disjuncts) besides simple truth-functionality. For example, both “and” and “or” convey *result* in

- (65) a. Throw another spit ball *and* you’ll regret it.

¹⁸ While in an earlier paper (Webber and Joshi, 1998), we discuss reasons for taking the lexical anchors of the initial trees in Figures 6 and 7 to be feature structures, following the analysis in (Knott, 1996; Knott and Mellish, 1996), here we just take them to be specific lexical items.

**Figure 8**

Initial trees for coordinate conjunction. These particular trees are for (a) “so” and (b) *relational coordination* on “and” expressing *consequence*.

**Figure 9**

Auxiliary trees for basic elaboration. These particular trees are anchored by (a) the punctuation mark “.” 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”.

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

while “and” can also convey *purpose*, as in

(66) Go to the shop *and* get me a quart of milk.

From a discourse perspective, *relational coordination* differs from what we are calling *scopal coordination*, in that the latter simply conveys that both conjuncts bear the same relation to the immediately left-adjacent discourse unit, whatever that may be. For example, in (67) the relation is *explanation* and each conjunct is a separate *explanation* for not trusting John, while in (68), the relation is *result*.

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

(68) a. John won the lottery. So his wife quit her job, *and* he bought a yacht.
b. John just won the lottery. So he will quit his job, *or* he will at least stop working overtime.

In (68a), each conjunct is a separate *result* of John’s winning the lottery, while in (68b), each disjunct conveys an alternative result of John’s good fortune.

We distinguish *relational coordination* and *scopal coordination* in the grammar by having initial trees for the former (Figure 8b) – one for each coordinator with its appropriate semantics – and auxiliary trees for the latter (Figure 9b), so that the compositional rules can treat the two cases distinctly. Note that this means that the lexical ambiguity of “and” and “or” corresponds to a structural ambiguity with respect to this aspect of discourse grammar.

4.1.2 Auxiliary Trees The grammar uses auxiliary trees in two ways: (a) for discourse units that continue a description in some way; and (b) for discourse adverbials. Again we describe each in turn.

First, auxiliary trees anchored by punctuation (e.g. *period*, *comma*, *semi-colon*, etc.) (Figure 9a) or by *scopal coordination* (Figure 9b) are used to provide further description

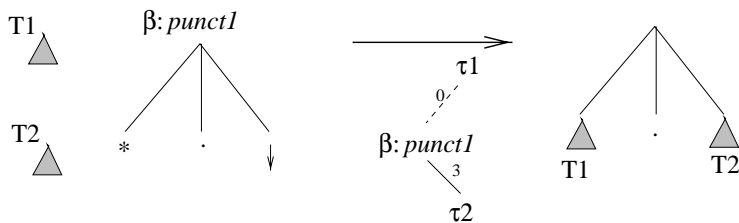


Figure 10
TAG derivation of Example 69

of a situation or of one or more entities (objects, events, situations, states, etc.) within the situation¹⁹ 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:

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

Figure 10 shows the TAG derivation of Example 69. To the left of \rightarrow are the elementary trees to be combined: T1 stands for the LTAG tree for clause 69a, T2 for clause 69b, and $\beta : punct1$, for the auxiliary tree. In the derivation, the foot node of $\beta : punct1$ is adjoined to the root of T1 and its substitution site filled by T2, resulting in the tree to the right of \rightarrow . (A standard way of indicating TAG *derivations* is shown under \rightarrow , where dashed lines indicate adjunction, and solid lines, substitution, with each line labelled by 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 lexicalised 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 9c. Adjoining such an adverbial to a clausal/sentential structure contributes to how information conveyed by that structure relates to the previous discourse.

Obviously, this discourse grammar suffers from lexical ambiguity. First, as already noted, we have different trees for “and” (and for “or”), depending on whether it contributes an independent relation (in which case, it anchors an initial tree), or whether it merely extends the “scope” of the previous clause, so that the same relation holds with the previous discourse. Secondly, 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 7), while in the second, they will anchor a simple auxiliary tree (Figure 9c). These lexical ambiguities correlate with semantic ambiguity.

4.2 Example Derivations and Interpretations

It should be clear by now that our approach aims to explain discourse semantics in terms of a product of

- compositional rules on syntactic structure
- anaphor resolution

¹⁹ 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).

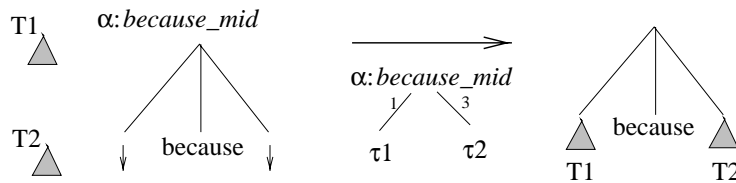


Figure 11

Derivation of Example 70a. The derivation tree is shown below the arrow, and the derived tree, to its right.

- inference triggered by adjacency

much as clausal semantics can be explained in this way. For the compositional part of semantics in LTAG (in particular, computing interpretations on derivation trees), we follow Joshi and Vijay-Shanker (1999). Roughly, they compute interpretations on the derivation tree by 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 non-deterministic: The resulting ambiguities in interpretation may be admitted as genuine, or they may be eliminated by a lexical specification.

Here we try to show rather informally how this lexicalised discourse grammar and an interpretation process on its derivations can explain the interpretations of several examples. To start with, consider the following variants on a familiar example:

- (70) 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.

We let $T1$ stand for the LTAG parse tree for “you shouldn't trust John”, $\tau1$, its derivation tree, and $\text{interp}(T1)$, the eventuality associated with its interpretation. Similarly, we let $T2$ stand for the LTAG parse tree for “he never returns what he borrows”, $\tau2$, its derivation tree, and $\text{interp}(T2)$, the eventuality associated with its interpretation.

Example 70a involves an initial tree ($\alpha:\textit{because_mid}$) anchored by “because” (Figure 11). Its derived tree comes from $T1$ substituting at the left-hand substitution site of $\alpha:\textit{because_mid}$ (index 1) and $T2$ at its right-hand substitution site (index 3). Through a compositional rules on the resulting derivation tree, we get that the interpretation of $T2$ is an explanation for the interpretation of $T1$ – i.e. $\textit{explanation}(\text{interp}(T2), \text{interp}(T1))$. (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 (70a), Example 70b employs an auxiliary tree ($\beta:\textit{punct1}$) anchored by “.” (Figure 12). Its derived tree comes from $T2$ substituting at the right-hand substitution site (index 3) of $\beta:\textit{punct1}$, and $\beta:\textit{punct1}$ adjoining at the root of $T1$ (index 0). Compositional interpretation on the resulting derivation tree yields merely that $T2$ continues the description of the situation associated with $T1$ – i.e., $\textit{elaboration}(\text{interp}(T2), \text{interp}(T1))$. Further inference triggered by adjacency leads to a defeasible conclusion of causality between them – i.e., $\textit{explanation}(\text{interp}(T2), \text{interp}(T1))$. That is, this conclusion can be denied without a contradiction – e.g.

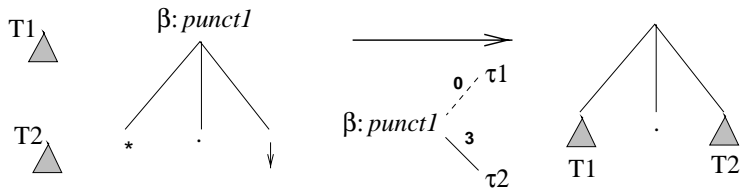


Figure 12
Derivation of Example 70b

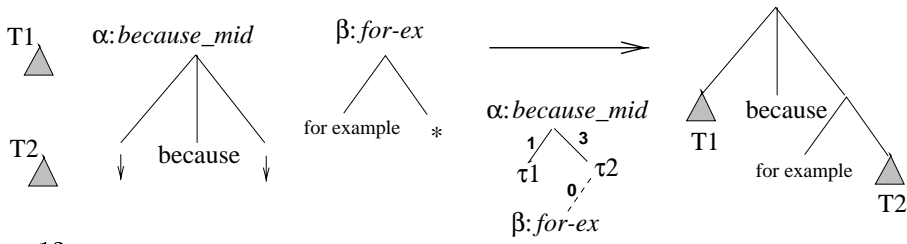


Figure 13
Derivation of Example 70c

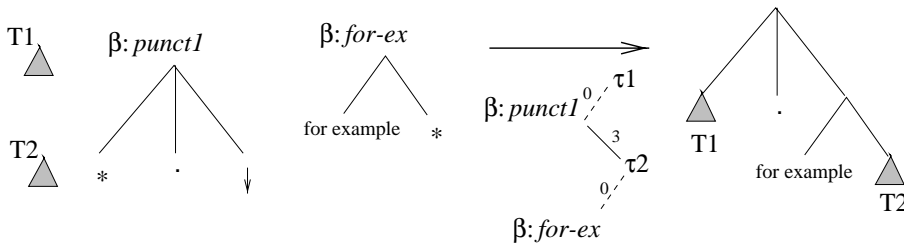


Figure 14
Derivation of Example 70d

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

Example 70c adds “for example” to (70a), which adds the auxiliary tree $\beta:for_ex$ to the set used in analysing it. $\beta:for_ex$ adjoins at the root of $T2$ (Figure 13). Since the relation between the interpretations of $T1$ and $T2$ is $explanation(interp(T2),interp(T1))$, “for example” contributes the interpretation

$$exemplification(interp(T2), \{X \mid explanation(X,interp(T1))\}).$$

That is, his never returning what he borrows is one instance of a set of explanations.

Example 70d adds “for example” to (70b). As in Example 70b, the adjacency-triggered relation between the interpretations of $T2$ and $T1$ is $explanation(interp(T2),interp(T1))$. So “for example” again contributes the interpretation

$$exemplification(interp(T2), \{X \mid explanation(X,interp(T1))\}).$$

Since we have referred to Example 10 (given here as Example 72) so often in the earlier parts of the paper, we now give its analysis in DLTAG.

- (72) 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 15, this example involves three initial trees ($\alpha:contrast$, $\alpha:so$, $\alpha:because_mid$) for the structural connectives, and one auxiliary tree ($\beta:then$) for the discourse adverbial “then”, as well as the initial trees for the four individual clauses $T1$ - $T4$. The interpretation contributed by “then”, after its anaphoric argument is resolved to $interp(T2)$ ²⁰, is

$$\iota4: after(interp(S4), interp(T2)).$$

The interpretations derived compositionally from the structural connectives are:

$$\begin{aligned} \iota1: & result(interp(T2), interp(T1)) \\ \iota2: & explanation(interp(S4), interp(S3)) \text{ or } explanation(\iota4, interp(S3)) \\ \iota3: & contrast(\iota1, \iota2) \end{aligned}$$

Finally, we should point out that discourses that seem to be close paraphrases of each other, such as those in Example 73, can (in this approach) get their interpretations in different ways:

- (73) a. If John took growth hormones for a year, he'd probably shoot up another three inches.

²⁰ How this is done is not addressed here.

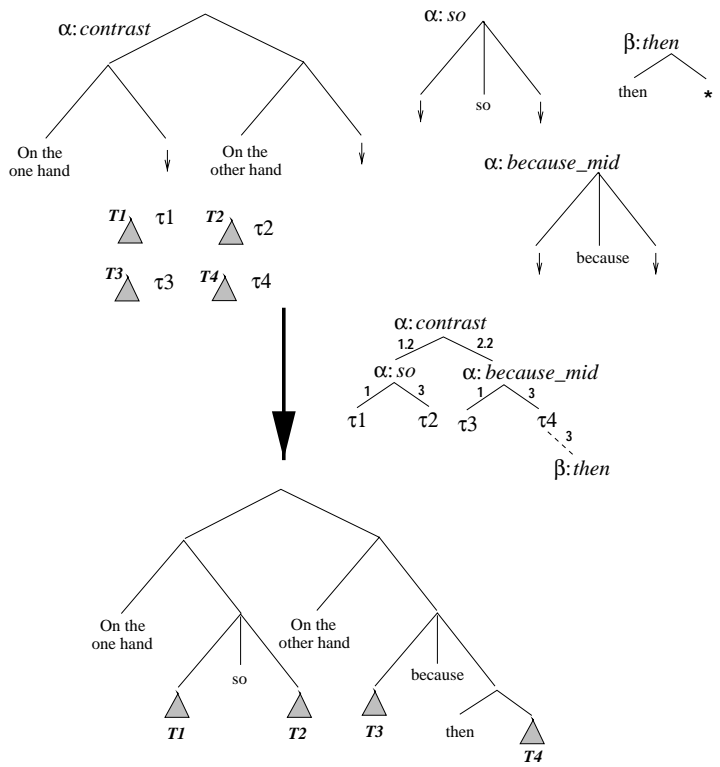


Figure 15
Derivation of Example 72

- b. Suppose John took growth hormones for a year. He'd probably shoot up another three inches.

Besides the trees corresponding to the individual clauses, (73a) would be analysed using an initial tree for “if”, while (73b) would use an auxiliary tree conveying simply *elaboration* (i.e., that shooting up another 3” further describes the situation in which John takes growth hormones). This is justified by the fact that sentences starting with “suppose” don't have to be followed (and very often are not followed) by any description of the (possible) consequences of the supposition. Nevertheless, one may want to recognize that the modal situation underlying the two examples is the same. However, in the current approach, this is not a fact about the compositional semantics of (73b), but derives from other processes in discourse interpretation, namely anaphor resolution and defeasible inference.

5 Conclusion

We hope by now to have convinced the reader of several things, including the benefits of treating various discourse adverbials as anaphors and, more generally, of seeing how discourse adverbials and structural connectives contribute to discourse semantics, rather than simply treating them as cues to particular discourse relations. We are clearly not the first to have proposed a grammatical treatment of low-level aspects of discourse semantics (Gardent, 1997; Polanyi and van den Berg, 1996; Scha and Polanyi, 1988; van den Berg, 1996). However, we believe that the key to the problem lies in recognizing discourse adverbials as anaphors and understanding the parasitic nature of adverbials like “for example”. This enables a degree of simplicity that has not before been possible.

Of course, a few roughly done examples do not make a complete grammar or syntax-semantics interface, and there is clearly a lot more that needs to be done in order to derive anything practical from this work. Still, we hope that we have convinced the reader of two main things:

- that one does not have to treat the systematic syntax and semantics of discourse in any way differently than clause-level syntax and semantics;
- that anaphora involves categories other than pronouns and definite NPs, and thereby plays a more significant role in the computation of discourse semantics.

Acknowledgments

The authors would like to thank Kate Forbes, Katja Markert, Natalia Modjeska, Rashmi Prasad, Mark Steedman, members of the University of Edinburgh Dialogue Systems Group, and participants at ESSLLI'01, for helpful criticism as the ideas in the paper were being developed.

References

- Asher, Nicholas and Alex Lascarides. 1999. The semantics and pragmatics of presupposition. *Journal of Semantics*, 15(3):239–300.
- Asher, Nicholas and Michael Morreau. 1991. Commonsense entailment. In *IJCAI'91, Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 387–392, Sydney, Australia.
- 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.
- Biernier, Gann. 2001a. Alternative phrases and natural language information retrieval. In *Proceedings of the 39th Annual Conference of the Association for Computational Linguistics*, Toulouse, France, July.
- Biernier, 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.
- Carter, David. 1987. *Interpreting anaphors in natural language texts*. Ellis Horwood Ltd, Chichester.
- Cosse, Michel. 1996. Indefinite associative anaphora in French. In *Proceedings of the IndiAna Workshop on Indirect Anaphora*, University of Lancaster, UK.
- Dale, Robert. 1992. *Generating Referring Expressions*. MIT Press, Cambridge MA.
- Evans, G. 1980. Pronouns. *Linguistic Inquiry*, 11:337–362.
- 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.
- Gardent, Claire. 1997. Discourse tree adjoining grammars. Claus report nr.89, University of the Saarland, Saarbrücken.
- Hahn, Udo, Katja Markert, and Michael Strube. 1996. A conceptual reasoning approach to textual ellipsis. In *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 572–576, Budapest, Hungary.
- Hardt, Dan. 1999. Ellipsis and the structure of discourse. In *Proceedings of the Berlin ZAS 1999 Workshop on Ellipsis and Information Structure*, Berlin.
- Hellman, Christina and Kari Fraurud. 1996. *Proceedings of the IndiAna Workshop on Indirect Anaphora*. University of Lancaster, UK.
- Hobbs, Jerry. 1985. Ontological promiscuity. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 61–69, Palo Alto, CA. Morgan Kaufmann.
- Hobbs, Jerry. 1990. *Literature and Cognition*. Center for the Study of Language and Information, Stanford CA. CSLI Lecture Notes No. 21.
- 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 Press, Dordrecht, 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. 1978. A note on partial matching of descriptions: Can one simultaneously question (retrieve) and inform (update)? In *Proceedings of 2nd Workshop on Theoretical Issues in Natural Language Processing (TINLAP-2)*, pages 184–186, University of Illinois.
- 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. 1999. Compositional semantics with lexicalized tree-adjoining grammar (LTAG): How much underspecification is necessary? In *Proceedings of the Third International Workshop on Computational Semantics*, Tilburg, Netherlands, January. Revised version appears in this volume.
- Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer, Dordrecht NL.
- Kehler, Andrew. 1995. *Interpreting Cohesive Forms in the Context of Discourse Inference*. Ph.D. thesis, Division of Applied Sciences, Harvard University.
- 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*. John Benjamins Publishing.
- Kruijff-Korbayová, Ivana and Bonnie Webber. 2001. 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. PhD Thesis, Catholic University of Brabant.
- Lakoff, George. 1971a. The role of deduction in grammar. In Charles Fillmore and Terence Lagedoen, editors, *Studies in Linguistic Semantics*. Holt, Rinehart and Winston, pages 62–70.
- Lakoff, Robin. 1971b. If's, and's and but's about conjunction. In Charles Fillmore and Terence Lagedoen, editors, *Studies in Linguistic Semantics*. Holt, Rinehart and Winston, pages 114–149.
- 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 DE.
- 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>.
- Modjeska, Natalia Nygren. 2001. Towards a resolution of comparative anaphora: A corpus study of 'other'. In *PAPACOL*, Italy.
- Neale, Stephen. 1990. *Descriptions*. MIT.
- 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.
- 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.
- Salmon-Alt, Suzanne. 2000. Interpreting referring expressions by restructuring context. In *ESSLLI'2000 Student Session*, Birmingham, UK.
- 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, Netherlands, January.
- Steedman, Mark. 1996. *Surface Structure and Interpretation*. Linguistic Inquiry Monograph 30, 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*, Cambridge MA. MIT Press.
- Traugott, Elizabeth. 1995. The role of the development of discourse markers in a theory of grammaticalization. Presented at ICHL XII, Manchester. Version of 11/97 available at <http://www.stanford.edu/~traugott/ect-papersonline.html>.
- Traugott, Elizabeth. 1997. The discourse connective after all: A historical pragmatic account. 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., pages 181–237.
- 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 DE.

- 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, Canada.
- Webber, Bonnie, Alistair Knott, and Aravind Joshi. 1999. Multiple discourse connectives in a lexicalized grammar for discourse. In *Third International Workshop on Computational Semantics*, pages 309–325, Tilburg, The Netherlands.
- 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.
- XTAG-Group, The. 2001. A Lexicalized Tree Adjoining Grammar for English. Technical Report IRCS 01-03, University of Pennsylvania. See <ftp://ftp.cis.upenn.edu/pub/ircs/technical-reports/01-03>.