Bridging

Nicholas Asher
Department of Philosophy,
University of Texas at Austin

Alex Lascarides
Centre for Cognitive Science and HCRC,
University of Edinburgh

Abstract

In this paper, we offer a novel method for processing given information in discourse, paying particular attention to definite descriptions. We argue that extant theories don't do justice to the complexity of interaction between the knowledge resources that are used. In line with Hobbs (1979), we claim that discourse structure—as defined by the rhetorical connections between the propositions introduced in the text—is an important source of knowledge for processing given information, because rhetorical relations can change the semantic content of it. We model the processing of given information as a byproduct of computing rhetorical structure in a framework known as SDRT (Asher, 1993), which formalises the interaction between discourse structure and compositional and lexical semantics in determining semantic content. We demonstrate that it provides a richer, more accurate interpretation of definite descriptions than has been offered so far.

1 Introduction

Utterances containing definite descriptions and other constructions like clefts give rise to a distinction between *given* and *new* information (Stalnaker 1978, Clark 1977, and others). Given information is conventionally required to convey information the speaker can assume the hearer already knows, while new information does not. Given information is also known as *presupposition*. For utterances containing a definite description or a cleft construction to be coherent, the hearer must be able to compute the unique antecedent that the speaker intended for the given information (Clark, 1977). But a suitable unique antecedent is not always present on the basis of what's explicitly been said. When this happens as in (1) or in (2),[1] the hearer must construct it.

(1) I met two interesting people last night at a party.
The woman was a member of Clinton's Cabinet.

(2) In the group there was one person missing. It was Mary who left.

In (1), *the woman* supplies the given information that a woman exists. The context fails to supply a unique antecedent explicitly. However, the hearer can construct it. He can draw the implicature that the woman is one of the two interesting people the speaker met last night, and therefore, to guarantee the uniqueness of this antecedent, the other person he met must

have been a man. These implicatures, which arise as the hearer attempts to construct the unique antecedent for the given information, are known as *bridging inferences*. Computing this bridging inference is essential to computing the coherence of the discourse. Without it, there is no connection at all between the sentences, no connection between the woman and the people the speaker met the previous night. The bridging inference also adds content to the definite description: the woman's one of the people the speaker met last night.

In (2) the *it*-cleft conveys *one and only one person left* as given information. But no leaving is explicitly mentioned in the context. Here, the hearer can draw an implicature that the reason the person is missing is because he/she left. The bridging inference is also essential for computing the coherence of the discourse in (2).

Extant theories on bridging have argued that they are influenced by a wide variety of knowledge sources. But we claim they don't do justice to the complexity of the interaction between given information and context. In particular, they all ignore what we will claim is a vital ingredient: *rhetorical relations*—such as *Elaboration*, *Parallel* and *Narration*, among others—which are used to connect together the propositions introduced in a text.

Each rhetorical relation has coherence constraints specific to it, which specify the particular relationship that must hold between the semantic contents of the constituents it connects.[2] For example, constituents related by *Background* and *Narration* must have a common topic. When the compositional semantic content of the constituents doesn't meet the coherence constraints of the rhetorical relation, it can be added in a constrained manner. This captures the intuition that the content of a sentence is determined in part by the way it connects to the previous discourse. Furthermore, the added semantic content is similar to, but more constrained than, Karttunen's (1974) and Stalnaker's (1978) notion of *accommodation*, which has become the standard way in which current theories of presupposition deal with cases where given information isn't in the context, and has to be added.

This inference from a rhetorical relation to additional semantic content is on occasion a bridging inference. Different rhetorical relations can trigger different bridging inferences. In this paper, we will demonstrate that this new perspective on bridging replaces accommodation; the analysis of several presupposition triggers can do without this central tenet of current presupposition theories.

Our approach to presupposition comes in two parts. First, we agree with Van der Sandt's (1992) insight that presuppositions are anaphoric. We also claim that presuppositions are underspecified in a particular way: the presupposition trigger introduces a binding relation B by means of which some element in the presuppositional content is related to some antecedently specified element. This relation B is typically under-determined by the presupposition trigger.

The second part, and the principal one that we will develop in this paper, is that the rhetorical relations used in the discourse context help specify the relation B introduced by a presupposition trigger. It can be the identity relation, but need not be. We will provide a precise and systematic account of how people compute bridging inferences, by formalising the process in a formal representation of discourse semantics known as SDRT (Asher, 1993), which incorporates rhetorical relations. An accompanying formal theory of pragmatics known as DICE (Lascarides and Asher, 1993) models how the construction of this discourse semantics is influenced by the reader's background knowledge. By mixing these ingredients, we hope to furnish a richer theory of how given information is processed than has been attempted so far, where both background knowledge and rhetorical relations play a central role.

Karttunen (1974), Heim (1992) and van der Sandt (1992) incorporate a notion of accommodation, where one adds the presuppositional material to the context if it's not already there (cf. Lewis, 1979). But there is no need to specify the connection between this added material and other objects in the discourse. In our approach, we reject these accommodation mechanisms for presuppositions of *familiarity*—familiar in the sense that the presupposed entity should have already been introduced in the discourse context—for reasons given in section 2. Rather, like van der Sandt, we view presuppositions as anaphoric; but in contrast to van der Sandt, connections between familiar presuppositions and objects in the context *must* be provided. We claim that if there are no ways of binding these presuppositions of familiarity to elements in the discourse context, then we have cases of presupposition failure.

However, not all presuppositions are familiarity ones. Presuppositions of verbs of creation and destruction (Asher 1996) seem to fall outside this class. These presuppositions are satisfied in the context if they hold just prior to the eventuality described by the verbs. In this paper, we will concentrate only on presupposition triggers that generate familiarity presuppositions. Definite descriptions are the best known presupposition triggers of this type.[3]

The technique of computing given information in SDRT was used in Lascarides and Oberlander (1993) to process the presuppositional content of *when*-, *before*- and *after*-clauses in multi-sentence discourse. They formalised an interaction between background knowledge and the way given information is incorporated in the discourse context. We adopt some of their ideas here, but we replace their procedural approach with a declarative one, and we cover a wider variety of presupposition triggers. Further, Lascarides and Oberlander ignored the information flow from rhetorical relations to semantic content. In contrast, this information flow will be a crucial ingredient in the theory presented here. In the next section, we will examine some simple examples with definite descriptions that motivate our approach.

2 Some simple Examples

2.1 Familiarity and Bridging

Many have argued that definite descriptions have familiarity presuppositions, in the sense that the entity denoted should be linked to something already introduced in the discourse context. Often the link is identity as in (3):

(3) a. Lizzie met a dog yesterday.
b. The dog was very friendly

Intuitively, the presupposition triggered by *the dog* is satisfied in this context, because the dog to which the definite description refers has already been introduced in (3a). But the link isn't always identity. Indeed, such cases are pervasive. Consider text (4).

(4) a. I took my car for a test drive.
b. The engine made a weird noise.

The natural interpretation of *the engine* in (4b) is that it's part of the car mentioned in (4a). This interpretation is enough to satisfy the presuppositions of the definite. But there isn't an engine

explicitly mentioned prior to (4b). The link between the car and the engine is inferred from background knowledge and the presence of certain elements in the context. Indeed, this link must be inferred to maintain discourse coherence, because otherwise there is no connection between the sentences, and no connection between the engine and the car.

Contrast (3) and (4) with (5), in which such a link cannot be constructed.

(5) a. It was morning.
b. A cat meowed.
c. ?The dog barked.

(5) is infelicitous; it is an instance of presupposition failure, because we can't construct any link between the discourse referent introduced by *the dog* and any entity mentioned in (5a-b). Finally, contrast these with (6) from Charniak (1983):

(6) Jack was going to commit suicide. He got a rope.

As Clark attests, referring expressions which aren't traditionally thought of as presupposition triggers can also express given information. In (6), Charniak claims there is a bridging inference connected with the indefinite description *a rope*: one infers that it is to be used in the suicide. Indeed, this bridging inference shares two properties that we observed in (1) and (4). First, it is essential to the coherence of the discourse. And second, the inference changes the content of the description: it's a rope to be used for suicide, and not simply a rope.

How do some existing theories of definite descriptions analyse these examples? Van der Sandt (1992) relates given information to context explicitly, by viewing it as anaphora with semantic content. He couches the theory in Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), so as to exploit the constraints on the accessibility of antecedents provided there. First, he demarcates the given information from the new in the DRS, to allow them to be processed differently. He first deals with the given information. If there isn't a suitable antecedent to which this information can bind in an accessible part of the DRS which represents the discourse context, then all is not lost. He adds one. This accommodation in the style of Lewis (1979) is subject to certain constraints, such as the result of the addition should be logically consistent. However, it's important to stress two things about his method of accommodation. First, it isn't constrained by any conditions regarding uniqueness. Second, it isn't influenced by background knowledge.

Van der Sandt has a nice account of (3). The discourse referent introduced by the definite must be identified with that introduced by the indefinite in (3a), since it's an accessible antecedent in the DRS. Our account of (3) will also boil down to this analysis. On the other hand, van der Sandt's account doesn't fare so well with (4), (5) or (6). Since there is no presupposition trigger in (6), there is no attempt to relate the rope mentioned to previous entities in the discourse. In (5), there isn't an available antecedent, so by his account we simply add an appropriate discourse entity to the context. Because accommodation is so unconstrained, van der Sandt's account, and the account of others such as Heim (1992), don't predict the presupposition failure of the sort exemplified in (5). They don't distinguish (5) from a similar discourse with the definite description replaced by the indefinite noun phrase *a dog*.

Finally, his analysis of (4) fails to be influenced by the background knowledge that cars typically have engines. An engine isn't found in the DRS representing the context (4a), and so one

is added. But it's not linked to the car. Moreover, if he were to add a uniqueness constraint on accommodation to distinguish (5a,b) vs. (5a,c)—there would be presupposition failure because one can't plausibly assume there's only one dog in the model—then he would wrongly predict that (4) was incoherent, unless in tandem with adding this uniqueness constraint, he also modelled the bridging inference. For without this bridging inference, the uniqueness constraint amounts to an assertion that there's only one engine in the model, rather than the constraint that the car referred to in (4a) has only one engine.

One might argue that bridging inferences are not part of truth conditional content, and therefore should not be modelled in a formal semantic representation of the sentence. Rather, one should add such content when the DRS representing (4) is sent to the pragmatic component for further processing. However, there are problems with such an approach. We've already mentioned that the bridging inference is central to satisfying the uniqueness constraint associated with given information. So, if we were to leave bridging inferences unaccounted for, then the uniqueness condition cannot be verified. This means that one shouldn't accommodate given information in a DRS, and consider bridging inferences after that. Bridging inferences are essential to verifying the conditions under which accommodation can occur in the first place. If bridging inferences happen after the DRS is constructed, then contrary to van der Sandt's theory, the DRS construction procedure should not incorporate mechanisms for accommodation at all.

Bos *et al.* (1995) address shortcomings in van der Sandt's theory by extending it with lexical knowledge. They assume a generative model of the lexicon (Pustejovksy 1991, 1995), where lexical semantic information and real world knowledge are not seen as necessarily distinct. Instead, linguistic processes have limited access to world knowledge, which could therefore interact with knowledge of language and become conventionalised in various ways. Conventionalising this information helps account for the way it influences certain linguistic phenomena such as thematic role structure (Sanfilippo 1992) and metonymic constructions (Pustejovsky 1991, Briscoe *et al.*, 1990). In particular, lexical entries for artifacts have a *qualia structure*, which represents a limited amount of information about the properties of that artifact: what it's made up of, what one does with it, and so on.

Bos *et al* use the qualia structure to perform bridging inferences. They amend van der Sandt's model of presupposition as follows: one first tries to bind it; if that fails, one tries to link it to elements of the qualia structure of entries in the accessible parts of the DRS; and only then, failing that, does one attempt to add it. So in (4), one attempts to link the engine before one attempts to add it. It links successfully to the QUALIA:CONSTITUENCY value of the lexical entry for *car*, which in turn is in the accessible DRS representing the discourse context (4a), because this value in the lexical entry contains an engine (to reflect the fact that cars have engines as parts).

However, this extension to van der Sandt's theory also has shortcomings. First, it suffers from some of the same problems as van der Sandt's with regard to the handling of presupposition failure. Accommodation is still just as unconstrained as before, and so they fail to model presupposition failure in examples like (5). Secondly, they fail to model bridging inferences in the absence of presupposition triggers (e.g., (6)). Finally, although lexical semantic knowledge is a useful source of information for modelling bridging, it isn't sufficient. The background information which underlies bridging inferences is open-ended, and there is no independent motivation for including *all* this arbitrary information in the lexicon. To illustrate the problem, consider (7):

(7) a. I arrived at 3pm today.

b. The camel is outside and needs water.

It's implausible to assume that the inference that I arrived by camel is achieved solely through lexical semantic information. If one were to assume this, then the lexicon would become unwieldy, and productive lexical phenomena would in general overgenerate word senses (cf. Verspoor, 1996).

Indeed, even in this simple example, there is a wide variety of knowledge that's used to support the bridging inference. One infers that the camel was the mode of transport in virtue of several things. First, one uses the meanings of the words; for example, *arrive* is a motion verb, and so it is plausible to assume that there was a mode of transport. Second one uses world knowledge; for example, camels can be used as a mode of transport. But crucially, one uses the above lexical knowledge and world knowledge, as opposed to other world knowledge about camels, because this knowledge must be utilised to meet the coherence constraints imposed by the way (7b) connects to (7a). (7a) is stative, and according to Lascarides and Asher (1993), states normally provide background information. If this were the case here, however, then the camel being outside would temporally overlap the arrival, thereby blocking the camel from being the mode of transport because of the lexical semantics of *arrive*. But another coherence constraint on *Background* is that the constituents must have a common topic (Lascarides and Asher, 1993). And if one is forced to assume that the camel has nothing to do with the arrival, then a suitable topic can't be constructed, leading ultimately to discourse incoherence. Intuitively, one tries to interpret constituents so that you obtain the best discourse coherence you can. Here, assuming the camel isn't the mode of transport leads to discourse incoherence. On the other hand, assuming the camel is the mode of transport allows us to interpret the discourse coherently—my arrival caused the camel to be outside, and so the propositions are connected by *Result*. Thus, if we formalise the coherence constraints of different rhetorical relations, together with the principle that you aim for discourse coherence, one can compute the link between the camel in (7b) and its discourse context.

Verifying coherence constraints imposed by the rhetorical relation that connects the sentences together has two important effects. First, it brings certain lexical knowledge and world knowledge into play. Second, it adds semantic content to the constituents that are connected (cf. Asher, 1993). We now know that the object described in (7b) isn't just a camel; it's a camel that I used as a mode of transport in the arrival event mentioned in (7a). Thus the description of the object, and hence its unique antecedent, have changed. At any rate, it's important to note that coherence constraints on rhetorical relations like *Background* and *Result*, as well as truth conditional content, lexical knowledge and world knowledge, affect the interpretation of definite descriptions.

There are some descriptions that seem resistant to presupposition failure. Compare (8a) to (8b) or to (5):

(8) a. The Kremlin is in Moscow.

b. ?The house is in Moscow.

In (8), the given information that there is a Kremlin and that there is a house can't be connected to objects mentioned in the preceding discourse, because there aren't any. But (8a) is better than (8b). van der Sandt's (1992) account of accommodation works for (8a): one adds a

Kremlin to the context before one computes that it is in Moscow. But (8b) is as bad as (5) is, and van der Sandt doesn't account for this. We think it can't simply be accommodation that satisfies the familiarity presupposition in (8a). There is something special about such descriptions. We tentatively propose that descriptions like *Kremlin*, *White House* and *House of Commons* behave like proper names. In contrast, the semantics of descriptions like *the present King of France* function in a particular way, such that the presuppositions of uniqueness and familiarity get satisfied in virtue of that semantics. We will return to this in section 4.2.

2.2 Uniqueness and Presuppositions

Russell (1905) argued that part of the semantic content of a definite description *the* ϕ was that there was one and only one ϕ. Since Russell's seminal argument, many have pointed out that this uniqueness requirement is too strong; at best uniqueness can hold only of a restricted domain, domain restriction being part of all quantificational structures.

Chierchia (1995) and von Fintel (1994) have suggested that the Russellian uniqueness condition suffices for definite descriptions, so long as one includes in the representation of quantificational structures in general, and definite descriptions in particular, an underspecified relation B which relates the element referred to by the description to something in the discourse context. This serves to restrict the domain. The exact properties of B are inferred by the way the description connects to the discourse context. The definite description *the N* is represented in (9a), which is equivalent to (9b):

(9) a. $\iota x(B(x,y) \wedge N(x))$

b. $\exists x \forall z((N(z) \wedge B(z,y)) \leftrightarrow (z = x))$

But this analysis fails to do justice to the interaction between resolving the binding relation B, and the impact of the uniqueness constraint. In particular, the Russellian uniqueness condition is still too strong for certain resolutions of B. To see this, consider (10).

(10) a. John dented the door of his car.

b. John had an car accident. The door was damaged.

c. John shot himself in the foot.

In (10a,b), one can define B so that the door is part of John's car, but then the Russellian condition amounts to the car having only one door, contrary to intuitions. Similarly, in (10b), one can define B so that the foot is John's foot, but the Russellian uniqueness condition amounts to the condition that John has only one foot, contrary to intuitions.

One could entertain three possible ways out of this problem, while preserving the Russellian uniqueness condition. The first is that in examples such as (10) the definite description refers to a property rather than a (token) object. Under this assumption, Russellian uniqueness requires there to be a unique set of objects satisfying the relevant property, rather than the relevant property denoting a singleton set. So in (10a), for example, there must be a unique set of objects that satisfies the property of being a door of John's car, but this set may contain more than one element (and so John's car has more than one door). However, this assumption is untenable. First, there will be a type clash in logical form, since the argument required by

the predicate introduced by the main verb will be of a different type. Second, if one can always apply this strategy, then the uniqueness condition—even the Russellian one—becomes vacuous, because *all* property expressions denote unique sets at a world (that may contain more than one element).

The second two ways we explore for resolving this problem with the uniqueness condition both involve changing the content of B. First, one could assume that in (10a) B includes the rest of the content in the sentence, making (10a) mean: John dented the door of John's car which John dented. Then the Russellian condition on uniqueness is satisfiable, even though John's car has more than one door. Alternatively, one could add to the content of B that the door in question is the (unique) one that the speaker had in mind when he said *the door*. One could modify the relation B for (10b) in similar ways. But both these strategies are problematic in general. If one can always add all the contextual information to B, or one can add the condition 'salient to the speaker' to the content of B, then once again, satisfying the uniqueness condition—even the Russellian one—becomes vacuous; it can always be satisfied if this information can be added to B. And so both these proposals will fail in general to model presupposition failure. In particular, it won't model the presupposition failure in (5)—B can be resolved so that (5c) means: the dog which is the agent of the barking event e that occurred that morning, barked. This interpretation of B ensures the Russellian uniqueness condition is satisfiable, since we can assume only one dog was the agent of the barking event e introduced by the VP; and we have also linked this to something in the context, namely *the morning*.

In contrast to Chierchia, van der Sandt's and Bos *et al*'s theories contain no constraints on uniqueness at all. Consequently, Bos *et al*'s analysis of (1) captures the inference that the woman is one of the two people I met last night, because the lexicon includes the information that women are a subtype of people. But it fails to model the further inference that the other person must have been a man. Let's examine how adding the following constraint to Bos *et al.*'s theory improves the situation:

- `A Simple Uniqueness Constraint:`
  A presupposition can bind or link to an antecedent x only if x was the only choice. That is, it was the one antecedent with conditions on it that match those in the presupposition.

This constraint is a simple way of stipulating Clark's (1977) observation that there must be an antecedent for a definite description that is uniquely salient in the discourse context.

One might think at first glance that this would be sufficient. It would enable them to compute that one of the people in (1) that the speaker met was a man.

(1) I met two interesting people last night at a party.
The woman was a member of Clinton's Cabinet.

However, like the uniqueness condition in Chierchia's and von Fintel's analyses, this fails to take into account the complex interactions between uniqueness conditions and the way the presupposition binds to the context. The above constraint, when added to Bos *et al*'s theory, would incorrectly predict that (10a,b,c) are odd.

We will adopt Chierchia's position that presuppositions introduce an underspecified binding relation B, which get resolved through discourse context. However, we will take this account further by spelling out a detailed formal theory of exactly how B gets resolved, and in contrast

to von Fintel (1994), we will use rhetorical relations to do this. Second, our analysis will be different from Chierchia's, in that we will assume that in some contexts, the uniqueness condition is relaxed to $\top$. To this end, we will replace the quantifier ι in Chierchia's analysis with a separate uniqueness condition—so the semantics of *the N* becomes something like $\exists x(B(x,y) \wedge N(x) \wedge \textit{unique}_N(x))$—where the semantics of $\textit{unique}_N(x)$ is constrained by the values of other elements on the syntax/semantics interface, as well as the specification of the binding relation B. $\textit{unique}_N(x)$ will be equivalent to a Russellian uniqueness condition in some contexts, but not all of them. (10a,b,c) will be cases where $\textit{unique}_N(x)$ amounts to $\top$. We will specify the conditions under which $\textit{unique}_N(x)$ is $\top$ in section 4.1.

2.3 Definite Descriptions and Discourse Structure

We suggested earlier that the rhetorical relations linking propositions introduced in the discourse course affects the interpretation of definite descriptions. The idea that discourse structure affects definite descriptions isn't knew. Grosz and Sidner (1986) constrain the antecedents to definite descriptions via the discourse segmentation of task oriented dialogues, which in turn is determined by the intentional structure of the plan that underlies the task described. Poesio (1993, 1994) merges Grosz and Sidner's framework with a situation theoretic semantics to account for how focus affects the denotation of definite descriptions. Tracking focus and allowing this to influence the available antecedents is a compelling idea. It enables one to capture the intuition that the uniqueness constraint on definite descriptions is closely related to the notion of saliency. For example, Poesio (1994) tracks the motion in (11) below, to infer that the focus of attention at the time when (11b) is processed is Dansville:[4]

(11) a. John took engine E1 from Avon to Dansville.
b. He picked up the boxcar and took it to Broxburn.

By doing this, he is able to infer that the boxcar is in Dansville—that is, he infers additional semantic content for (11b) as a result of tracking focus through the discourse structure. Hence, the uniqueness of the referent is correctly predicted to amount to there being one and only one boxcar in Dansville, as opposed to one and only one boxcar.

Such an account is fine as far as it goes. However, it lacks a detailed formal, *general* theory of how the semantic content of constituents can be modified in the light of the way they connect together in the discourse structure.[5] But the flow from discourse structure to the addition of further semantic content to the constituents is an essential feature of the phenomena described here. Moreover, Poesio's account of how motion determines focus produces the wrong results for other examples that feature other rhetorical relations. This is because his model of discourse structure is based on Grosz and Sidner's, which includes only two discourse relations—*dominance* and *satisfaction precedence*. This is too coarse grained to handle the different semantic effects that different rhetorical relations can have. So, for example, the rhetorical relation in (11a,b′) is *Parallel* rather than *Narration*:

(11) a. John took the engine E1 from Avon to Dansville.
b′. He also took the boxcar.

In contrast to (11a,b), the natural reading of (11a,b′) is one where the boxcar is in Avon. Presumably this is because of the different way that the sentences connect together, which in turn

results in different spatio-temporal effects in the semantic content. But these spatial differences between *Narration* and *Parallel* aren't represented in the theory of discourse structure that Poesio adopts. Just as before, tracking the motion in (11a) leads to the focus of attention being Dansville at the point when (11b′) is processed. And so as in (11a,b), this predicts that the boxcar mentioned in (11b′) is unique in Dansville, contrary to intuitions. Computing that the boxcar was in Avon by recognising John's commonsense plan won't help here either, since to recognise this plan involves computing the rhetorical connection that we've described between the sentences, and yet in Grosz and Sidner's theory, recognising commonsense plans is primary to constructing discourse structure.

One can view the phenomenon that the semantic content of constituents change in the light of the way they connect together as closely related to the concept of focus. The added content affects what's being talked about, and hence what's salient. So a general theory of how discourse structure affects semantic content can be viewed as contributing towards a general theory of focus. We will use this feature in this paper to model bridging inferences, by formalising the process in SDRT (Asher, 1993).

We've given texts where different rhetorical relations have different effects on the semantic content of given information, and hence the unique antecedent. Text (12) provides evidence that constraints on rhetorical coherence can even override default world knowledge when processing given information.

(12) a. John moved from Brixton to St. John's Wood.

b. The rent was less expensive.

Matsui (1995) tested subjects' judgements on where the rent was less expensive in (12). All the subjects knew the world knowledge that rents tend to be less expensive in Brixton than in St. John's Wood. But in spite of this, the majority of informants judged that in (12), the rent being talked about was in St. John's Wood; thereby drawing conclusions which conflicted with their world knowledge. Arguably, information about how the sentences connect together in the discourse structure conflicts with the world knowledge, and ultimately wins over it. So if processing given information ignores discourse structure, then the world knowledge would trigger the wrong results in (12).

We will explain (12) in terms of the rhetorical relation that's used to connect the constituents. (12b) is stative, and so supports a *Background* relation. However, intuitively, one prefers explanations of intentional changes (in this case, moving house), to simple background information that sets the scene for the change. Assuming that we *always* want to maximise discourse coherence, then even if default world knowledge conflicts with this, we infer both *Background* and *Explanation* for these texts. But the *Explanation* that John moved because the rent was less expensive is plausible only if the rent was less expensive in the place he went to: St. John's Wood.

As we've seen, the above texts where rhetorical information affects the content of given information pose challenges for extant theories. They all fail to model the differences between (11a,b) vs (11a,b′), because they don't model how the rhetorical relations connecting propositions in a discourse affect the semantic content of those propositions. In addition, none of these theories model what happens when discourse coherence principles conflict with world knowledge, and so they don't give an account of (12).

We need to analyse definite descriptions in a theory where information flow from rhetorical relations to the semantic content of constituents is taken into account. So we propose to use SDRT (Asher, 1993), where this information flow is a distinguishing feature. SDRT is a theory of discourse semantics designed to explore systematically the interface between semantics, pragmatics and discourse structure. To date it has been used to model the effects of discourse structure on pronominal and temporal anaphora (e.g., Asher 1993, Lascarides and Asher 1993), cognitive states (e.g., Asher and Lascarides 1994, 1996), spatio-temporal interpretation (Asher et al. 1996), lexical interpretation (Asher and Lascarides 1995, Lascarides *et al* 1996, Lascarides and Copestake 1995), and the presuppositional content of the temporal connectives *when*, *before* and *after* (Lascarides and Oberlander, 1993). Here, we will use the pragmatic knowledge resources and constraints on discourse coherence supplied in SDRT to interpret definite descriptions and to offer a new picture of presuppositions in general.

SDRT has three main advantages for our purposes. First, the way discourse structure affects and is affected by semantic content has already been studied extensively in this framework, and an adequate account of definite descriptions must make use of these effects. Second, the basic semantic framework which underlies SDRT (DRT), has already proved useful in specifying constraints on the interpretation of definite descriptions (van der Sandt 1992, Bos *et al* 1995). We will build on this work here. Finally, one of the main features of SDRT is the underlying axiomatic theory DICE (Discourse in Commonsense Entailment) which allows us to infer rhetorical relations, using semantic content and the interpreter's background knowledge as clues (Lascarides and Asher, 1993). DICE is distinctive in that it deals in a principled way with cases where different knowledge sources give conflicting clues about how to interpret a text. We will use this axiomatisation to provide a novel analysis of bridging that records the influence of background knowledge on the process, and we will use DICE's tools for conflict resolution to model why the default world knowledge is 'ignored' in (12).

3 A Crash Course in SDRT

SDRT, which is an extension of DRT (Kamp and Reyle, 1993), is a theory of discourse structure and content. An NL discourse is represented by a segmented DRS (SDRS), which is a recursive structure of DRSs that represent the clauses, linked together with rhetorical relations such as *Narration* and *Parallel* (cf. Hobbs (1985), Polanyi (1985), Thompson and Mann (1987) and others).[6] Building an SDRS involves computing a rhetorical relation between the representation of the current clause and the SDRS built for the discourse context so far. DICE (Lascarides and Asher, 1993) specifies how various background knowledge resources interact to provide clues about which rhetorical relation holds.

To make technical definitions about SDRT below more tractable, we will adopt the labelling techniques for SDRSs from Asher (1996, forthcoming); that is, each SDRS will have a label, and rhetorical relation symbols will take labels of SDRSs as constituents. We make the general notational convention that π labels the SDRS K_π. Using labels in the arguments to rhetorical relations will allow us to simplify the notion of SDRS update—that is the process of extending an SDRS with new information—and it will also enable us to compare it with the simpler DRT update notion, as we'll see shortly.

As in DRT, SDRT uses a level of semantic representation, viz. that of SDRSs, to model the dynamic aspects of discourse interpretation. The way this is done is to define an update func-

tion on semantic representations of a given context and information to be integrated into that context. The update function returns, if successful, a new representation—or a model of a discourse context within which further information can be integrated.

As we've mentioned, updating the discourse context with new information involves computing a rhetorical relation via DICE. We can think of DICE as a 'glue' logic or a logic of information packaging, by means of which we arrive at a more completely specified representation of the content of the discourse.

The glue logic differs from the logic of the "information content", which operates on the representations themselves and whose consequence relation is defined relative to the model-theoretic structures for those representations. Asher (1996, forthcoming) offers several versions of SDRT with several different versions of the logic of information content, following the approach to the logic underlying dynamic semantics developed in Fernando (1994). But these logics have a validity problem that is at least recursively enumerable (r.e.), for the expressive power of the SDRT language themselves is at least that of first order logic.

The glue logic DICE, on the other hand, exploits a glue language, which is decidedly less expressive than that of SDRT itself. In this and in earlier papers (Lascarides and Asher, 1993), it suffices to make the glue language a quantifier free fragment of a first order language augmented by a weak conditional operator $>$, which formalizes generic or defeasible rules of interpretation ($P > Q$ means *If P, then normally Q*), and the logic is shown to be decidable.

There is an information transfer function μ from the SDRS representation itself into the weaker language of DICE, and this allows us to infer rhetorical relations using compositional semantic content and the interpreter's other knowledge resources as clues. This transfer function μ, formalized in Asher (forthcoming), associates the SDRS labels that feature as arguments to rhetorical relations with information in the SDRS that the label refers to; roughly, it takes conditions inside the SDRS and turns them into predicates of labels. So $\mu K_\alpha(\alpha)$ is a set of formulae of the DICE language of the form $\phi(\alpha)$, where ϕ is any SDRS condition. The DICE rules about which rhetorical relation holds between two constituents then exploit these predicates on labels. These default rules enable the interpreter to come to nonmonotonic conclusions about which rhetorical relation to use to bind the constituents in an SDRS together. DICE also contains axioms concerning the semantic effects of rhetorical relations. These rhetorical relations and their semantic effects, that are inferred via DICE, are then used to define the appropriate update function in SDRT.

For example, `Narration` states: If β is to be attached to α with a rhetorical relation, where α is part of the discourse structure τ already (i.e., $\langle \tau, \alpha, \beta \rangle$ holds), and α and β describe events (as opposed to states) (i.e., *event*(e_α) and *event*(e_β) hold), then normally, the rhetorical relation is *Narration*.[7] The `Temporal Consequence of Narration` is a coherence constraint on *Narration* in that it constrains the relationship between the semantic content of the constituents connected by *Narration*; it states that if α and β are related by *Narration*, then α's event precedes β's:

- `Narration:` $(\langle \tau, \alpha, \beta \rangle \wedge$ *event*$(e_\alpha) \wedge$ *event*$(e_\beta)) >$ *Narration*(α, β)
- `Temporal Consequence of Narration:` *Narration*$(\alpha, \beta) \rightarrow e_\alpha \prec e_\beta$

Together, these axioms record the normal rhetorical role played by event sentences in discourse, and they capture the Gricean-style pragmatic maxim *be orderly*: in the absence of in-

formation to the contrary, one assume that the events are described in the order in which they occur. *Narration* also has implications on the spatio-temporal trajectories of objects (Asher *et al.*, 1996). If we assume that every eventuality has a beginning and end location for the actors within it (a source and goal), then we can derive the following constraint (ignoring the effects of adverbials) from the additional assumption that nothing can be in two distinct places at the same time:

- `Spatial Consequence of Narration:`
  $(\textit{Narration}(\alpha, \beta) \wedge \textit{actor}(x, \alpha) \wedge \textit{actor}(x, \beta)) \rightarrow \textit{loc}(x, \textit{source}(e_\beta)) = \textit{loc}(x, \textit{goal}(e_\alpha))$

In words, if $\textit{Narration}(\alpha, \beta)$ holds and α and β share an actor x then the location of x is the same at the end of e_α and the onset of e_β.[8] This theorem, which is derived in DICE from more basic axioms, will play an important role in interpretating the definite description in (11a,b), for example. There's also an axiom which states that narratives have a distinct common topic. We will introduce further DICE axioms in later sections of this paper.

DICE exploits a nonmonotonic notion of validity ($\mid\!\approx$) with several nice properties, which we have discussed in detail elsewhere (Lascarides and Asher 1993). There are three that are relevant for our purposes. First, $\mid\!\approx$ validates Defeasible Modus Ponens (DMP): when the default laws whose antecedents are verified all have consequents that are consistent with the KB and with each other, then all the consequents are nonmonotonically inferred. Second, $\mid\!\approx$ validates Specificity, also known as the Penguin Principle: when conflicting default rules apply, the consequent of the most specific default rule (if there is one) is inferred. And finally, $\mid\!\approx$ is robust in that if $\Gamma \mid\!\approx \phi$ then ϕ will survive as a consequence of the premises Γ augmented with logically independent information.

A distinctive feature of SDRT is that if the DICE axioms yield a nonmonotonic conclusion that the discourse relation is R, and information that's necessary for the coherence of R isn't already in the constituents connected with R (e.g., $\textit{Narration}(\alpha, \beta)$ is nonmonotically inferred, but $e_\alpha \prec e_\beta$ and information about the spatial location of actors is not in α or in β), then this content can be added to the constituents in a constrained manner through the SDRS *Update* process. Informally, $\textit{Update}(K_\tau, K_\alpha, K_\beta)$ is an SDRS which includes (a) the discourse context K_τ, plus (b) the new information K_β, and (c) an attachment of β to α (which is part of τ) with a rhetorical relation R that's computed via DICE, where (d) the content of K_τ, K_α and K_β are modified so that the coherence constraints on R are met.[9] There is some indeterminacy on which constituent α the new information β attaches to: β can attach to any available constituent on the right frontier of τ, where the subordinating relations are *Elaboration*, *Explanation* and $\Downarrow$ ($\alpha \Downarrow \beta$ means that α is a topic for β). Thus there is some indeterminacy in computing the update of old with new information (this is why the *Update* function has three arguments, rather than just the two arguments K_τ for the old information, and K_β for the new).

This process of discourse update is much more complex than the DRT notion of update ($\textit{Update}_{drt}$), which amounts to set union on the discourse referents and DRS-conditions of the DRSs K and K' denoting the old and new information:[10]

- `DRT's Update Function:`
  $\textit{Update}_{drt}(K, K') = (U_K \cup U_{K'}, C_K \cup C_{K'})$

However, the formal definition of SDRT update is defined in terms of $\textit{Update}_{drt}$ and DICE. Let

$\mid\!\sim$ be the proof theoretic counterpart of the DICE consequence relation. Then the $Update_{sdrt}$ function is defined as:

- `SDRT's Update Function:`
  $Update_{sdrt}(K_\tau, K_\alpha, K_\beta)$ is the SDRS $K^\dagger$ such that:
  1. $(\mu(K_\alpha)(\alpha), \mu(K_\beta)(\beta)) \mid\!\sim (R(\alpha,\beta) \wedge \varphi)$; and
  2. $K^\dagger = Update_{drt}(K, [[\beta][\beta : K_\beta[\varphi], R(\alpha,\beta)]])$, where $K_\beta[\varphi] = K_\beta$ together with those conditions specified in φ, where φ is that information needed to satisfy the coherence constraints on R inferred in DICE.

In what follows, we will specify constraints on SDRT's update function. In certain cases, we will replace one update task with another. So in what follows, $Update(K_\tau, K_\alpha, K_\beta) := Update(K_{\tau'}, K_{\alpha'}, K_{\beta'})$ means: replace the task of updating K_τ with K_β via attachment to K_α with the task of updating $K_{\tau'}$ with $K_{\beta'}$ via attachment to $K_{\alpha'}$.

Asher (1996, forthcoming) shows that SDRS updates so defined correspond to transitions on "more model-theoretic" conceptions of discourse contexts than SDRSs. This replicates the correspondence between DRS update and the transition function on information states as sets of model embedding function pairs proved in Fernando (1994). However, updates in SDRT are more complex than in DRT; there is more than set union going on. First, there is indeterminacy on where new information attaches in the old SDRS. Second, the way it attaches is determined by axioms in DICE. SDRT adds two things over and above normal DRSs (or standard dynamic contexts if these are chosen to replace DRSs): it divides up the context into a set of contexts related to each other by rhetorical relations; and it shows how updating such a relational structure may affect several elements in the structure, including semantic content, in a non-local way.

Specifically, our analysis of (11a,b″) involves the techniques for revising semantic content through SDRT's update procedure:

(11) a. John took engine E1 from Avon to Dansville.
b″. He picked up a boxcar
c. and took it to Broxburn.

First, we build the DRSs α and β for the sentences (11a,b″). We can resolve the pronoun in β to *John* because in SDRT, the only available antecedents to pronouns are those that are DRS-accessible in the current constituent (in this case, the DRS β), or those that are DRS-accessible in the constituent to which this constituent is going to be attached. Using DMP on `Narration` and then Modus Ponens on `Axiom on Narration` will correctly predict that the text is a narrative, and John's taking engine E1 from Avon to Dansville precedes his picking up a boxcar. In addition, the antecedent to the `Spatial Consequence on Narration` is satisfied here, since (11a) and (11b′) share an actor (*John*). Therefore, by Modus Ponens on `Spatial Consequence on Narration`, we infer that John is in Dansville (because this is the location of the goal of e_α) when he begins to pick up a box car. What's more, by the lexical semantics of *picking up* (see Asher and Sablayrolles, 1995), the location of the source of this event is the same as the location of its goal, and the object that's picked up is at this location. So, through inferring in DICE that the discourse is a narrative, we

have learnt more about semantic content: the boxcar which is picked up is in Dansville. Thus because of the way $Update_{sdrt}$ is defined, this additional content about the boxcar appears in the SDRS which represents the discourse (11a,b″). We will use this in our analysis of (11a,b) in section 5.2. This inference about the content of the referring expression in (11b″) can be viewed as a bridging inference, similar to that in text (6), since what we have computed here is essentially a relation between an object mentioned in the current constituent, and one mentioned previously. Thus unlike other theories for processing given information, SDRT can model bridging inferences in the absence of presupposition triggers.

DICE can also predict when a text sounds odd. An SDRS is well-defined (written $\downarrow K$) if there are no conditions of the form $x =?$ (i.e., there are no unresolved anaphoric elements), and every constituent is attached with a rhetorical relation. A discourse is incoherent if $\neg \downarrow$ $Update(K_\tau, K_\alpha, K_\beta)$ holds for every available attachment point α in τ. An example where the update is not well-defined is given in (13):

(13) Max opened the door. Mary scores a goal!

Even though the antecedent to `Narration` is verified by (13), one can't attach the constituents with *Narration* because SDRT has a constraint that a *Narration* must have a non-vacuous common topic,[11] and this is violated in (13). Hence, the consequent of `Narration` is inconsistent with the knowledge base (KB) and so it can't be inferred. Furthermore, the other rules in DICE for computing rhetorical relations fail to supply an alternative, for either the antecedent of the relevant default rule isn't verified, or coherence constraints on the rhetorical relation are violated. So a rhetorical relation cannot be inferred, which by the above definition of well-definedness on SDRSs, means that the discourse is incoherent (for details see Lascarides and Asher (1993)).

4 Bridging with SDRT

We will build on DRT's existing analysis of definite descriptions in two ways. First, we will propose a treatment that makes presupposed information the same in kind as new information, except that it may be underspecified in particular ways. Second, given information will introduce anaphoric equations that must be resolved during the SDRS update process, rather than simply in terms of DRT accessibility.

Our theory is similar in spirit to that given in Hobbs (1979), where definite descriptions are resolved as a byproduct of connecting sentences together. However, there are several important differences between his approach and ours. First, Hobbs ignored compositional semantic information and lexical semantics in computing the antecedents to given information. Indeed, he doesn't define the compositional semantics of definite descriptions at all, and he doesn't specify how to translate NL definite descriptions into logical form. We aim to rectify this.

Second, Hobbs *et al.* (1993) aim to produce the cognitively best interpretation of the discourse by means of *abduction* on background knowledge and the content of the discourse given so far; in particular, they use abduction to process given information. The process isn't compositional, in the sense that the meaning of the discourse is a systematic function of the meaning

of the propositions expressed by the clauses and the meanings of the rhetorical relations between them. Rather, Hobbs *et al.* compute assumptions that explain the data at least cost, from a knowledge base that includes *all* information, both linguistic and non-linguistic. This knowledge base doesn't distinguish the various sources of information at all.

In contrast to Hobbs, our theory of discourse semantics is compositional: the meaning of discourse is a systematic function of the meaning of the propositions and the meanings of the rhetorical relations between them, as we described in the previous section.

Third, the nature of the DICE axioms that we use to compute connections between propositions in discourse are very different from Hobbs *et al.*'s axioms for interpreting discourse. In SDRT, the aim is in line with that in DRT. We build a linguistically justified representation of discourse, which represents the dynamic context change potential when it is added to a given context. But the resources by means of which this context change potential is defined, for both linguistic and computational reasons, exploit a logic that is distinct from the logic of information content; that is, the logic that includes the definition of truth for these discourse representations. Indeed, the former logic is not only separate, but weaker than the latter logic. In contrast, in Hobbs *et al*'s abductive framework, the logic of the information content contained in the propositions related in the discourse structure, and the logic for computing rhetorical relations are one and the same.

There are two reasons for splitting the logics of information content and information cohesion in the way we do. First, all the nonmonotonic frameworks, including Hobbs *et al*'s abductive one, require some appeal to consistency tests to draw conclusions. But if one's base logic of information content is already that of first order logic, then adding consistency tests to this logic will take it beyond the boundary of what is recursively enumerable. Our framework for computing rhetorical relations is nonmonotonic. But the base logic is propositional rather than first order logic, because it is kept separate from the logic of information content of discourse (which is first order logic). In contrast to the abductive framework Hobbs *et al.* use, the logic for information packaging we use here is decidable.

Second, by modelling compositional semantics, background knowledge and discourse coherence principles within a single logic, one cannot separate the process of anaphora binding from the semantic content of the discourse as one would wish. If, as appears reasonable, abduction requires some additive measure of cost on the various assumptions made to compute a proof of the discourse, then inconsistent interpretations will always have the highest overall cost, and will be avoided if possible. Consequently, such a framework cannot account for discourses where definite descriptions receive an *unambiguous* interpretation, which results in an inconsistency in the semantic content of the discourse (thereby making the discourse sound odd). For example, *the woman* and *the election* in (14b) unambiguously denote one of the people I met last night and the vote denoted in (14a) respectively, even though this results in an inconsistency that makes the discourse sound strange:

(14) a. I met two interesting people last night who voted for Clinton.

b. The woman abstained from voting in the election.

Hobbs *at al*'s abductive framework can't account for examples like these, because the account will prefer accommodating the definite descriptions to the binding, in order to preserve consistency. In our account, binding definite descriptions to the discourse context is essential, because the compositional semantics of the definite article will demand it. In the above exam-

ple, one would infer *Elaboration* between the constituents because of the relationship between the woman and the two people. But the coherence constraints on this relation *won't* be violated by the fact that one can't abstain and vote at the same time.

4.1 The Presupposition of a Definite Description

We assume DRSs for clauses are built monotonically and compositionally, using the bottom up procedure described in Asher (1993). Definite descriptions introduce two anaphoric conditions: one on a binding relation B, which has as one of its arguments the discourse referent introduced by the definite description; the other anaphoric condition is on the second argument of the relation B. Furthermore, in order to do justice to the intuition that given information must obtain prior to the new information, we assume that this second argument of B is asserted to exist in the pre-state of the main eventuality of the constituent. We suppose that pre-states and post-states are defined for all eventualities. For those eventualities that are changes of state, the assumption seems commonsensical. The pre-state is the state that holds just before the changes take place. For states themselves, and perhaps also for processes, the assumption may be a little more problematic. But see Asher *et al.* (1996) for a discussion. Finally, we add a uniqueness condition, whose meaning is dependent in part on the way the underspecified binding relation B is resolved. While this part of our proposal is specific to definite descriptions, the main ideas, we hope, will carry over to other presuppositional triggers.

In detail, *the N* produces the partial DRS (15), and so sentence (16) has the logical form (16′):

(15) $\lambda e \lambda Q$

x, u, B, U
$Q(x, e)$
$N(x, \mathit{pre}(e))$
$u \in \mathit{Dom}(\mathit{pre}(e))$
$B(x, u)$
$B = ?$
$u = ?$
$\mathit{unique}_N(x)$

(16) The man walked

(16′)

x, u, B, U, e, t, n
$\mathit{walk}(e, x)$
$\mathit{hold}(e, t)$
$t \prec n$
$\mathit{man}(x, \mathit{pre}(e))$
$B(x, u)$
$B = ?$
$u = ?$
$u \in \mathit{Dom}(\mathit{pre}(e))$
$\mathit{unique}_{\mathit{man}}(x)$

(16′) contains the underspecified and anaphoric elements mentioned earlier. In words, there is an object x that's a man at the time when the pre-state of the eventuality e holds, where e is an

event of x walking which holds at a time t before now; furthermore, there is some (anaphoric) relation B between x and a discourse referent u that was introduced in the prior (accessible) discourse context, where u is within the domain of the pre-state of the event e, and this man x is unique. We will shortly explicate this latter condition $\textit{unique}_N(x)$.

For the sake of brevity, we haven't defined conventional constraints which determine when a definite description is licensed. For example, we have ignored syntactic constraints such as Anti C-command. Anti C-command could be added to the DRS construction procedure, to block anaphoric bindings in sentences such as (17), thereby explaining its anomaly:

(17) ?Every man thinks that the man is a genius.

However, we remain agnostic on the validity of binding constraints such as Anti C-command, because of counterexamples like (18):

(18) The pilot who shot at it destroyed the MIG that chased him.

There appear to be, however, other highly idiosyncratic constraints on the use of definite descriptions. The examples in (19) illustrate that even if pragmatics permits one to compute a binding relation B to a unique antecedent, the use of the definite description may still be odd: compare (19a,f) with (19b,c,d,e,g):

(19)
a. John had an accident. ?The nose was broken.
b. John had an accident. His nose was broken.
c. The plane had a serious fault. The nose was damaged.
d. John had an accident. The arm he depended on most was broken.
e. John had an accident. The next day he went to see the doctor. He told him that the arm was permanently damaged.
f. John shot himself. ?The foot was injured.
g. John shot to kill. He aimed for the chest but hit the leg.

We capture the difference in acceptability in these sentences by assuming the following constraint: if the bridging relation *part-of* could have been replaced with a possessor, then there is a preference for that, so long as the definite would not have generated other discourse effects that the possessor does not. In other words, it's better to say *his nose*, *his briefcase*, *his leg* and so on, so long as the definite wouldn't have produced additional discourse effects. Intuitively, definites are more expensive to process than anaphoric pronouns and this constraint reflects this fact. Our formal analysis of definites also reflects the fact that they are more expensive to process, because to fully interpret a definite one must compute resolutions for the underspecified anaphoric elements, B and u.

This constraint predicts that (19b) is better than (19a). Furthermore, given that objects possess things only through sense extension, it also predicts that (19c) is better than (19a). In general, the above constraint will only apply when the antecedent to the definite is human, since it's most natural to talk of humans possessing things. The constraint isn't violated by (19d) either, because of the anaphoric pronoun which appears within the description, which

would ultimately make the possessive anaphor sound odd in this case; and similarly, it's not violated in (10c) because of the presence of the reflexive. This contrasts with (19f), where the possessive pronoun would be acceptable, and so the constraint predicts that (19f) is odd. (19e) is acceptable, because there appear to be additional discourse effects produced by the definite, and so the constraint isn't violated in this case. This discourse seems to be of a specific genre, typical of medical diagnosis, and the bridging relation triggered by the definite isn't simply *part-of*; rather, one also computes the bridging implicature from the definite that the arm is the item that the doctor examined. Finally, the definite is acceptable in (19g), because the possessive anaphor would not have a linguistically expressed anatecedent, and so the constraint isn't violated in this case.

With this in mind, let's return to the semantics of the condition $\textit{unique}_N(x)$. It appears that all the cases where the uniqueness condition is vacuous (e.g., (10a,b,c)) are ones where the binding relation B is *part-of*. So this is a necessary condition for $\textit{unique}_N(x)$ being resolved to $\top$. Indeed, we believe that *all* cases where the bridging relation is *part-of*, and the definite is licensed by the above constraint (it's licensed in (10a,b) because the antecedent isn't human, and it's licensed in (10c) because the human antecedent is bound within the same clause), are cases where the uniqueness condition becomes vacuous. Those cases such as (20), where one attempts to resolve the definite to a non-unique *part-of* an antecedent, are all cases where a definite which resolves to a unique *part-of* an antecedent are also odd (e.g., (19a)).

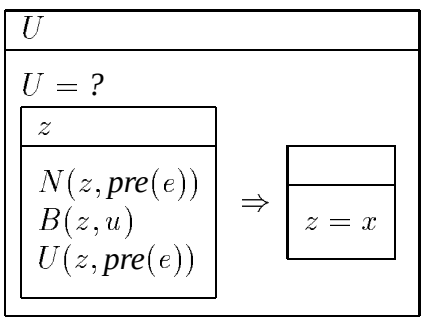
(20) John had an accident. ?The leg was broken.

Therefore, we rule out the acceptability of (20) via the above DICE constraint on the use of definites, rather than via a violation of the uniqueness constraint. Hence, we assume resolving B to *part-of* is also a sufficient condition for $\textit{unique}_N(x)$ being resolved to $\top$.

The uniqueness axioms below capture this. As is usual with the SDRT axioms of DICE, we will indicate that a constituent β has a particular condition C in it with the formula $\beta[C]$, and $\beta[C/C']$ labels the constituent that is just like β, save that condition C is replaced with C'. So in words, axiom (a) below stipulates that if the binding relation B is resolved to *part-of* in β, then the uniqueness condition is 'vacuous', in that β itself is equivalent to the DRS β with that uniqueness condition replaced with the tautology $\top$. Axiom (b) states that in any other circumstance, the uniqueness condition in β amounts to the condition given in the DRS $\mathcal{U}$ below.

- Uniqueness Axioms:
  - (a) $(\beta[B = \lambda x \lambda y \textit{part-of}(x, y)] \wedge \beta[\textit{unique}_N(x)]) \rightarrow$
    $(\beta \leftrightarrow \beta[\textit{unique}_N(x)/\top])$
  - (b) $(\beta[\textit{unique}_N(x)] \wedge \beta[B \neq \lambda x \lambda y \textit{part-of}(x, y)]) \rightarrow$
    $(\beta \leftrightarrow \beta[\textit{unique}_N(x)/\mathcal{U}])$

($\mathcal{U}$)



$\mathcal{U}$ amounts to a weak Russellian uniqueness condition, where the antecedent of the quantification is relativised to a predicate U that can be filled in by the discourse context. U is a relation between an object and a pre-state, which specifies conditions on the object that hold at that pre-state. As we've mentioned, we ensure these conditions hold in the pre-state of the eventuality described by the new information in the sentence, to capture the intuition that given information must be true *prior* to the addition of new information.

4.2 Building the Bridges

We've argued that rhetorical relations determine how presuppositions are anaphorically bound to elements in the context. To put this into a slogan: presupposition accommodation is a byproduct of SDRS update. This differs from the standard dynamic notion of presupposition satisfaction. The latter assumes a narrative sequence and threatens not to get the presuppositions right in the above non-narrative examples, such as (12).

We now define how the anaphoric binding relation B and antecedent u, which are introduced by the compositional semantics of the presupposition trigger, are resolved in terms of the function *Update* introduced in section 3.[12] There will be three rules that define this behaviour. They are not part of the DICE language; rather, they are meta-rules about how the semantic content of underspecified constituents and the function *Update* interact. The first rule captures the intuition that one binds presupposition in preference to inferring bridging implicatures; the second captures the intuition that if updating the discourse with the (underspecified) new information triggers changes to the semantic content of that new information that can act as a bridging implicature, then this information is treated as a bridging implicature. And the last rule captures the intuition that we favour bridging implicatures that maximise discourse coherence.

The first rule is given below. It states that if SDRS update with the binding relation B specified to identity is well-defined, then SDRS update must set B to identity.

- `If Possible Use Identity:`
  $(K_\beta[B =?] \wedge \downarrow \textit{Update}(K_\tau, K_\alpha, K_\beta[B/\lambda x \lambda y x = y])) \rightarrow$
  $(\textit{Update}(K_\tau, K_\alpha, K_\beta) := \textit{Update}(K_\tau, K_\alpha, K_\beta[B/\lambda x \lambda y x = y]))$

This axiom reflects the preference noted by van der Sandt, for standard anaphoric binding over other forms of presupposition satisfaction, which include bridging. However, the condition that `If Possible Use Identity` imposes on standard anaphoric binding are stronger than van der Sandt's. In van der Sandt's theory, a presupposition will bind in any context where an accessible discourse referent satisfying the same content is to be found. In contrast, `If Possible Use Identity` permits this binding only if van der Sandt's condition holds, *and* one can compute a rhetorical relation with the result. Van der Sandt's weaker condition on binding is problematic in an example such as (21):[13]

(21) a. Boggs stood calmly by as Ryan struck out the hitter with a 95-mph pitch,
b. then he stepped up to the plate and
c. he hit the pitch out of the park.

In van der Sandt's analysis *the pitch* in (21c) will bind to the 95mph pitch mentioned in (21b), because his theory fails to account for the affects of temporal constraints on narrative discourse. If one were to augment his theory with the compositional semantic account of temporal information in discourse provided in Kamp and Reyle (1993), then the definite description would no longer bind, because the result would be inconsistent. However, we have shown elsewhere (Lascarides and Asher 1991, 1993; Asher *et al.* 1996) that Kamp and Reyle's account of temporal information in discourse is deficient, because it doesn't take into account the constraints on temporal information imposed by non-narrative rhetorical relations. In contrast, our theory will detect that the binding relation B in the representation of *the pitch* in (21b) cannot be identity, because the result will violate the temporal coherence constraints on *Narration* which, by DMP on `Narration`, binds the propositions together in this discourse. Instead, of B resolving to identity, it will resolve to 'thrown-by', and the second argument of the relation will be Ryan.

Note that `If Possible Use Identity` is monotonic rather than default. Giles Fauconnier (pc.) has offered (22) as a potential counterexample to its monotonicity:

(22) A foreign president visited the White House, but the President was busy.

He claims that resolving the binding relation to identity is possible in this case, but it isn't the intended reading. We disagree. If we do identify *the President* with the president mentioned in the first sentence, then the coherence constraints required by the relation *Contrast*, which is monotonically inferred from the cue word *but*, are violated, much in the same way as they're violated in (23), if one assumes that *he* refers to the foreign president.

(23) ?A foreign president$_i$ visited the White House, but he$_i$ was busy.

In our terms, SDRS update is not well-defined when B is set to identity.

We also have the task of specifying the the parameter U in the uniqueness constraint $\mathcal{U}$, which potentially adds semantic content to the presupposition. However, we don't want to add unnecessary semantic content. This is a special case of the principle of Relevance discussed in Sperber and Wilson (1986): compute connections between things in the discourse at minimum cost. In this special case of definite descriptions, this rule amounts to the following: if $\mathcal{U}$ is satisfiable when U is resolved to a tautology $\top$, then normally U is indeed resolved to this. For the sake of brevity, we don't formalise this Relevance Principle here, since it would involve an extensive discussion of reasoning with cognitive states. We simply call it `Minimum Cost` and assume it has the desired effect of resolving U to $\top$ by default. Note that when B is resolved to identity, $\mathcal{U}$ is guaranteed to be true when U is resolved to $\top$. So by `Minimum Cost`, U is $\top$ when B is identity.

However, as we've seen, specifying the binding relation as identity doesn't always yield a complete SDRS constituent; some of the underspecified parameters don't get filled in. In this case, we allow the discourse context to guide us to a suitable specification for the binding relation, and for U. All the following rules suppose that $\neg \downarrow (\mathit{Update}(\tau, \alpha, \beta[B/\lambda x \lambda y x = y]))$ holds.

In general, there are many ways the underspecified parameters could be made precise and so account for the presuppositions; some of these may be more plausible than others. It is here that we see an important role for world knowledge in presupposition calculation (cf. Beaver

(1994)). We suppose that the KB of the interpreter will specify certain plausible ways of filling in the underspecified parameters in the presupposed material. So we will introduce a conditional operator: $A >_\diamond B$ should be read as "If A, then it's plausible to assume B". This specifies a weaker connection than $>$; it stipulates what is plausibly the case, rather than what is normally the case. In essence our rules below will restrict the strength of the bridging inference as follows: the bridge must be built from $>_\diamond$ consequences of the semantic content of the constituents. That is, a bridge must be plausible.

An axiomatisation of $>_\diamond$ would involve extensive discussion of commonsense reasoning, and so we gloss over it here. However, we do suggest a truth definition of $A >_\diamond B$. This is defined in terms of a function $*_{>\diamond}(w, p)$, from worlds and propositions to sets of propositions; that is, it's a function from $W \times \mathcal{P}(W)$ to $\mathcal{P}(\mathcal{P}(W))$. The truth definition of $A >_\diamond B$ is then as follows:

- `Truth Definition of` $A >_\diamond B$`:`
  $M \models_w A >_\diamond B$ if and only if there is a set of worlds $W' \in *_{>\diamond}(w, [\![A]\!])$ such that $W' \subseteq [\![B]\!]$.

We assume that $*_{>\diamond}$ is defined so that $\cup *_{>\diamond}(w, p) \subseteq *(w, p)$. This guarantees that all normal things are plausible.[14]

We can now use this conditional to represent constraints on the specification of the underspecified parameters. `Bridges are Plausible` specifies that the binding relation and accompanying uniqueness constraint must resolve to something plausible:

- `Bridges are Plausible:`
  $(\beta[B = \phi] \wedge \beta[U = \psi] \wedge \langle \tau, \alpha, \beta \rangle \wedge R(\alpha, \beta)) \rightarrow$
  $((\mu(K_\tau)(\tau) \wedge \mu(K_\beta)(\beta) \wedge R(\alpha, \beta)) >_\diamond (\phi \wedge \psi))$

In words, if B and U are resolved to ϕ and ψ in β, where β is attached to the constituent α in τ with a rhetorical relation R, then the semantic content of this discourse context τ and constituent β (i.e., $\mu(K_\tau)(\tau)$ and $\mu(K_\beta)(\beta)$) and the rhetorical connection R between α and β must make it plausible to assume both ϕ and ψ. We'll see in section 5.3 that this rule will prove important when distinguishing (7a,b) from (7a,b′) (it's not plausible to assume fleas were the mode of transport):

(7) a. I arrived here at 3pm.
b. The camel is outside and needs water.
b′. ?The fleas are outside and need water.

A further important rule governing bridging inferences is `Discourse Structure Determines Presupposition Satisfaction` (or `DS determines PS`). This rule captures the intuition that when the rhetorical relation used to connect the constituents gives us a particular way of resolving the presupposition, we do it that way. More formally, let $\mu(K_\beta)(\beta) \rightsquigarrow_* \mu(K_\phi)(\phi)$ mean: K_ϕ is a DRS which represents one plausible way of resolving the underspecification in K_β. Then `DS Determines PS` is given below:

- `Discourse Structure Determines presuppositions satisfaction (DS Determines PS):`
  Suppose: (a) $\mu(K_\tau)(\tau) \wedge \mu(K_\beta)(\beta) \wedge \langle \tau, \alpha, \beta \rangle \ \mid\!\approx\ R(\alpha, \beta)$
  (b) $\mid\!\approx\ \mu(K_\beta)(\beta) \rightsquigarrow_* \mu(K_\phi)(\phi)$; and
  (c) $\mid\!\approx\ (R(\alpha, \beta) \wedge \mu(K_\tau)(\tau)) > \mu(K_\phi)(\phi)$
  Then *Update*$(K_\tau, K_\alpha, K_\beta)$:= *Update*$(K_\tau, K_\alpha, K_\phi)$

In words, if we can infer the rhetorical connection R from the discourse context τ and the *underspecified* constituent β, and this relation R allows us to infer a particular resolution ϕ of the underspecified elements in β, then ϕ is incorporated into the SDRS update.

`DS Determines PS` models the information flow from discourse structure to the content of definite descriptions. This rule will be used to model (11), for example.

(11) a. John took engine E1 from Avon to Dansville.
b. He picked up the boxcar and took it to Broxburn.

We can use DICE to infer (11a,b) is narrative even before determining the underspecified elements B and u in (11b); we then use the narrative to infer the boxcar is in Dansville, and this added content suffices to produce a plausible way of resolving these underspecified elements. `DS Determines PS` ensures we resolve them this way. The details of this analysis are given in the next section.

`DS Determines PS` deals with the case when the coherence constraints imposed by the rhetorical relation that's inferrable from the *underspecified* constituent β produces a plausible bridging inference. But the underspecified constituent β doesn't always contain sufficient information to determine the rhetorical relation; hence it may not be enough to determine the bridging inference. To deal with such cases, we state a rule which captures the intuition that people interpret text so as to maximise discourse coherence. It is a more restricted version of the Interpretation Constraint in DICE that was introduced in Lascarides *et al* (1996) for modelling word sense disambiguation, and this more restricted rule suffices for our purposes.

As background to this rule, we assume that rhetorical relations between constituents may be partially ordered with respect to the semantic content of the context. In other words, depending on the context τ, and the attachment site α that is part of τ, *Explanation* $>_{\tau,\alpha}$ *Background* may hold when attaching β to α; indicating that we would prefer to interpret β as an explanation for α, rather than background information—although both alternatives are coherent, one is better than the other. The following rule then captures the following: resolve the underspecified elements to as to maximise discourse coherence:

- `Maximise Discourse Coherence:`
  If (a) $\mu(K_\beta)(\beta) \rightsquigarrow_* \mu(K_{\beta_1})(\beta_1)$; and
  (b) $\langle \tau, \alpha, \beta_1 \rangle \wedge \mu(K_\tau)(\tau) \wedge \mu(K_\beta)(\beta) \ \mid\!\approx\ R_1(\alpha, \beta_1)$; and
  (c) R_1 is the $>_{\tau,\alpha}$ maximal rhetorical relation of attachment
  Then *Update*$(K_\tau, K_\alpha, K_\beta)$:= *Update*$(K_\tau, K_\alpha, K_{\beta_1})$.

`Maximise Discourse Coherence` will be used in the analysis of (1) and (7) in section 5.2.

This account of presupposition amounts to one where one always attempts to bind the given information to a unique antecedent. Lewis' (1979) view of accommodation (i.e., the simple addition of the presupposition to the context) is ruled out. Hence, presupposition failure amounts to: one cannot find a unique antecedent to which the given information binds. Two comments on this notion of failure are in order. First, one might think that this theory would predict that all true (i.e., not namelike) definite descriptions that appear in the absence of a prior clause amount to presupposition failure. However, this is not the case. We assume that some types of descriptions are able to specify, by their own *compositional* semantic content, the underspecified elements B, u and U in the presuppositional information. In fact, we assume that if it's possible to resolve B, u and U via the compositional semantics, then this is the way it is done. For example, the compositional semantics of descriptions like *King of France* has this effect, in that a description of the form *X of Y* amounts to resolving B to *of* and u to the discourse referent introduced by complement of the PP (i.e, Y). *The X of Y* will thus be acceptable in the absence of any discourse context, so long as the uniqueness condition $\mathcal{U}$ is verifiable with these resolutions of B and u. This is the basis of the distinction between (24a,b), which are acceptable, and (24c), which is odd:

(24) a. The King of France lived at Versailles.

b. The prime minister of some country will visit tomorrow.

c. ?The man lived at Versailles.

d. ?The voter of some constituency will visit tomorrow.

(24d) is odd because the uniqueness constraint $\mathcal{U}$ isn't verifiable when the bridging relation is *of* and u is the constituency. That is, we cannot compute a plausible value for U in this case.

Second, we also assume there are certain cases, in particular definite descriptions at the start of a novel, where the interpreter is happy to resolve the underspecified elements in the given information in the light of *subsequent* information, that appears later in the narrative. We leave as an open question the constraints that model when underspecified elements may remain unresolved for a while. Presumably, psychological experiments would provide evidence of exactly when interpreters can store underspecified information for some period of time without adverse effects on parsing.

5 Applications to Examples

5.1 Binding

Our analysis of presuppositions accounts for their projection in sentences containing negation and modality. It's complementary to van der Sandt's (1992) account in those cases where an anaphoric link can be made; i.e., where the binding relation B is specified to be identity. To see this, consider an example taken from Bos *et al.* (1995):

(25) If I invite a celebrity, then the celebrity doesn't come.

First, we must build the DRSs α and β of the two sentences *I invite a celebrity* and *The celebrity doesn't come* respectively:[15]

(α)

i, n, e_1, t_1, x
celebrity(x), *invite*(e_1, i, x) *hold*(e_1, t_1) $t_1 \subseteq n$

(β)

$n, B, u, y,$	
$\neg$	e_2, t_2 *come*(e_2, y), *hold*(e_2, t_2), $t_2 \subseteq n$ *celebrity*(y) $B =?$ $u =?$ $B(y, u)$ $u \in$ *Dom*(*Pre*(e_2)) *unique*$_{\textit{celebrity}}(y)$

According to `If Possible Use Identity`, if we can set the binding relation B to identity, we must. In this case we can. First, there is an available discourse referent in α that can be identified with u—namely x—such that if B is identity then we identify y, who is a celebrity, with the celebrity x. As we mentioned before, resolving B to identity also entails that *unique*$_N(x)$ is true when U is $\top$. So, for `If Possible Use Identity` to apply, we need only check that the result of assuming this binding between y and x produces a coherent discourse (i.e., the discourse update is well-defined). In other words, we must be able to compute a rhetorical relation between α and β' given below:

(β')

n	
$\neg$	e_2, t_2 *come*(e_2, x) $x \in Dom(Pre(e_2))$ *hold*(e_2, t_2) $t_2 \subseteq n$

Assuming that there is a rule in DICE which encodes the fact that the cue words *if. . .then* is a clue that the rhetorical relation is *Condition*, we infer by DMP on this rule that the constituents are attached together with *Condition* (so α enables β').[16] Since this update is well-defined, `If Possible Use Identity` applies, and therefore by modus ponens, the celebrity mentioned in β is the same one that's mentioned in α; i.e., *the celebrity* refers to whichever celebrities I invite. Because of this binding, (25) *doesn't* entail there's a unique celebrity in the model, as required.

This analysis is very similar to van der Sandt's (1992) and Bos *et al.*'s (1995). The only difference is that the final representation of the sentence is more complex in that it features

rhetorical relations, which give us information about the semantic relationship between the two clauses (e.g., the contingent relations between the events described in α and β).

More generally, this richer notion of discourse structure allows us to treat the *Projection Problem* differently in certain respects from DRT. In DRT, accessible discourse referents are defined via the embedding relations among the DRSs. In SDRT, however, the available discourse referents are limited to only those in the current DRS (i.e., the representation of the clause currently being processed), and the DRSs which are related to it by a subordinating relation (i.e., *Explanation*, *Elaboration*, and *Topic*). Typically, this is a smaller set. Furthermore, one could use the semantics of the rhetorical relations to affect presupposition projection. For example, consider the conditionals (26) and (27), taken from Beaver (1994):

(26) If David wrote the article, then the knowledge that no good logician was involved will confound the editors.

(27) If David wrote the article, then the knowledge that David is a computer program running on a PC will confound the editors.

The presupposition of (26) is (28), whereas for (27) it's (29):

(28) If David wrote the article, then no good logician was involved.

(29) David is a computer program running on a PC.

The presuppositional item in both (26) and (27) is the factive nominal knowledge, and this is generated in the consequent of the conditional. The treatment of the Projection Problem via accommodation in DRT tells us that we have to add the presupposition to the context, but this doesn't explain the difference between these two examples, and the various preference strategies based on structure alone (local vs. global accommodation) fail here.

The difference in the presuppositions must come from the *contents* of the presupposed material, and the way these interact with the background context. Beaver suggests that the beliefs that interpreters attribute to the speakers form the basis of the difference. An interpreter will not assume that the speaker believes that the truth of David being a computer program running on a PC is dependent on his writing the article; i.e., on the semantics of the antecedent. However, they will assume that the truth of no decent logician being involved with the article is somehow dependent on David's being an author of it. Discourse structure as construed in SDRT can help model these differences, because inferences about what to believe can be set up as a general rule about conditional contexts, via the semantics of the rhetorical relation *Condition*. We leave as an open question whether other rules will tell us how to deal with other logical connectives.

5.2 Bridging Through Discourse Attachment

Now consider a case where the presupposition is accommodated through discourse attachment:

(11) a. John took engine E1 from Avon to Dansville.
b. He picked up the boxcar
c. and took it to Broxburn.

The DRSs representing (11a) and (11b) are α and β respectively:

(α)

$j, E1, a, d, e_1, t_1, n$
John(j)
engine-E1$(E1)$
Avon(a)
Dansville(d)
take$(e_1, j, E1)$
from(e_1, a)
to(e_1, d)
hold(e_1, t_1)
$t_1 \prec n$

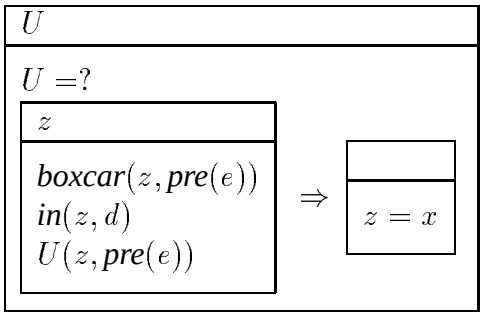
(β)

n, B, u, y, e_2, t_2, n
pick-up(e_2, j, y)
hold(e_2, t_2)
$t_2 \prec n$
$B =?$
$u =?$
$B(y, u)$
$u \in$ *Dom*(*Pre*(e_2))
boxcar(y)
unique$_{\textit{boxcar}}(y)$

Note that we have assumed that the pronoun *he* in β refers to John. We can assume this because α is the only open attachment site in the preceding SDRS, and so John is the only possible antecedent, regardless of the rhetorical relation that's used to connect α and β.

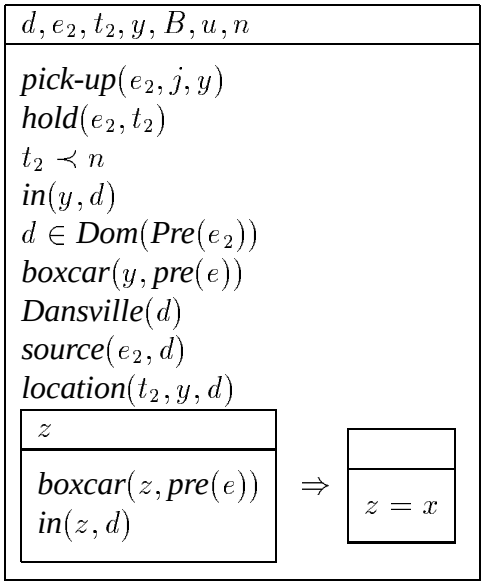
In this example, resolving B to identity leaves the update undefined, because given that there is no boxcar in the DRS α, we cannot find an acceptable antecedent to u. According to `DS Determines PS`, however, we should check whether the rhetorical structure of the context will help us in finding a suitable bridge; i.e., a specification of B and antecedent for u. In attaching β to α with a rhetorical relation, via the rules in DICE, the antecedent to `Narration` is verified, since both e_α and e_β are events. Moreover, the consequent of *Narration* is consistent with the reader's KB. So by DMP on `Narration`, *Narration*(α, β) is inferred.

Now, given that *Narration*(α, β) has been inferred, we can achieve further inferences about the semantic content of the constituents. First, by Modus Ponens and the `Temporal Consequent of Narration`, e_α occurs before e_β; that is, the taking of the engine from Avon to Dansville occurs before a boxcar is picked up. Furthermore, as we showed in section 3, by the semantics of the phrases *take to* and *pick up* and the `Spatial Consequence of Narration`, one infers that the source of the picking up event is in Dansville and the object that is picked up is therefore also in Dansville. Hence, the boxcar is in Dansville. Thus, the coherence constraints on *Narration* allows us to infer a particular way of satisfying the presuppositions—viz. specifying the binding relation B as *in* and antecedent for u as d or Dansville (hence the boxcar y is in Dansville just before it's picked up). We must also resolve the parameter U in the uniqueness condition given below (this is the DRS representing *unique*$_{\textit{boxcar}}(y)$):



As we argued earlier, there appears to be a default specification of `Minimum Cost`, which amounts to: don't infer unnecessary semantic content in the bridging inference. So, if the uniqueness constraint is consistent when U is resolved to $\top$, it should indeed be resolved to $\top$. In this case, resolving U to $\top$ produces a satisfiable DRS: it is consistent that there is one and only one boxcar in Dansville in the pre-state of the picking up event. Thus `DS Determines PS` allows us to infer the following revision of β, and this gets attached to α with *Narration*:

(β_1)



Note that our final result β_1 includes added content to the given information, that was conventionally triggered by *the boxcar*. This added content was inferred in order to meet constraints on discourse coherence. It amounts to: the boxcar is located in Dansville at the time when it's picked up, and moreover, this was the only boxcar in Dansville at this time.

Van der Sandt (1992) and Bos *et al*'s (1995) fail to account for the fact that the boxcar is in Dansville, and it's the only boxcar in Dansville. Poesio's analysis fails to model the different spatial constraints imposed by different rhetorical relations, thereby failing to predict that the boxcar in (the *Parallel* text) (11a,b′) is in Avon:

(11) a. John took the engine E1 from Avon to Dansville.

b′. He also took the boxcar.

In contrast, our analysis captures the intuitive interpretation of (11a,b′). Briefly, as in the previous example, the attempt to specify the binding relation B to identity fails. The similarity

in syntactic structure and the cue word *also* are clues in DICE that the discourse relation between (11a) and (11b) is *Parallel*. This doesn't have a spatial constraint like that represented in `Spatial Consequence of Narration`. Rather, the spatial constraints are computed on the basis of the way the different parts of the DRSs related in the parallel relation are mapped onto each other. This mapping is an essential feature of the coherence constraints on *Parallel*, since this relation requires a partial structural isomorphism between the constituents (Asher, 1993). For the sake of brevity, we omit the details of constructing this mapping here, but informally, the taking event in (11b′) is matched with that in (11a). The consequence is that, by the spatial constraints on *Parallel*, their sources and goals are taken to be the same, unless there is information to the contrary. This adds semantic content to the DRS representing (11b′); the source of the taking event in (11b′) is Avon. So by lexical semantics, the boxcar is in Avon at this source. One adds this to the representation of the given information via `DS Determines PS` as before. And so one obtains an interpretation where the boxcar that's referred to in (11b′) is in Avon rather than Dansville, and at the time of the event referred to in (11b′), it is the only boxcar in Avon, as required by the uniqueness condition.

5.3 Bridging Before Discourse Attachment

We have looked at several cases where inferring a rhetorical relation helps specify how the presuppositions of a definite description are satisfied. Let's now briefly look at the information flow in the other direction. The rule `Maximise Discourse Coherence` specified in section 3 enables us to infer added semantic content to given information so as obtain discourse coherence, where it wouldn't be coherent otherwise.

In example (1), we fail to get a well-defined update if we specify the binding relation to identity. Furthermore, in contrast to texts like (11a,b), there isn't enough information in the *underspecified* constituent β representing the second sentence, for us to infer a particular rhetorical relation between it and the first sentence.

(1) I met two interesting people last night at a party.
The woman was a member of Clinton's Cabinet.

This is because only `Background` in DICE applies, and so the only candidate relation is *Background*. But constituents related by *Background* must have a common topic. We can compute this using the technique discussed in Grover *et al.* (1994). That is, we generalise over the predicates and arguments in the propositions. Since we haven't resolved B and u, the woman is unconnected with the two people. And so computing a common topic in this way isn't possible, because the result is too general; something like *things that were true yesterday*.[17] Hence *Background* can't be inferred between α and the underspecified β. Neither can any other relation. Hence `DS Determines PS` won't apply.

Instead, we must use `Maximise Discourse Coherence`. That is, we must investigate which resolution of β produces the best discourse, and resolve β to that. So, let's entertain some possible resolutions of β and investigate the consequences on discourse coherence. Suppose that β_2 is a resolution of β where B and u are defined so that *the woman* y is *separate* from the two people mentioned in the first sentence. Then this produces just as bad a discourse as that between α and β itself, for the same reasons. On the other hand, suppose that β_1 is the resolution of β where the woman y in the DRS β is one of the two people I met last night.

In other words, the binding relation B in β_2 resolves to *member-of*, and u resolves to the discourse referent denoting the two people I met in α. Then the uniqueness constraint given by $\mathcal{U}$ in section 4.1 is satisfiable when the underspecified parameter U is resolved to $\top$: If U is $\top$, then $\mathcal{U}$ stipulates that if z is one of the people I met las night, and moreover z is a woman, then z is identified with the woman mentioned in β. This is consistent with the knowledge base. And thereby by `Minimum Cost`, we infer that U is $\top$. Moreover, the rules in DICE given in Asher and Lascarides (1995) allow us to compute *Elaboration* between these constituents α and β_2. This comes with different coherence constraints from *Background*: the topic is α. The discourse coherence is therefore much improved relative to the representations β and β_1, which failed to produce a well-defined update. So, the antecedent to `Maximise Discourse Coherence` applies with these resolutions β_1 and β_2, and the discourse context α is updated via *Elaboration* with β_2.

As before, we gain further information through applying `Maximise Discourse Coherence`, which allowed us to infer the resolution β_2 of β: we now know that the woman is one of the two people I met last night, and only one of the people I met last night was a woman. So the other one must have been a man.

Our analysis of (7a,b) also uses the principle `Maximise Discourse Coherence`.

(7) a. I arrived at 3pm today.

b. The camel is outside and needs water.

b′. ?A camel is outside and needs water.

b″. The fleas are outside.

Again, we can't fill in the binding relation with identity. Moreover, just as in (1), we can't compute a coherent discourse with the underspecified constituent. The antecedent to `Background` is verified. *Background* requires a distinct common topic. However, the consequences of this are at best poor coherence. Generalising the two propositions produces something very general, like *things that are the case today*. The point to stress is that the *Background* relation doesn't tell us on its own that the camel had anything to do with the arrival; it is because of this that the topic constructed is so poor. This is in contrast from the analysis of (11a,b), where the discourse relation *Narration* which was inferred added content to the definite description through its spatial constraint. Here, *Background* doesn't constrain the interpretation of camel. And since we currently lack any connection between the camel and the arrival in the KB, it yields a poor topic. Therefore, just as in (1), we must entertain various resolutions of the underspecified parameter in β and see which option maximises discourse coherence. Suppose B, u and U are resolved so that the camel had some role in the arrival. By the constraint `Bridges are Plausible` given in section 4.1, this must be a plausible role. The only one is that the camel is the mode of transport by which I arrived. Therefore, $B(y, u)$ and $U(y, pre(e))$ in β (where y is the camel) resolve to: y is the mode of transport which i used in the arrival event e. This content enables us to infer a new rhetorical relation, with improved discourse coherence. We can infer that the camel being outside was caused by my arrival thanks to the spatial information in the compositional semantics of the change of location phrase *arrive here*, and so the rhetorical relation is *Result*. So `Maximise Discourse Coherence` is used to infer this new content to the definite description *the camel*, together with the *Result* relation between the constituents.

(7a,b′) is odd because the indefinite description doesn't require one to resolve anaphoric relations and underspecified elements. So `Maximise Discourse Coherence` doesn't apply (since clause (a) doesn't hold), and one can't use it to infer the camel is the mode of transport and the rhetorical relation is *Result*. (7a,b″) is odd because one cannot infer that the fleas are the mode of transport. This is implausible, and so it's ruled out by `Bridges are Plausible`. Indeed, there is no plausible resolution of B and u that produces a coherent discourse, and so the SDRS can't be updated.

Now consider the text (12):

(12) a. John moved from Brixton to St. John's Wood.

b. The rent was less expensive.

Let the sentences (12a,b) be represented by the DRSs α and β respectively. Once again attempting to resolve the binding relation B to identity fails. But *rent* is a functional noun, and it in and of itself suggests a binding relation B. B should be the relation *of* and the other term of the binding relation should be some sort of object that can have rents. Nevertheless, there are still no places that are mentioned in (12a) that have rents. So we attempt attachment between α and β.[18]

As in the previous examples, one cannot compute a rhetorical relation between α and the (underspecified) β, because we require further content: we need to know more about the connection between the rent mentioned in β and the content of α.

There are at least two possible resolutions of u in β. The first, β_1, is such that the constituent means: *the rent of the place that John moved to, which is in St. John's Wood, is less expensive than the rent he paid in Brixton*. The second, β_2, is such that the constituent means: *the rent he paid in Brixton is less expensive than the rent of the place he moved to, which is St. John's Wood*. β_1 together with the content of α yield *Explanation*(α, β_1) in DICE. They also yield *Background*(α, β_1), because *Background* is compatible with *Explanation*, and β_1 describes a state (i.e., the rent in St. John's would being less expensive). Moreover, in contrast to α and β, we can compute a good topic for α and β_1, since we now know the rent is connected to St. John's Wood. In contrast, β_2 and the content of α yields only *Background*(α, β_2), but it cannot support *Explanation* (since moving to a more expensive house doesn't explain why one moved; at least, not on its own). Intuitively, one prefers an interpretation of a discourse that offers explanations of intentional behaviour that's described in the text—such as moving house—to an interpretation of the discourse where such behaviour is left unexplained. In essence, interpreters don't like miracles, or unexplained changes. We can model this via the partial order of rhetorical relations: *Explanation* $>_{\tau,\alpha}$ *Background*, in this case. Therefore, the antecedent to the monotonic rule `Maximise Discourse Coherence` is verified with β resolving to β_1, and so one updates β to β_1. In other words, one infers the rent referred to in (12b) is the rent that John pays in the place he moved to, which is in St. John's Wood.

This consequent of `Maximise Discourse Coherence` is incompatible with the default world knowledge that rents in Brixton are typically less expensive than those in St. John's Wood. However, since `Maximise Discourse Coherence` is a monotonic rule, it overrides this default world knowledge, which is as required, given the evidence in Matsui's experiments. In essence, `Maximise Discourse Coherence` guarantees that maintaining discourse coherence takes priority over default world knowledge; a principle of discourse interpretation for which we have argued elsewhere in modelling word sense disambiguation

(Lascarides and Copestake 1995, Lascarides *et al* 1996).

6 Conclusion

Bridging inferences involve a complex interaction between lexical and compositional semantics, world knowledge and discourse structure. We have shown in this paper that the coherence constraints imposed by different rhetorical relations have an effect on the resolution of presuppositions in discourse, which cannot be accounted for purely in terms of constraints on tracking focus or on domain knowledge.

We have modelled this effect in SDRT, a theory of discourse structure with the distinguishing feature that rhetorical connections can trigger a change to the semantic content of the propositions introduced in the text. Bridging inferences are a byproduct of computing in SDRT how the current sentence connects to the previous ones in the discourse. Our account fully integrates compositional and lexical semantics and discourse structure to compute presuppositions. We use a well-defined logic which combines these various knowledge sources to compute bridging inferences, paying particular attention to when these knowledge resources conflict. We demonstrated that by integrating compositional semantics and pragmatic reasoning in this way, we provide a more refined account of bridging inferences than either compositional semantic accounts or AI accounts that exploit background knowledge in discourse interpretation can achieve on their own.

This paper has concentrated on presuppositions triggered by definite descriptions. In future work, we hope to show that the strategy adopted here for resolving presuppositions of definite descriptions applies to presupposition triggers in general. If the strategy of underspecification and binding proposed here generalises to all other presupposition triggers, then this may form the foundations of a theory that rejects Lewis' notion of accommodation all together. However, if not, then we may have the basis for distinguishing between the various types of presupposition triggers, which traditionally have always been recognised as a heterogeneous set.

Acknowledgements

This paper was presented at the International Workshop on Underspecification which was held at Berlin in 1996, and also at University of Texas at Austin and at CUNY. We would like to thank the people that attended these talks for their feedback. We would also like to thank David Beaver, Janet Hitzeman, Ali Knott and Frank Veltman.

Notes

[1] (2) is taken from Clark (1977).

[2] For detailed discussion of these constraints, see Asher 1993, Lascarides and Asher 1993, Asher and Lascarides 1996, Asher *et al.* 1996.

[3]For the purposes of this paper, we ignore the use of definite descriptions in generics such as *The back of the leg has large muscles known as leg biceps* (taken from the BNC corpus).

[4]In fact, this is a slightly modified version of the example in Poesio (1994), in that we have put it in the past tense, rather than having a sequence of instructions. We modify the example here because we want to ignore speech acts in this paper.

[5]Perhaps more seriously, these accounts also lack a general inference procedure for computing intentional structures from commonsense plans, and hence the ultimate discourse segmentation, which is assumed to be isomorphic to this intentional structure, is inferred by theory bound intuitions. For a detailed critique of this, see Asher and Lascarides (1996).

[6] For our purposes, one could replace the DRSs with labelled dynamic contexts, i.e., sets of world embedding function pairs, together with certain information about which variables are available for anaphoric reference.

[7]Note that these rules have evolved since the theory was presented in Lascarides and Asher (1993), for reasons stipulated in Asher *et al.* (1996). But the spirit of the approach remains the same.

[8]In some narratives, Asher *et al.* (1996) observe a spatiotemporal interpretation that implies that the actors are in motion and so there is a "transition" of the location of x at the end of e_α by the time of the onset of e_β. Our hypothesis is that these transitions are due to the presence of frame adverbials. Asher *et al.* are currently verifying this hypothesis at least for French with an extensive corpus-based search for counterexamples.

[9]Note that if R's coherence constraints can't be inferred, then the logic underlying DICE guarantees that R won't be nonmonotonically inferred.

[10]Note that this definition holds only for DRSs without conditions of the form $x =?$; we gloss over the complications in the DRT update procedure that occur when we countenance such conditions. They aren't relevant for our purposes here, since we resolve conditions such as these in different ways from those in DRT.

[11]See Lascarides and Asher (1993).

[12]We assume that *Update* is a partial function, and so ignore problems of ambiguity.

[13]Thanks to Geoff Nunberg for this example.

[14]However, note that the semantics of $>_\diamond$ is different from that of $>$, because we want it to be possible that $A >_\diamond B$ and $A >_\diamond \neg B$ without it following that $A >_\diamond \bot$.

[15]In line with the analysis of indexicals in DRT, the indexicals such as *now* (n) and *I* (i) are promoted to the top-level box in the DRS-structure.

[16]The antecedent to `Narration` doesn't hold because e_β is not identified with the event e_2, since β is a negated sentence.

[17]We don't formalise here the conditions under which a topic is poor. For such a formalisation, see Lascarides *et al.* (1996).

[18]For the sake of simplicity, we ignore the comparative nature of *less*, and gloss over the way one computes from the discourse context the set over which the comparison (or rental cost) is measured.

7 References

Asher, N. (1993) *Reference To Abstract Objects in Discourse*, Kluwer Academic Publishers.

Asher, N. (1996) Mathematical Treatments of Discourse Contexts, in P. Dekker and M. Stokhof (eds.) *Proceedings of the Tenth Amsterdam Colloquium on Formal Semantics*, pp. 21–40, ILLC publications, University of Amsterdam.

Asher, N. (forthcoming) The Logical Foundations of Discourse Interpretation, in J. M. Larrazabal (ed.) *Logic Colloquium 1996*, University of Chicago Press.

Asher, N., M. Aurnague, M. Bras, L. Vieu (1996) De l'espace-temps dans l'analyse du discours", *Semiotique : Numero Special "Theories sémantiques et modalisation*, **9**.

Asher, N. and A. Lascarides (1994) Intentions and Information in Discourse, in *Proceedings of the 32nd Annual Meeting of the Association of Computational Linguistics*, Las Cruces, USA, June 1994, pp34–41.

Asher, N. and A. Lascarides (1995) Lexical Disambiguation in a Discourse Context, *Journal of Semantics*, **12** number 1, pp69–108, Oxford University Press.

Asher, N. and A. Lascarides (1996) Questions in Dialogue, Research report number HCRC/RP–81, Human Communication Research Centre, University of Edinburgh; also submitted to *Linguistics and Philosophy*.

Asher, N. and P. Sablayrolles (1995) A Typology and Discourse Semantics for Motion Verbs and Spatial PPs in French, *Journal of Semantics*, **12**.

Beaver, D. (1994), An Infinite Number of Monkeys, Technical Report, ILLC, Universiteit van Amsterdam.

Bos, J., A-M. Mineur and P. Buitelaar (1995) Bridging as coercive accommodation, technical report number 52, Department of Computational Linguistics, Universität Saarbruücken.

Briscoe, E. J., A. Copestake and B. Boguraev (1990) Enjoy the Paper: Lexical Semantics via Lexicology, *13th International Conference on Computational Linguistics (*COLING-*90)*, Helsinki, pp. 42–47.

Charniak, E. (1983) Passing Markers: A Theory of Contextual Influence in Language Comprehension, *Cognitive Science*, **7**, pp. 171–190.

Chierchia, G. (1995) *Dynamics of Meaning : anaphora, presupposition, and the theory of grammar*, University of Chicago Press.

Clark, H. (1977) Bridging, in P. N. Johnson-Laird, P. C. Wason (eds.) *Thinking: Readings in Cognitive Science*, pp. 411–420, Cambridge University Press.

Fernando, T. (1994) What is a DRS?, in R. Cooper and J. Groenendijk (eds.) *Integrating Semantic Theories II, Dyana Deliverable R2.1.B*, 1994, also appeared in the Proceedings of

the First International Workshop on Computational Semantics, Tilburg 1994.

von Fintel, K. (1994) *Restrictions on Quantifier Domains*, PhD Thesis, University of Massachussetts, Amherst Mass.

Grosz, B. and C. Sidner (1986) Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, **12**, 175–204.

Grover, C., C. Brew, M. Moens, S. Manandhar (1994) Priority union and generalisation in discourse grammar, *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94)*, Las Cruces, USA, pp 17–24.

Heim, I. (1992) Presupposition Projection and the Semantics of Attitude Verbs, *Journal of Semantics*, **9** (3).

Hobbs, J. (1979) Coherence and Coreference, *Cognitive Science*, **3**, pp.67–90.

Hobbs, J. (1985) On the Coherence and Structure of Discourse. Report No. CSLI-85-37, Center for the Study of Language and Information, October, 1985.

Kamp, H. and U. Reyle (1993) *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers.

Karttunen, L. (1974) Presupposition and Linguistic Context, *Theoretical Linguistics* **1**, pp. 181–194.

Lascarides, A. and N. Asher (1993) Temporal Interpretation, Discourse Relations and Commonsense Entailment, in *Linguistics and Philosophy*, **16**, pp.437–493.

Lascarides, A. and A. Copestake (1995) The Pragmatics of Word Meaning, in T. Galloway and M. Simons (eds.) *Salt V Proceedings*, Cornell University Press, Ithaca, NY.

Lascarides, A., A. Copestake and E. J. Briscoe (1996) Ambiguity and Coherence, *Journal of Semantics*, **13** number 1, pp41–65, Oxford University Press.

Lascarides, A. and J. Oberlander (1993) Temporal Connectives in a Discourse Context, in *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL93)*, pp260–268, Utrecht, The Netherlands.

Lewis, D. (1979) Scorekeeping in a Language Game, *Journal of Philosophical Logic*, **8**, pp. 339–59.

Matsui, T. (1995) *Bridging and Relevance* University of London Ph.D. Thesis

Poesio, M. (1993) A Situation-Theoretic Formalization of Definite Description Interpretation in Plan Elaboration Dialogues, In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, (eds.), *Situation Theory and its Applications*, vol. 3, 1993, p.339–374.

Poesio, M. (1994) Definite Descriptions, Focus Shift and a Theory of Discourse Interpretation, in *Proceedings of the Conference in Focus on Natural Language*, June 1994.

Polanyi, L. (1985) A Theory of Discourse Structure and Discourse Coherence. In *Papers from the General Session at the Twenty-First Regional Meeting of the Chicago Linguistics Society*, pp. 25–27.

Pustejovsky, J. (1991) The Generative Lexicon, *Computational Linguistics*, **17**(4), pp. 409–

441.

Pustejovsky, J. (1995) *The Generative Lexicon*, MIT Press, Cambridge, Mass.

Russell, B. (1905) On Denoting, *Mind*, **14**, pp. 479–493.

van der Sandt, R. (1992) Presupposition Projection as Anaphora Resolution, *Journal of Semantics*, **19**(4).

Sanfilippo, A. (1992) Grammatical Relations in Unification Categorial Grammar, *Lingua & Stile*, Fall issue.

Sperber, D. and Wilson, D. (1986) *Relevance*, Blackwells.

Stalnaker, R. (1978) Assertion, in Cole, Peter (ed.) *Syntax and Semantics volume 9: Pragmatics*, Academic Press.

Thompson, S. and W. Mann (1987) Rhetorical Structure Theory: A Framework for the Analysis of Texts. In *IPRA Papers in Pragmatics*, **1**, 79–105.

Verspoor, C. (1996) Lexical Limits on the Influence of Context, in G. W. Cottrell (ed.) *Proceedings of the 18th Annual Conference of the Cognitive Science Society* pp 116–120, Lawrence Erlbaum Associates.

Nicholas Asher
Department of Philosophy,
University of Texas at Austin,
Austin, Texas 78712,
USA
nasher@bertie.la.utexas.edu

Alex Lascarides
Centre for Cognitive Science and HCRC,
University of Edinburgh,
2, Buccleuch Place,
Edinburgh, EH8 9LW,
Scotland, UK
alex@cogsci.ed.ac.uk