Agreement, Disputes and Commitments in Dialogue

Alex Lascarides
School of Informatics,
University of Edinburgh

Nicholas Asher
IRIT
Université Paul Sabatier, Toulouse

Abstract

This paper provides a logically precise analysis of agreement and disputes in dialogue. The semantics distinguishes among the public commitments of each dialogue agent, including commitments to relational speech acts or *rhetorical relations* (e.g., *Narration*, *Explanation*, *Correction*). Agreement is defined to be the shared entailments of the agents' public commitments. We show that this makes precise predictions about implicit agreement. The theory also provides a consistent interpretation of disputes and models what content is agreed upon when a dispute has taken place.

1 Introduction

A semantic framework for interpreting dialogue should provide an account of what is agreed upon and what is in dispute. In spite of the interest that dialogue interpretation has attracted, *implicit agreement* remains difficult to analyse in logically precise terms. For instance, consider the following real dialogue, described in (Sacks et al., 1974, p.717):

(1) a. Mark (to Karen and Sharon): Karen 'n' I're having a fight,
b. Mark (to Karen and Sharon): after she went out with Keith and not me.
c. Karen (to Mark and Sharon): Wul Mark, you never asked me out.

Intuitively, Mark and Karen agree that they had a fight, and that this was caused by Karen going out with Keith and not Mark. Thus *implicatures can be agreed upon*—that (1b) explains (1a) goes beyond compositional semantics. Furthermore, *agreement can be implicated*—Karen does not repeat (1a), (1b) or utter *OK* to indicate agreement. Finally, (1c) is not (yet) agreed upon. A theory of dialogue interpretation should predict these facts.

In principle, the *Grounding Acts Model* (GAM, Traum (1994), Poesio and Traum (1997)) supports implicit agreement. But the particular rules that are currently specified demand the recognition of an *acceptance* act for agreement to take place (Matheson et al., 2000), and the rules they offer don't predict such an acceptance act of (1a) and (1b) from Karen's utterance (1c). Segmented Discourse Representation Theory (SDRT, Asher and Lascarides (2003)) errs in the opposite direction. It stipulates that lack of disagreement implies agreement, and so (1c) is incorrectly predicted to be agreed upon. Thus, SDRT needs modification to deal with (1), just as GAM needs supplementation.

Agreement can also occur in the context of corrections or disputes. So disputes must receive a consistent interpretation; otherwise it will lead to inconsistent inferences about agreement. Consider dialogue (2) from the Toulouse-Stuttgart Procorpe corpus.

(2)
1.1. A: Je suis tombé en panne.
1.2. Peux-tu m'aider?
I have had a breakdown. Can you help me?
2.1. B: Où es-tu?
Where are you?
3.1. A: Je suis devant le refuge qui se trouve un km après Couiza.
3.2. Il y a la une cabine téléphonique.
I'm in front of the refuge that's 1 km after Couiza. There's a telephone booth.
4.1. B: Il y a plusieurs refuges aux alentours de Couiza.
4.2. Dans quelle direction es-tu parti de Couiza?
There are several refuges around Couiza. In which direction did you leave Couiza?
5.1. A: Je suis sorti par la route Paul Sabatier.
5.2. Puis j'ai roulé vers la montagne.
5.3. A une clairière j'ai tourne a droite.
I left by the Paul Sabatier road. I then drove toward the mountains. At a clearing I turned right.
6.1. B: Au grand carrefour?
At the big intersection?
7.1. A: Non, après, là ou on commence à avoir une belle vue sur la mer.
No afterwards, where you begin to have a nice view of the sea.
8.1. B: Ah, je vois, au Rocher du diable.
Oh, I see, at the Devil's cliff.
9.1. A: C'est possible, il y avait un gros rocher.
It's possible; there was a big cliff.
10.1. B: Donc tu es devant la refuge de la Maison de l'aigle et j'arrive tout de suite.
So, you're in front of the refuge at the house of the eagle. I'm coming right away

In (2), B's utterance (4.1) denies a uniqueness presupposition arising from the definite *the refuge that's 1 km after Couiza* in (3.1). But utterance (10.1) intuitively implicates B's agreement about the undisputed parts of this utterance; i.e., that A is in front of a refuge. In other words, some content that is uttered before the dispute takes place is agreed upon after it.

The content of utterance (4.1) in dialogue (2) is inconsistent with the uniqueness presupposition from utterance (3.1) that it denies. But a denial can also be implicated by an utterance that is consistent with the denied content, as the dialogue (3) from Walker (1996) and dialogue (4) show:

(3) a. A: The new student is brilliant and imaginative.
b. B: He's imaginative.

(4) a. A: John is not a good speaker
b. A: because he's hard to understand.
c. B: I agree that he's hard to understand.

Intuitively, by agreeing with a strict part of a prior contribution, B implicates a dispute with the other part—following Hirschberg (1985), Walker argues that (3b) generates a scalar implicature that B does not believe that the new student is brilliant. Similarly, B agrees in (4) with (4b) but implicates that he does not believe (4a).

To our knowledge, there is currently no formally precise account of dialogue that yields accurate interpretations of corrections and agreement. We aim to rectify this here. We will say that a proposition p is *grounded* just in case p is agreed by the dialogue agents to be true. This follows Clark's terminology, in particular the concept of grounding a joint action at level 4 (Clark, 1996, p388). Clark focusses mainly on grounding at the 'lower' levels; how agents ground what was meant, for instance. By contrast, in order to study grounding at the higher level, we will assume a highly idealised scenario where dialogue agents understand each other perfectly, resolving all ambiguities to the same specific values. One of Clark's main claims is that grounding at all levels occurs only when there is positive evidence for it, and here we aim to explore in a logically precise manner what evidence suffices for grounding a proposition. In future work, we intend to demonstrate that our definitions can be extended to model grounding at the lower levels too; this will involve modelling misunderstandings.

The rest of the paper is as follows. In Section 2 we use existing approaches to motivate general criteria for any adequate theory of agreement and disputes. These criteria could be formalised within any Information State Update (ISU) approach to dialogue (see Larsson and Traum (2000) for an overview). But in Section 3 we argue for SDRT as a starting point, and we describe how to extend it to meet the criteria. This extension is then formalised in detail: Section 4 defines the syntax and dynamic semantics of the language in which logical forms are expressed and Section 5 defines the *glue logic* that uses linguistic form and contextual information to construct logical form. We analyse several real dialogues and explore how agreements and disputes interact with anaphora.

2 Motivation

Many current theories of dialogue adopt the Information State Update (ISU) paradigm (Larsson and Traum, 2000): an utterance triggers an update to the information state representing the dialogue context to form a new information state. The differences among ISU approaches reside in three areas: the type of information that's recorded in an information state; the type of update operations that are permissible; and the controls over their application. We will examine two ISU theories that already offer an account of agreement: the Grounding

Acts model (GAM, Traum (1994), Poesio and Traum (1997, 1998)) and Segmented Discourse Representation Theory (SDRT, Asher and Lascarides (2003)).[1]

The Grounding Acts Model (GAM) links the speech acts performed with their semantic and cognitive effects, including effects on grounding. Poesio and Traum (1998) formalise this: as the dialogue proceeds, each dialogue agent builds the *conversational information state* (CIS) shown in (5).

(5)

$$\begin{array}{|l|}\hline \mathit{G},\ \mathit{DU1},\ \mathit{DU2},\ \mathit{DU3},\ \mathit{UDU},\ \mathit{CDU} \\ \hline G = \ldots \\ \mathit{DU1} = \ldots \\ \mathit{DU2} = \ldots \\ \mathit{DU3} = \ldots \\ \mathit{UDU} = \langle \mathit{DU1}, \mathit{DU3} \rangle \\ \mathit{CDU} = \mathit{first}(\mathit{UDU}) = \mathit{DU1} \\ \hline \end{array}$$

A CIS is a DRS (Kamp and Reyle, 1993) where each of its referents G (for "ground"), *DU* (for "discourse unit") etc. are also DRSs. The currently pending discourse units, which require further attention in the dialogue, are grouped under *UDU*, the top element being the current discourse unit (*CDU*). The update to a CIS borne from a particular speech act (and the conditions under which the act can be performed) are then specified as changes to (and conditions on) the DRSs G, *UDU* etc. For example, the speech act where B **asserts** K to A updates the common ground G to include an event e' that B intends A to believe K and a conditional event e'' that should A accept the assertion, then A would be socially committed to B to believe K (shown via the attitude **SCCOE**):

(6)

Name:	**Assertion**
Condition on update:	$G :: [e : \mathbf{Assert}(B, A, K)]$
Update	$G{+} = [e']e' : \mathbf{Try}(B, \lambda s'.s' : \mathbf{Bel}(A, K)),$ $[e'']e'' : \mathbf{Accept}(A, e) \Rightarrow [s \mid s : \mathbf{SCCOE}(A, B, K)]$

The update rules form an inheritance hierarchy, and so in addition to the effects specified in (6) **Assertion** inherits from its supertype speech act **Directive** an **obligation** on A to address e, and from its supertype **Statement** a **social commitment** (**SCCOE**) of B to A to K. The application of update rules are then controlled by *decision trees* that use the linguistic analysis of the utterance and the prior CIS to predict which speech acts were performed.

The hallmark of content p being mutually agreed upon is that the relevant agents A and B are socially committed to each other to p; in other words, (7) is a part of the CIS (since we ignore misunderstandings we can assume that A's and B's CISs are identical):

(7)

$$G :: \begin{array}{l} \mathbf{SCCOE}(A, B, p) \\ \mathbf{SCCOE}(B, A, p) \end{array}$$

[1] Agreement and disputes aren't examined in Ginzburg's (2008) ISU theory, and so we don't examine it here.

A social commitment to p can be created either by stating (or asserting) p, or by accepting a prior assertion that p. With this in mind, consider (1). While it is possible in principle to provide decision trees in GAM that will recognise (1c) as an *acceptance* act of Mark's prior assertions, the rules they actually provide only recognise (1c) as an assertion. Consequently, GAM as it stands does not recognise that Karen is socially committed to (1a), (1b) and a causal relation between them. GAM needs to be supplemented with rules for inferring that Karen was *implicitly accepting* Mark's contribution (it also needs rules that recognise Mark's contribution as conveying a causal relation between (1a) and (1b)).

Let's compare this with SDRT's analysis of (1) (Asher and Lascarides, 2003). In SDRT, an information state is a Segmented Discourse Representation Structure (SDRS)—or a set of them when ambiguities remain unresolved (see Section 5). An SDRS consists of a set of *labels* that each represent a unit of discourse, and a function that associates each label with a formula representing the unit's interpretation. These formulae include rhetorical relations between labels. The hierarchical structure on labels that this creates constrains anaphoric dependencies (see Sections 3 and 5 for details). We will explore the effects of agreement and disputes on anaphora in the course of this paper. But for now we focus on SDRT's predictions about agreement in (1). Its logical form is (1′), where $\pi_{1.1}$, $\pi_{1.2}$ and $\pi_{2.1}$ label the contents of the clauses (1a–c) respectively (we will often use the convention that the n^{th} utterance in the m^{th} turn is indexed $m.n$), and π labels the content of the dialogue segment that is created by the rhetorical connections:

(1′) $$\pi : \mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$$

For reasons of space, the semantic representations of the clauses are omitted from (1′). In fact, we will often gloss the content of a label π as K_π. But assuming that $K_{\pi_{1.1}}$ to $K_{\pi_{2.1}}$ are expressed appropriately, (1′) entails following: Karen and Mark were having a fight, this was caused by Karen going out with Keith and not Mark, and this in turn was caused by Mark not asking Karen out. In short, the presence of rhetorical relations in logical form captures content that's linguistically implicit (here, the causal relations). The update rules and the processes that control their application are rendered in a default *glue logic* which constructs logical forms like (1′) via axioms that validate default inferences about rhetorical connections among the discourse units (see Section 5).

The logical form (1′) captures Karen's commitments but loses Mark's. This contributes to SDRT's problematic analysis of agreement. (Asher and Lascarides, 2003, p363) stipulates that in the absence of divergent rhetorical relations (in other words, speech acts of denial such as *Correction* and *Counterevidence*), all the content is agreed upon. So SDRT wrongly predicts that (1c) is agreed upon. In essence, SDRT takes silence to mean assent—a mistake, given Clark's (1992) empirical findings that grounding requires positive evidence.

While SDRT's current model of agreement is wrong, we believe that rhetorical relations like *Explanation* are a crucial ingredient in any adequate model. Rhetorical relations are types of *relational speech acts* (Asher and Lascarides, 2003): they are speech acts because explaining something or continuing a narrative are things that people *do* with utterances; and they are relational because the successful performance of the speech act *Explanation*, for instance, is logically dependent on the content of the utterance (or sequence of utterances) that is being explained. When Karen utters (1c), she is explaining (1b). So even though the compositional semantics of (1c) does not entail (1b), its illocutionary contribution does entail it—or more accurately entails that Karen is publicly committed to it.[2] This shows how recognising an implicit acceptance is logically dependent on recognising the particular relational speech act(s) that the agent performed. An implicit acceptance act follows if that speech act is *left-veridical*—in other words, it entails the content of its first argument—and that first argument labels an utterance (or sequence of utterances) that were spoken by another agent.

While Karen is committed to the speech act $\mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$, Mark is committed to $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$. So if agreement is defined as shared public commitment, then (1b) is agreed upon. But clearly there is still something missing. By performing the speech act $\mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$, Karen (implicitly) accepts more of Mark's contribution than just (1b). The agreed content also includes (1a) and the fact that (1b) caused (1a). But this does not follow from the semantics of Karen's speech act $\mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$. Furthermore, it would be implausible to interpret Karen's speech act as explaining the content of Mark's entire turn—i.e., as $\mathit{Explanation}(\pi_{1M}, \pi_{2.1})$, where π_{1M} is the dialogue segment (1a–b), associated with the content $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$. It's simply not plausible to assume that Mark not asking her out explains why her going out with Keith and not Mark caused a fight. So Karen's commitment to Mark's entire turn must arise in some other way.

We believe, like GAM, that agreement should be defined as shared public or social commitment. This explains why positive evidence is necessary for it, as Clark and Schaefer (1989) claim: both agents must perform a speech act with appropriate semantic consequences in order to form a shared public commitment. But (1) shows that an agent can commit to more than just the rhetorical speech act that they performed. Here, Karen must be committed to more than (1c) explaining (1b). Her commitment that (1b) also explains (1a) must arise as a *consequence* of the fact that she explained (1b) and that Mark conveyed that (1b) explains (1a).

What this suggests, then, is that commitments persist from prior turns and are even transferred from one speaker to another. The basic intuition regarding Karen's endorsement of Mark's contribution is the following principle:

> *If one implicitly accepts a prior utterance, one normally also accepts its illocutionary effects.*

We will explain shortly why we believe that this principle should apply only to *implicit* endorsements, (i.e., it won't apply to utterances of the form *OK*, *I agree*, repeating content, and the like). We will also explain why it is a default (note the use of the word *normally*). But first let's see how this principle has the desired effect in (1). It predicts that Karen commits not only to the speech act $\mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$ that she performed, but also through doing so she commits to the illocutionary effects $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$ that Mark committed to when he uttered (1b). So the shared entailments of Karen's and Mark's public commitments match what they intuitively agree upon, as desired.

[2] We think Hamblin's (1970) notion of public commitment is the appropriate speaker's attitude to the moves he makes in dialogue (see also Gaudou et al. (2006)). We explore the links between commitments and other attitudes like belief in Asher and Lascarides (2008). Poesio and Traum (1998) don't provide truth conditions for **SCCOE**, but it too can be viewed as public commitment.

Exactly which commitments persist from a prior information state to the current one depends on the speech acts that are performed in the current turn, and how they are semantically related to the prior commitments. One might wonder, for instance, why the above principle does not apply to explicit acceptance acts. This is because deciding what's agreed upon by an explicit acceptance amounts to simply identifying the first argument of the (relational) acceptance act (or equivalently, determining the semantic scope of the acceptance act). In GAM, this speech act is called **Accept**; in the version of SDRT from Asher and Lascarides (2003) it is called *Acknowledgement* (while acknowledging an understanding of what was said is represented with the so-called metatalk relation *Acknowledgement**). In contrast, GAM and Clark (1996) use *Acknowledgement* to represent grounding an understanding. Therefore, to avoid confusion, we will use the rhetorical relation *Acceptance* to represent an explicit endorsement (even within SDRT).

If an utterance like *OK* is intended to endorse the illocutionary effects of an utterance and not just its semantics, then this should be represented by making the first argument of the *Acceptance* relation the discourse segment whose content entails those illocutionary effects. In other words, the first argument to *Acceptance* should be the discourse unit π that outscopes both π_1 and π_2, where $R(\pi_1, \pi_2)$ represents the illocutionary effects that are being explicitly endorsed. So making implicatures and not just compositional semantics accepted follows from the truth conditional interpretation of the *Acceptance* relation that is part of logical form, so long as the first argument to the acceptance act (or, equivalently, the *semantic scope* of the acceptance) is chosen correctly. This makes any persistence axioms for explicit acceptance acts redundant. Rather, what is required are principles for identifying the first argument of the *Acceptance* relation.

Moreover, examples (3) and (4), repeated here with labels for particular discourse units, show that it would be *wrong* for the above principle to apply to explicit endorsements.

(3) a. A: [The new student is brilliant]$_{\pi_{1.1}}$ and [he's imaginative]$_{\pi_{1.2}}$.

b. B: [He's imaginative]$_{\pi_{2.1}}$

(4) $\pi_{1.1}$. A: John is not a good speaker

$\pi_{1.2}$. A: because he's hard to understand.

$\pi_{2.1}$. B: I agree that he's hard to understand.

Walker (1996) argues that (3), which features an explicit acceptance, triggers a scalar implicature that B does *not* believe that the new student is brilliant. However, this could not be modelled straightforwardly if the above principle applied to explicit endorsements: A's (prior) commitment in (3) to *Continuation*$(\pi_{1.1}, \pi_{1.2})$ is consistent with *Acceptance*$(\pi_{1.2}, \pi_{2.1})$, and so if the above principle applied to explicit acceptances, B would be (wrongly) committed to *Continuation*$(\pi_{1.1}, \pi_{1.2})$ and thus to the new student being brilliant. To capture the right predictions about B's beliefs would then require us to conclude that B is somehow violating Sincerity conditions; i.e., that he does not believe something that he has publicly committed to. This would be highly counterintuitive. In fact, by being *very specific* about exactly which part of A's contribution B accepts, he is also conveying a *lack* of commitment to the other parts of A's contribution. Similarly in (4), the compositional semantics of $\pi_{2.1}$, which makes the content of $\pi_{1.2}$ an argument to the predicate *agree* given the syntactic complement of the verb, seems to suggest that B has accepted $\pi_{1.2}$ and *no more* than this. So while *Explanation*$(\pi_{1.1}, \pi_{1.2})$ is consistent with accepting $\pi_{1.2}$, it would be wrong to assume that B becomes committed to it.

Of course, these scalar implicatures about B's beliefs apply, as Walker (1996) attests, only when the compositional semantics of the acceptance act is highly specific. If it is expressed in an unspecific way, like *OK*, a default seems to apply that the first argument to *Acceptance* is the *entire* last turn. In other words, if $\pi_{2.1}$ in (4) were *OK*, then the update rules should predict a 'wide-scope' interpretation of the endorsement act, making B committed to *Acceptance*$(\pi_{1A}, \pi_{2.1})$, where π_{1A} is the discourse unit corresponding to A's entire first turn, that is associated with the content *Explanation*$(\pi_{1.1}, \pi_{1.2})$. Even if B's response in (4) is as shown below (so that $\pi_{2.1}$ is no longer a syntactic complement to *agree*), a different interpretation is more salient:

(4) $\pi_{2.1}$. B: OK,
$\pi_{2.2}$. B: he's hard to understand

Namely, *OK* (explicitly) endorses all of A's turn, while $\pi_{2.2}$ explains why that acceptance act was performed. In terms of rhetorical relations, this is expressed with *Acceptance*$(\pi_{1A}, \pi_{2.1}) \wedge$ *Explanation**$(\pi_{2.1}, \pi_{2.2})$ (*Explanation** is a metatalk relation, where the second argument explains why the speech act expressed by the first argument was performed).

We proposed making the principle that the implicit endorsement of an utterance is also an endorsement of its illocutionary effects a *default* principle, as opposed to monotonic. This is motivated by the need to account for dialogues like (8), where C implicitly endorses $\pi_{2.2}$, but does so with a speech act whose illocutionary effects conflict with those of $\pi_{2.2}$:

(8) $\pi_{1.1}$. A: How is James doing?
$\pi_{2.1}$. B: Actually, not so well.
$\pi_{2.2}$. B: His wife left him.
$\pi_{3.1}$. C: And he lost his job.

Intuitively, C uses $\pi_{3.1}$ as a continuation of $\pi_{2.2}$, and the resulting *segment* explains $\pi_{2.1}$. This implies that $\pi_{2.2}$ is *not* the sole cause for James not doing so well. In contrast, the illocutionary effect *Explanation*$(\pi_{2.1}, \pi_{2.2})$ that B commits to by uttering $\pi_{2.2}$ implicates, via a scalar implicature, that $\pi_{2.2}$ is the sole cause. This conflict among B's and C's intended illocutionary effects should block C from committing to *Explanation*$(\pi_{2.1}, \pi_{2.2})$, even though he has implicitly endorsed $\pi_{2.2}$. Making our principle default achieves this, so long as the consistency checks that accompany the default reasoning are based on what follows non-monotonically from the individual premises, rather than what follows from them monotonically.

Now let's consider dialogues that feature (explicit) disputes. Dialogue (9) is taken from a chat forum,[3] and it involves two agents—PianoCraft and WildWind—discussing David Bowie's album *Outside* (we have labelled discourse units that are relevant to our discussion).

[3] `www.teenagewildlife.com/Interact/cp/showflat.pl?Cat=&Board=interp&Number=251307&page=17&view=collapsed&sb=7&part=`.

(9) $\pi_{1.1}$. Pianocraft: It's a safe bet that most of us love *Outside*, the dark themes, the death abyss, the mutilations, murder, and ignorance at the turn of the century.

$\pi_{1.2}$. Pianocraft: That's exactly the album's theme, that violence is now a form of entertainment, and we watch the destruction as a form of beauty.

$\pi_{2.1}$. Wildwind: Much like the statement that was attempted in *Natural Born Killers.*

$\pi_{2.2}$. Wildwind: However, I don't think we like the album because of the theme, that is, I don't think it's the violence and death that makes the album worth listening to.

$\pi_{2.3}$. WildWind: I also know at least one person who can't stand the album *because* of the theme (Wildwind's emphasis).

This example provides further evidence for representing an agent's commitments with rhetorical connections. This is because intuitively, WildWind agrees with $\pi_{1.1}$ and $\pi_{1.2}$, but disputes that the latter *explains* the former (see $\pi_{2.2}$). This illocutionary effect is something that PianoCraft intended to convey, but left linguistically implicit—there is no cue phrase such as *because* in his utterances. However, a representation of PianoCraft's commitment must include this intended effect, so that the dispute between PianoCraft and WildWind can be reflected in their conflicting commitments.

As we mentioned in Section 1, one basic requirement is that disputes receive a consistent interpretation. Individuating each agents' commitments in the dialogue's information state helps to achieve this, because it is consistent for two agents to make mutually inconsistent commitments. However, an agent's beliefs can change as the dialogue progresses, and such changes can surface in the dialogue by an agent committing to something that is inconsistent with his earlier commitments. For instance, consider the simple (constructed) dialogue (10):

(10) $\pi_{1.1}$. A: It's raining.

$\pi_{2.1}$. B: No it's not.

$\pi_{3.1}$. A: Oh, you're right (*uttered after A looks out the window*)

A's utterance $\pi_{3.1}$ is an (explicit) *Acceptance* of B's utterance $\pi_{2.1}$, and given the way the anaphoric elements in $\pi_{2.1}$ are resolved, this commits A to *it's not raining*, contrary to his commitments from the first turn. If one simply adds this new commitment to the representation of his old one via conjunction, then contrary to intuitions A's commitments will become inconsistent, making the entire information state inconsistent.

There are several ways one could maintain consistency in A's commitments. First, one could assume that updating an agent's commitments involves *truth maintenance*, thereby incorporating downdating and revision. In dialogue (10), this would trigger the removal of the prior and conflicting commitment to $\pi_{1.1}$. However, while this might work in principle, modelling downdating and revision is an unsolved problem for dynamic first-order models. So we will take a different approach in this paper that avoids revision, both in the model theory and in the update rules for constructing information states (see Section 3 for an overview). But the general point to make here is that an adequate model of disputes must incorporate principles that identify which prior commitments are preserved and which are dropped.

Dialogue (11) shows that should A accept B's correction of his prior turn, then the principles for identifying ongoing prior commitments must validate that A remains committed to the parts of that turn that B did not dispute (even if B did not endorse them either).

(11) $\pi_{1.1}$. A: John went to jail.
$\pi_{1.2}$. A: He embezzled the pension funds.
$\pi_{2.1}$. B: No, it was BILL who stole the pension funds.
$\pi_{2.2}$. B: I was at the trial.
$\pi_{3.1}$. A: Oh, OK.
$\pi_{4.1}$. B: John did go to jail though.

In this dialogue, B uses the second turn ($\pi_{2.1}$ and $\pi_{2.2}$) to dispute $\pi_{1.2}$, and hence also the speech act $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$ that A performed. A accepts this correction in the third turn $\pi_{3.1}$. And finally, B uses utterance $\pi_{4.1}$ to accept $\pi_{1.1}$. Crucially, this is sufficient for $\pi_{1.1}$ to be agreed between them; A need not repeat his commitment to $\pi_{1.1}$. And so $\pi_{1.1}$ must be a part of A's commitments at the third turn, even though by accepting B's correction A can no longer be committed to $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$. Thus when A endorses a dispute of his prior contribution, A should remain committed to those parts of his contribution that B did *not* deny. Here, the moves B makes in the second turn make him neutral about John going to jail (although his later turn commits him to this); and so by accepting B's denial A maintains a commitment to John going to jail. More generally, the principles that establish which prior commitments persist should ensure the following:

> *A speaker remains committed to those parts of his prior turn(s) that he does not disavow.*

To summarise this section, we have argued that any theory of dialogue should meet the following criteria:

1. Information states should individuate among the commitments of each dialogue agent, with agreement defined to be shared public commitment.

2. An agent's commitments should include rhetorical connections. This is required for modelling implicit acceptance (e.g,. (1)) and denials when what is denied is only the rhetorical connection (e.g., (9)).

3. The task of computing what's agreed upon when an endorsement is explicit is logically equivalent to identifying the first argument of the speech act *Acceptance*.

4. The update rules for computing the current information state should identify which prior commitments persist in the current state, and are even transferred from one agent to another. In particular, the rules should ensure that:

   (a) When one implicitly endorses a prior utterance, one normally also endorses its illocutionary effects.

   (b) An agent remains committed to those parts of his prior turn(s) that he does not disavow.

3 An Account of Dialogue in SDRT

SDRT as it stands fails to distinguish among the agents' commitments, but it includes rhetorical relations. GAM distinguishes among each agent's commitments, but its current rules don't link acceptance acts with relational speech acts and their semantics. Neither GAM nor SDRT include update rules that effect a transfer of a commitment from one agent to another in appropriate contexts (see criterion 4 above), which we argued is required for an adequate analysis of (1).

We now propose a formal theory within SDRT that meets all the above criteria. We use SDRT for two main reasons. First, logical form in SDRT is defined more abstractly than in GAM, and this allows us to think of logical forms as first order models, as Asher and Lascarides (2003) discuss. We can therefore exploit standard preservation results from model theory when stipulating which commitments persist as the dialogue proceeds. In fact, we'll take full advantage of this by ensuring that updates to logical form always *extend* the prior logical form and never revise it, even when an agent drops a prior commitment (in which case, the updated logical form adds a non-truth preserving operator with semantic scope over that prior commitment). The dynamics in the interpretation of logical forms will model how commitments change as the dialogue proceeds.

Secondly, SDRT makes specific predictions about which parts of a dialogue context can be antecedents to subsequent anaphora. This gives us the opportunity to explore how agreement and disputes interact with anaphora. SDRT's model of anaphora relies on separating the representation of dialogue content (i.e., an SDRS) from the conversational implicatures which it gives rise to. For instance, the implicature in (3) that B does not believe that the new student is brilliant is not a part of its SDRS—the only place from which antecedents to surface anaphora can be chosen. And this helps us to explain anomalous subsequent anaphora such as A uttering *Why not?* (meaning *Why don't you believe that the new student is brilliant*) or B uttering *That's why we shouldn't accept him on the course* (meaning "we shouldn't accept him on the course because I don't believe he's brilliant"). It also predicts that B cannot respond to an assertion that p with *I do too* (meaning "I believe that too"). In GAM, all effects on content and cognitive states are expressed in the same information state, and so it would need further refinements to explain this anaphoric data.

To make SDRT meet the criterion that it distinguishes among each agent's public commitments, we make the logical form of each dialogue turn a tuple of SDRSs, one for each agent. An SDRS for a given agent at a given turn will represent all of his current commitments, including the ongoing commitments from prior turns (this is a way to avoid revision in the model theory, as discussed in Lascarides and Asher (2008)). The logical form of the dialogue overall is the logical forms of each of its turns. All dialogue participants build all the agents' SDRSs and not just those representing his own commitments (and since we ignore misunderstandings we can assume they all build the same SDRSs). We assume an extremely simple notion of turns, where turn boundaries occur whenever the speaker changes (even if this happens mid-clause).

The proposed logical form of (1) is shown in Table 1. As before, $\pi_{1.1}$, $\pi_{1.2}$ and $\pi_{2.1}$ label the contents of the clauses (1a) to (1c) respectively (these semantic representations are omitted for reasons of space). We also from now on adopt a convention that the label of the dialogue

Turn	Mark's SDRS	Karen's SDRS	Sharon's SDRS
1	π_{1M} : *Explanation*$(\pi_{1.1}, \pi_{1.2})$	$\emptyset$	$\emptyset$
2	π_{1M} : *Explanation*$(\pi_{1.1}, \pi_{1.2})$	π_{2K} : *Explanation*$(\pi_{1.1}, \pi_{1.2}) \wedge$ *Explanation*$(\pi_{1.2}, \pi_{2.1})$	$\emptyset$

Table 1: A representation of dialogue (1).

segment of turn n with speaker d is π_{nd} (here, M stands for Mark, K for Karen and S for Sharon). We call this tuple of SDRSs a Dialogue SDRS or DSDRS.

There is a sharing of labels across the SDRSs in a DSDRS. This reflects the reality that a speaker may perform a relational speech act whose first argument is part of someone else's turn, or part of his own previous turns. As a special case, it captures the fact that through speaking an agent can reveal his commitments (or lack of them) to content that another agent conveyed, even if this is linguistically implicit.

Leaving aside for now how this DSDRS is *constructed*, let's focus instead on its dynamic semantic interpretation. Asher and Lascarides (2003) define precisely the context change potential (CCP) of an individual SDRS. Since the logical form of a dialogue turn is now a tuple of SDRSs, its CCP is the *product* of the CCPs of the the individual SDRSs. In other words, the context of evaluation C_d for interpreting a dialogue turn is a set of dynamic contexts for interpreting SDRSs—one for each agent $a \in D$, where D is the set of all dialogue agents:

$$C_d = \{\langle C_a^i, C_a^o \rangle : a \in D\}$$

Thus C_a^i and C_a^o are world assignment pairs, given the definitions from Asher and Lascarides (2003). Equation (12) defines the dynamic interpretation of veridical relations (e.g. *Narration*, *Explanation*, *Acceptance*), where meaning postulates then constrain the illocutionary effects $\varphi_{R(\alpha,\beta)}$—e.g., for *Narration* they stipulate the spatio-temporal progression of the events (m in $[\![.]\!]_m$ stands for monologue). Equation (13) defines the dynamic interpretation of *Correction* (see Section 4 for details).

(12) $(w, f)[\![R(\alpha, \beta)]\!]_m(w', g)$ iff $(w, f)[\![K_\alpha \wedge K_\beta \wedge \varphi_{R(\alpha,\beta)}]\!]_m(w', g)$

(13) $(w, f)[\![\mathit{Correction}(\alpha, \beta)]\!]_m(w', g)$ iff $(w, f)[\![(\neg K_\alpha) \wedge K_\beta \wedge \varphi_{\mathit{Corr}(\alpha,\beta)}]\!]_m(w', g)$

The CCP of a dialogue turn $T = \{S_a : a \in D\}$ is the product of the CCPs of its SDRSs:

$$C_d[\![T]\!]_d C_d' \text{ iff } C_d' = \{\langle C_a^i, C_a^o \rangle \circ [\![S_a]\!]_m : \langle C_a^i, C_a^o \rangle \in C_d, a \in D\}$$

Accordingly, entailments from a dialogue turn can be defined in terms of the entailment relation $\models_m$ for SDRSs afforded by $[\![.]\!]_m$:

$$T \models_d \phi \text{ iff } \forall a \in D, S_a \models_m \phi$$

This makes $\models_d$ the shared entailment of each agent's public commitments. And we assume that content ϕ is grounded or agreed upon by all the dialogue agents by a dialogue turn T iff $T \models_d \phi$. Similarly, ϕ is agreed on by a subgroup $G \subseteq D$ of dialogue agents iff for all agents $a \in G$, $S_a \models_m \phi$ (where $S_a \in T$). In line with intuitions, this definition predicts that by the end of the second turn of dialogue (1), Mark, Karen and Sharon don't agree on anything other than logical truths (because Sharon isn't committed to anything). But Mark and Karen agree that they had a fight, which was caused by Karen going out with Keith and not Mark. Finally, given that the SDRSs for a dialogue turn reflect *all* an agent's current commitments, the CCP of the dialogue overall is the CCP of its last turn.

But what are the update rules for constructing this DSDRS-style information state? Asher and Lascarides (2003) provide a detailed *glue logic* for constructing a single SDRS. This includes default axioms for identifying rhetorical connections among discourse units given their linguistic form and context of utterance (see Section 5). The rules are defaults because one never has complete information about the context, including speaker intentions. Clearly, these default axioms are needed for constructing DSDRSs. But in addition, the glue logic must now incorporate principles for computing which commitments persist from prior turns (see criterion 4 from Section 2).

The data from Section 2 showed that the type of the speaker's current speech act influences which parts of the prior commitments persist and which don't. For convenience, we state one `Persistence Principle`, together with separate axioms that for each type of (current) speech act computes the so-called *undenied commitments* of a prior constituent:[4]

- `The Persistence Principle:`
  The undenied commitments of a constituent β in turn n persist to turn $n+1$.

For instance, the implicit endorsement in (1) concerns *simple left veridical relations* (e.g., *Explanation* and *Background*)—they entail the content of the first argument but are not *Acceptance* (which is an explicit endorsement via utterances like *OK* or repetition of content):

- `Undenied Commitments` for Simple Left Veridical Relations:
  If A commits to $R(\pi_j, \pi_i)$ where R is simple left veridical, then for any R' and any π_k such that B is (already) committed to $R'(\pi_k, \pi_j)$ (or to $R'(\pi_j, \pi_k)$), *as long as it's consistent to do so*, the undenied commitments of π_j include $R'(\pi_k, \pi_j)$ (or $R'(\pi_j, \pi_k)$).

`Undenied Commitments` ensures that when A implicitly endorses π_j, he normally also endorses the illocutionary effects that B had intended π_j to have. This applies when constructing the DSDRS for (1). Mark's first turn is constructed via existing glue logic axioms. Similarly, Karen's utterance $\pi_{2.1}$ will be recognised via these axioms as an *Explanation* of $\pi_{1.2}$. The result satisfies the premise to `Undenied Commitments` (substitute Karen and Mark for A and B, $\pi_{1.1}$, $\pi_{1.2}$ and $\pi_{2.1}$ for π_k, π_j and π_i respectively, and *Explanation* for R and R'). The consequence—that *Explanation*$(\pi_{1.1}, \pi_{1.2})$ is an undenied commitment of $\pi_{1.2}$—is consistent with the premises of the glue logic and therefore inferred. So by the `Persistence Principle` Karen's SDRS is as shown in Table 1.

`Undenied Commitments` means that when A implicitly endorses a part of the prior speaker's turn, he normally commits to all of that prior turn. If that prior turn endorsed A's turn prior

[4] The term *undenied commitment* isn't ideal. As we saw in (3) $\pi_{2.1}$ is consistent with the utterance $\pi_{1.1}$ and so one might want to think of $\pi_{1.1}$ as an undenied commitment. But that is not the intended interpretation of the term here: $\pi_{1.1}$ must not be thought of as an undenied commitment (even though it is undenied *content*), because as Walker (1996) attests $\pi_{2.1}$ implicitly rejects $\pi_{1.1}$.

to that, then a sequence of applications of this principle ensures that A remains committed to all the commitments he made the last time he spoke. In other words, this principle models as a special case the fact that A remains committed to his prior commitments in those dialogue contexts where each speaker endorses the prior speaker's turn. To see how this works, consider how `Undenied Commitments` contributes to the construction of the logical form for dialogue (14) (this is an extract from dialogue r008c from the Verbmobil corpus, (Wahlster, 2000)).

(14) $\pi_{1.1}$. A: Can I meet with you sometime in the next two weeks?

$\pi_{1.2}$. A: What days are good for you?

$\pi_{2.1}$. B: Well, I have some free time on almost every day except Fridays.

$\pi_{2.2}$. B: Fridays are bad.

$\pi_{2.3}$. B: So any day besides Friday we can probably work out a time.

$\pi_{3.1}$. A: Well next week I am out of town Tuesday, Wednesday and Thursday.

$\pi_{3.2}$. A: So perhaps Monday afternoon?

In turn 1, A's speech act in uttering the question $\pi_{1.2}$ is *Q-Elab*$(\pi_{1.1}, \pi_{1.2})$—that is, $\pi_{1.2}$ is a question that's designed to gather information that helps to achieve the goal behind the question $\pi_{1.1}$ (to meet in the next two weeks). So by the semantics of *Q-Elab*, the illocutionary contribution of the question $\pi_{1.2}$ is paraphrased as "What days *in the next two weeks* are good for you?". B's utterances in the second turn are an indirect answer, so in SDRT the segment they form attaches to $\pi_{1.2}$ with the left-veridical relation *IQAP* (Indirect Question Answer Pair). `Undenied Commitments` therefore predicts that B is also committed to *Q-Elab*$(\pi_{1.1}, \pi_{1.2})$, as shown in Table 2. This correctly predicts that B's response must be interpreted so that its temporal expressions (e.g., *every day* in $\pi_{2.1}$) denote times within the next two weeks. In turn 3, A uses $\pi_{3.1}$ to elaborate a plan to achieve the goals set out by $\pi_{2.3}$—to meet on any day (in the next two weeks) besides Friday. This type of speech act is rendered with the left-veridical relation *Plan-Elab* in SDRT. Therefore, by `Undenied Commitments`, the illocutionary effects of $\pi_{2.3}$ become a part of A's commitments. This makes A committed to the content π of B's indirect answer to A's original question, shown in Table 2. This triggers another application of `Undenied Commitments` (for π), resulting in A's initial commitment *Q-Elab*$(\pi_{1.1}, \pi_{1.2})$ being added to his current SDRS, as shown in Table 2 (for reasons of space, we have not re-iterated the content associated with label π in A's third SDRS).

Clark (1992) argues that positive evidence is needed for grounding, but he doesn't make precise exactly what counts as sufficient positive evidence. Similarly, Poesio and Traum (1998) don't provide rules for inferring when a speaker has performed an implicit acceptance. We have made the quantity of positive evidence that's needed logically precise, in terms of the (relational) speech acts that both speakers perform, and the logical relationships between the semantics of those speech acts. The `Persistence Principle` and `Undenied Commitments` capture a general class of examples involving implicit acceptance: there is sufficient positive evidence to ground p if p follows from the speech act that B performed in uttering π_j, and A performs a (left-veridical) speech act other than *Acceptance* that connects to π_j, whose semantic consequences are consistent with p. Sufficient positive evidence for grounding in the context of *explicit* endorsements rests on the formal semantic interpretation of the relevant

Turn	A's SDRS	B's SDRS
1	$\pi_{1A} : \mathit{Q\text{-}Elab}(\pi_{1.1}, \pi_{1.2})$	$\emptyset$
2	$\pi_{1A} : \mathit{Q\text{-}Elab}(\pi_{1.1}, \pi_{1.2})$	π_{2B}: $\mathit{Q\text{-}Elab}(\pi_{1.1}, \pi_{1.2})\wedge$ $\mathit{IQAP}(\pi_{1.2}, \pi)$ π : $\mathit{Explanation}(\pi_{2.1}, \pi_{2.2})\wedge$ $\mathit{Result}(\pi_{2.1}, \pi_{2.3})$
3	π_{3A}: $\mathit{Q\text{-}Elab}(\pi_{1.1}, \pi_{1.2})\wedge$ $\mathit{IQAP}(\pi_{1.2}, \pi)\wedge$ $\mathit{Plan\text{-}Elab}(\pi_{2.3}, \pi_{3.1})\wedge$ $\mathit{Q\text{-}Elab}(\pi_{3.1}, \pi_{3.2})$	π_{2B} : $\mathit{Q\text{-}Elab}(\pi_{1.1}, \pi_{1.2})\wedge$ $\mathit{IQAP}(\pi_{1.2}, \pi)$ π : $\mathit{Explanation}(\pi_{2.1}, \pi_{2.2})\wedge$ $\mathit{Result}(\pi_{2.1}, \pi_{2.3})$

Table 2: A representation of dialogue (14).

Turn	A's SDRS	B's SDRS	C's SDRS
1	$\pi_{1.1} : K_{\pi_{1.1}}$	$\emptyset$	$\emptyset$
2	$\pi_{1.1} : K_{\pi_{1.1}}$	π_{2B}: $\mathit{IQAP}(\pi_{1.1}, \pi_{2.1})\wedge$ $\mathit{Explanation}(\pi_{2.1}, \pi_{.2.2})$	$\emptyset$
3	$\pi_{1.1} : K_{\pi_{1.1}}$	π_{2B}: $\mathit{IQAP}(\pi_{1.1}, \pi_{2.1})\wedge$ $\mathit{Explanation}(\pi_{2.1}, \pi_{.2.2})$	π_{3A}: $\mathit{IQAP}(\pi_{1.1}, \pi_{1.2})\wedge$ $\mathit{Explanation}(\pi_{1.2}, \pi)$ π: $\mathit{Continuation}(\pi_{2.1}, \pi_{3.1})$

Table 3: The logical form of (8).

speech act *Acceptance* and the rules for choosing the first argument to this relation. We will examine in detail the evidence for grounding in the context of disputes in Section 3.1.

As explained in Section 2, making `Undenied Commitments` a default principle helps to construct the right logical form of dialogue (8), which is shown in Table 3.

(8) $\pi_{1.1}$. A: How is James doing?
$\pi_{2.1}$. B: Actually, not so well.
$\pi_{2.2}$. B: His wife left him.
$\pi_{3.1}$. C: And he lost his job.

As explained there, there's a (nonmonotonic) semantic conflict between $\mathit{Explanation}(\pi_{2.1}, \pi_{2.2})$ and $\mathit{Explanation}(\pi_{2.1}, \pi)$ (where π is $\mathit{Continuation}(\pi_{2.2}, \pi_{3.1})$), a speech act that the glue logic must identify as one that C intended to perform by uttering $\pi_{3.1}$. The former nonmonotonically entails, via a scalar implicature, that James losing his wife is the sole reason he's not doing well, while the latter (monotonically) entails it was a strict part of it. Thus even though these speech acts satisfy the antecedent to `Undenied Commitments`, its consequent isn't inferred.

We saw in Section 2 that different ways of expressing an explicit endorsement seem to come with different preferences for resolving the scope ambiguity as to what prior commitments are being endorsed. The logical form for (3) that's shown in Table 4 reflects the 'narrow-scope' acceptance discussed earlier. The joint entailment of the SDRSs for the second turn

Turn	A's SDRS	B's SDRS
1	π_{1A} : *Continuation*$(\pi_{1.1}, \pi_{1.2})$	$\emptyset$
2	π_{1A} : *Continuation*$(\pi_{1.1}, \pi_{1.2})$	π_{2K} : *Acceptance*$(\pi_{1.2}, \pi_{2.1})$

Table 4: A Representation of dialogue (3).

ensures that the new student being imaginative is agreed upon, but his being brilliant is not. This logical form, however, does *not* express an explicit dispute (i.e., it features no divergent relation like *Correction*). This is because we argued earlier that the denial in (3), while implicated, is not a part of what B *said*, because it cannot act as an antecedent to subsequent anaphora. SDRT models the cognitive effects of utterances in a separate but related logic in which axioms of agent rationality and cooperativity are encoded. Here, that cognitive logic should ensure that by failing to endorse $K_{\pi_{1.1}}$ (as shown in B's SDRS), B does not believe $K_{\pi_{1.1}}$. We forego details of this cognitive reasoning here (but see Asher and Lascarides (2008)).

We now examine explicit disputes in Section 3.1, focussing in particular on which commitments persist when a dispute takes place, and how disputes affect anaphoric interpretation.

3.1 Corrections

Corrections are highly complex. Like explicit endorsements, they have various scope possibilities. Furthermore, as detailed in van Leusen (1994) and Asher (2004), the focus structure of the correction reveals which element in the context the speaker denies, and this element may be only part of even a minimal discourse unit (e.g., see earlier discussion of dialogue (2)). We argued in Section 2 that to compute the right commitments and get the facts right about agreement, an agent who accepts a correction of his prior commitments must remain committed to those parts of his contribution that were *not* denied. In this section, we briefly recount the decidable method from Asher and Lascarides (2003) for computing these 'undenied' parts of a corrected discourse unit, and then we'll exploit it to specify the necessary update rules on commitments.

To illustrate SDRT's method for computing the 'undenied' or 'neutral' parts of a corrected discourse unit, consider an extract from (11) (pitch accents are shown with small caps):

(11) $\pi_{1.2}$. A: John embezzled the pension funds.
$\pi_{2.1}$. B: No, it was BILL who stole the pension funds.

B's utterance commits him to *Correction*$(\pi_{1.2}, \pi_{2.1})$, and hence to the negation of $K_{\pi_{1.2}}$. But B has not denied every aspect of $K_{\pi_{1.2}}$. For instance, he remains neutral about whether the funds were *embezzled* (as opposed to merely stolen), with the funds entrusted to Bill. Asher and Lascarides (2003) exploit the focus structure of $\pi_{2.1}$ to predict this.

Following Krifka (1991), Rooth (1992), Steedman (2000) and others, the linguistic form partitions the content of $\pi_{2.1}$ into a pair of formal representations, at least one of which is a λ-abstract: the focus (we'll label this $\pi_{2.1}^{f,\lambda}$) and the topic or 'background' ($\pi_{2.1}^{b,\lambda}$). It's a partition in that applying the λ-abstract to the other expression yields the content of $\pi_{2.1}$. In this example, the *it*-cleft and placement of the pitch accent make $\pi_{2.1}^{f,\lambda}$ equal to $\lambda P.P(\mathit{bill})$, and therefore $\pi_{2.1}^{b,\lambda}$ must be (roughly) $\lambda x(\mathit{steal}(e,x,y) \wedge \mathit{unique}(x))$, where $\mathit{unique}(x)$ is a gloss for the uniqueness conditions from the *it*-cleft, and y co-refers with the pension funds mentioned in $\pi_{1.2}$.[5] The propositional content of the background replaces the λ-abstracts with existential quantifiers (Rooth, 1992); we'll refer to this as $\pi_{2.1}^{b}$. So the background proposition $\pi_{2.1}^{b}$ can be paraphrased as *Someone (unique) stole the pension funds.*

This focus structure determines which parts of $\pi_{1.2}$ are denied and which are not: a mapping ζ maps the focus and background of $\pi_{2.1}$ into a partition of $\pi_{1.2}$ of its denied and undenied parts respectively (Asher and Lascarides, 2003, p.351). One can identify ζ while avoiding undecidable consistency checks by exploiting the fact that $\pi_{1.2}$ and $\pi_{2.1}$ have similar predicate argument structures (we'll see shortly what happens when the predicate argument structures aren't alike). In other words, $\zeta(\pi_{2.1}^{f,\lambda})$ is $\lambda P.P(\mathit{john})$, and therefore, $\zeta(\pi_{2.1}^{b,\lambda})$ must be $\lambda x(\mathit{embezzle}(e',x,y))$. The undenied part of $\pi_{1.2}$—which by convention we call $\pi_{1.2}^{b}$—is also computed by replacing the λ-abstracts in $\zeta(\pi_{2.1}^{b,\lambda})$ with existential quantifiers. In other words, $\pi_{1.2}^{b}$ is *someone embezzled the pension funds.* Intuitively, B is 'neutral' about this content; he is neither committed (yet) to the stealing being an embezzlement, nor committed to its negation. This neutrality means $K_{\pi_{1.2}^{b}}$ should not be *entailed* by $\mathit{Correction}(\pi_{1.2},\pi_{2.1})$—the speech act that B is committed to. But as we'll see shortly, $K_{\pi_{1.2}^{b}}$ will affect A's commitments, should he subsequently endorse B's correction.

Sometimes, corrections are much less specific than in (11):

(15) a. A: John embezzled the pension funds.

b. B: You're WRONG.

Arguably the focus of (15b) is now the whole sentence (see the discussion of all-rheme utterances in Steedman (2000)) and so all of (15a) is denied.[6]

Like explicit endorsements, a *Correction* can include more than just the last clause within its scope. When it does, *recursion* as well as prosody and paraphrase cues enable us to compute exactly what is denied and what is not. The full dialogue (11) demonstrates this. By denying $\pi_{1.2}$, B must also be committed to denying $\mathit{Explanation}(\pi_{1.1},\pi_{1.2})$. So arguably, $\mathit{Correction}(\pi_{1A},\pi_{2.1})$ as well as $\mathit{Correction}(\pi_{1.2},\pi_{2.1})$ express B's commitments after turn 2. We've already computed the part of $K_{\pi_{1.2}}$ that B is neutral about. The 'neutral' part of $K_{\pi_{1A}}$ is then determined *recursively*: roughly, one replaces all occurrences of $\pi_{1.2}$ in $K_{\pi_{1A}}$ with a new label $\pi_{1.2}^{b}$ that labels the 'neutral' part of $\pi_{1.2}$, and all relations R to which $\pi_{1.2}$ is an argument (in this case, *Explanation*) with dynamic conjunction. In words, this predicts that B is (currently) neutral about the following: John went to jail and someone embezzled the funds. Or to put it another way, B has not committed yet to this content nor to its negation. The reason R is replaced with dynamic conjunction is because one cannot assume that the illocutionary effects of uttering $\pi_{1.2}$ also accompany its neutral counterpart: for instance, "John went to jail; someone embezzled the funds" is not a coherent *Explanation.* We will see

[5] We explain how co-reference is inferred in the glue logic in Section 5.

[6] We ignore the complexities rendered by presuppositions, such as that John exists. These are represented in SDRT via separate labels from the asserted content, and so the correction denies a presupposition only if the label of the presupposed content is outscoped by the first argument to *Correction.*

shortly that it is sometimes convenient to express dynamic conjunction as a relation V over labels:

$$C[\![V(\pi_i, \pi_j)]\!]C' \text{ iff } C[\![K_{\pi_i} \wedge K_{\pi_j}]\!]C'$$

The neutral content of π_{1A} in (11) can then be expressed as $V(\pi_{1.1}, \pi^b_{1.2})$, where $\pi^b_{1.2}$ labels the content *someone embezzled the pension funds*.

A's utterance $\pi_{3.1}$ in (11) endorses B's utterance. And thus A's new commitment should not entail that John embezzled the funds, and instead entail that Bill stole those funds (and he was the only one to do so). But as we argued in Section 2, it should continue to entail that John went to jail (and that the stealing was an embezzlement). That way, John's going to jail will be agreed upon as an effect of B's utterance $\pi_{4.1}$.

We can capture these effects on commitments with two general principles. First, if B's correction contains within its scope an utterance π_i, then it also corrects all labels that outscope π_i. This reflects the intuition that when B corrects something, he also corrects the content of any dialogue segment whose semantics is dependent on the denied content. Given that in SDRT we infer all the discourse relations that are consistent with an attachment point, we will automatically include these additional corrections in the logical form. For the analysis of (11), this means that B's SDRS for the second turn is (16).

(16) $\pi_{2B} : \mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge \mathit{Correction}(\pi_{1.2}, \pi_{2.1}) \wedge \mathit{Explanation}^*(\pi_{2.1}, \pi_{2.2})$

Crucially, no rules validate an inference that the 'neutral' content of $\pi_{1.2}$ (or π_{1A}) is an undenied commitment, making the `Persistence Principle` irrelevant when constructing B's SDRS in (11). The lack of such a rule reflects intuitions about grounding—in (11), we accurately predict that John's going to jail isn't agreed upon after the second turn.

A's third turn is an acceptance of the entire segment π_{2B}; i.e., $\mathit{Acceptance}(\pi_{2B}, \pi_{3.1})$. Furthermore, as a side-effect of this speech act, A is also committed to $\mathit{Correction}(\pi_{1A}, \pi_{3.1})$. SDRT's glue logic captures this plurality in A's dialogue move already: the semantic consequences of $\mathit{Acceptance}(\pi_{2B}, \pi_{3.1})$ resolves the content of $\pi_{3.1}$ to be equivalent to that of π_{2B}; this commits A to the negation of his first turn, thereby fulfilling the necessary consequences of a corrective move; and so the general glue-logic principle that the necessary consequences of a speech act are normally sufficient for that speech act to be inferred applies. So A is now committed to speech acts that are semantically incompatible with his earlier commitments to $K_{\pi_{1A}}$. But his commitments are consistent because no update rules make $K_{\pi_{1A}}$ an undenied commitment and hence a part of A's current commitments. Nevertheless, we need to ensure that A now maintains commitments to the parts of π_{1A}'s content that B was neutral about.

We therefore articulate an update rule that 'imports back' A's prior commitment to the undenied bits of the dispute. This 'minimises' the commitments that A drops, and they derive from the scope possibilities for corrections and endorsements. A speaker who accepts a correction adjusts the scope of the acceptance and his interpretation of the correction appropriately to the commitments that he continues to hold. Similarly a speaker who corrects a correction will adjust the scope of his correction and his interpretation of the first correction in light of what commitments he continues to hold. This idea of minimising dropped commitments in response to disputes is captured informally in the principle Accepting Corrections (`AC`) below. This is the second of our update principles, and we formalise it in Section 5:

- `Undenied Commitments for Acknowledgements of Corrections (AC):`
  If A endorses π_j, where B's SDRS (already) contains *Correction*(π_i, π_j) and A was committed to π_i in the prior turn, then the undenied commitments of π_j include the 'neutral' or undenied content of π_i, as long as it's consistent to do so.

We saw earlier that the undenied content of $K_{\pi_{1A}}$ in (11) is $V(\pi_{1.1}, \pi^b_{1.2})$, where $\pi^b_{1.2}$ is a (new) label that expresses *someone embezzled the pension funds*. So according to `AC` and the `Persistence Principle`, $V(\pi_{1.1}, \pi^b_{1.2})$ is a part of A's SDRS for the third turn of (11).

The relation V is not a coherence relation. And yet it is part of the representation of A's third turn. The need to introduce V into our inventory of relations is an inevitable consequence of two things. First, in order to avoid revision in the model theory, our SDRSs at a given turn represent *all* of an agent's current commitments from the beginning of the dialogue to the end of that turn. Secondly, when an agent accepts a correction, his remaining commitments need not, on their own, be related in a coherent way. But a dialogue is coherent so long as each utterance *as it's interpreted* attaches to its context with a coherence relation.

To capture A's commitments correctly, this undenied content must be placed in a veridical part of A's SDRS, by which we mean that it must be connected to the root label via a sequence of zero or more veridical relations. In general, we can ensure the appropriate effect by putting the undenied content in the same scopal position as the acceptance of the correction. Thus in dialogue (11), since *Acceptance*$(\pi_{2B}, \pi_{3.1})$ is labelled with the root label π_{3A}, so is the relation $V(\pi_1, \pi^b_2)$. Finally, it makes semantic sense to make this undenied content a *Background* to the segment consisting of corrective moves, since the undenied parts of disputes behave semantically like 'given' or background information (van der Sandt, 2001). We achieve this by making the last label of the undenied content a second argument to the relation *Background*, where the first argument is the segment consisting of the corrections (π_{2B} in our example). Thus we arrive at the following, more specific definition of `AC`:

- `Undenied Commitments for Acknowledgements of Corrections (AC):`
  If A's SDRS contains $\lambda : R(\pi_j, \pi_k)$ where R is left-veridical, and B's SDRS contains *Correction*(π_i, π_j), and A was committed to the content of π_i in the prior turn, then the undenied commitments of π_j is computed as follows:

  (a) the undenied content of π_i is computed recursively by replacing corrected labels that are outscoped by π_i with their undenied parts and veridical rhetorical relations that involved these labels with V;

  (b) the resulting representation of the undenied content of π_i is assigned the label λ;

  (c) A *Background* relation, also labelled with λ, between the label of the correcting segment π_i and the last label in this undenied content is added (in dialogue (11), this introduces π_{3A} : *Background*$(\pi_{2B}, \pi^b_{2.1})$).

The entire representation for (11) is thus depicted in Table 5. It exhibits the plurality of speech acts that an agent can utter with a single discourse unit. However, in light of this plurality, SDRT assumes a *compactness* principle whereby speech acts that are inferred to be a part of one's commitments, but which are semantically and structurally redundant given his other commitments, are omitted from the representation. In the analysis of (11) this

Turn	A's SDRS	B's SDRS
1	$\pi_{1A} : \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$	$\emptyset$
2	$\pi_{1A} : \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$	$\pi_{2B} :$ $\mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge$ $\mathit{Correction}(\pi_{1.2}, \pi_{2.1}) \wedge$ $\mathit{Explanation*}(\pi_{2.1}, \pi_{2.2})$
3	$\pi_{3A} :$ $V(\pi_1, \pi_2^b) \wedge \mathit{Background}(\pi_{2B}, \pi_{1.2}^b) \wedge$ $\mathit{Acceptance}(\pi_{2B}, \pi_{3.1})$	$\pi_{2B} :$ $\mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge$ $\mathit{Correction}(\pi_{1.2}, \pi_{2.1}) \wedge$ $\mathit{Explanation*}(\pi_{2.1}, \pi_{2.2})$
4	$\pi_{3A} :$ $V(\pi_{1.1}, \pi_{1.2}^b) \wedge \mathit{Background}(\pi_{2B}, \pi_{1.2}^b) \wedge$ $\mathit{Acceptance}(\pi_{2B}, \pi_{3.1})$	$\pi_{4B} :$ $\mathit{Acceptance}(\pi_{1.1}, \pi_{4.1}) \wedge$ $\mathit{Contrast}(\pi_{2B}, \pi_{4.1})$

Table 5: A representation of dialogue (11).

means that *Correction*$(\pi_{1A}, \pi_{3.1}) \wedge$ *Correction*$(\pi_{1.2}, \pi_{3.1})$ is omitted from the representation of $K_{\pi_{3A}}$ (because *Acceptance*$(\pi_{2B}, \pi_{3.1})$ entails the same content and predicts the same available labels, as we'll see shortly).

In Section 4 we'll unpack the dynamic interpretation of the representation in Table 5. Here, we simply state that each turn has a consistent interpretation (i.e., the output state is not $\perp$). But the SDRSs in turn 2 are not consistent with each other (one entails John embezzled the funds and the other that he didn't). At the end of turn 3, that John didn't embezzle the funds and that Bill stole the funds is agreed upon. B remains neutral about whether that stealing was an embezzlement and whether John went to jail, while A is committed to it. By the end of turn 4, B commits to John going to jail, and hence A and B agree on it.

B's fourth turn $\pi_{4.1}$ has been interpreted as accepting $\pi_{1.1}$. But those readers who are familiar with SDRT may notice that $\pi_{1.1}$ is not on the right frontier of the discourse structure and therefore it is not available for attachment. We now examine this issue in some detail, to explain how we predict in SDRT that this attachment to $\pi_{1.1}$ is legitimate.

Asher and Lascarides (2003) argue that in general, the available labels are those on the right frontier of the discourse structure: that is, the label of the last clause, and any label that is related to it via a sequence of subordinating discourse relations (e.g., *Explanation*) and/or the semantic outscoping relation (recall that π_i outscopes π_j if K_{π_i} includes π_j). Asher and Lascarides (2003) and Asher (1993) highlight particular kinds of examples where this right-frontier constraint breaks down (e.g., in the presence of structural relations like *Contrast* and *Parallel*). Here, we observed via (11) that it also appears to break down in the face of certain acceptances (i.e., $\pi_{4.1}$). The same is true of corrections: for instance, instead of uttering $\pi_{4.1}$ in (11), B could have uttered $\pi'_{4.1}$.

(11) $\pi'_{4.1}$. B: John didn't go to jail either.

Intuitively, this should attach with *Correction* to $\pi_{1.1}$ and hence attach off the right frontier. Alternatively, instead of B saying $\pi_{4.1}$, A may have said it. This should attach not only to A's prior acceptance act $\pi_{3.1}$, but also arguably to $\pi_{1.1}$ directly.[7] Similarly, consider the two alternative responses B might make to A's turn in (17):

[7] One might also wonder why A would utter something he is already committed to. But such redundant

(17) $\pi_{1.1}$. A: John embezzled a pension fund,

$\pi_{1.2}$. A: and then he was convicted of tax evasion.

$\pi_{2.1}$. B: It was BILL who embezzled the funds.

$\pi'_{2.1}$. B: It wasn't a pension fund.

A's commitments are to *Narration*$(\pi_{1.1}, \pi_{1.2})$. So $\pi_{1.1}$ is not on the right frontier. But arguably, $\pi_{2.1}$ should commit B to *Correction*$(\pi_{1.1}, \pi_{2.1})$. This not only gets the facts right about commitment, but it also allows us to finesse what he denies and what he doesn't (see the exploitation of recursion in the analysis of (11)), which can prove important for getting the facts right for subsequent agreement. Similarly, the anaphoric data in $\pi'_{2.1}$ also suggest that $\pi_{1.1}$ should be available to utterances of a certain form.

We propose that these variants of (11) and (17) constitute a special case of what Asher (1993) called *discourse subordination*: there is enough compositional semantic information in the utterance, derivable from syntax and prosodic cues, to guide the attachment to a particular point in the discourse structure. In contrast, a highly underspecified compositional semantics doesn't guide attachment off the right frontier: responding to A's move in (17) with *you're wrong* (or *OK*) would be interpreted as a correction (or an acceptance) of at least $K_{\pi_{1.2}}$. Thus in dialogues where the compositional semantics of an utterance is sufficiently specific, the availability constraints should be relaxed to include in principle *all* constituents in the previous turn. This is also what happens in our various versions of (11): B's very specific utterance(s) $\pi_{4.1}$ (and it's alternative $\pi'_{4.1}$) suffice to pick out $\pi_{1.1}$ as an attachment point, even though it's not on the right frontier. Similarly, if A instead of B were to say $\pi_{4.1}$ it can attach to $\pi_{1.1}$. In fact, many types of speech acts allow discourse subordination, at least in principle (see Asher (1993) for examples involving *Elaboration*).

Because discourse subordination is only possible when an utterance has a certain specific form, the right frontier constraint on availability is not rendered entirely impotent. For instance, dialogue (18) illustrates `AC`'s predictions about anaphoric dependencies (imagine that A and B are telling an agent C what happened in the pub last night):

(18) $\pi_{1.1}$. A: John came in the pub.

$\pi_{1.2}$. A: He sat down on an old coat.

$\pi_{1.3}$. A: He drank a beer.

$\pi_{2.1}$. B: I definitely saw that he didn't drink ANYTHING.

$\pi_{3.1}$. A: Oh, OK.

$\pi_{3.2}$. A: ??It was made of tweed.

The SDRS for utterances $\pi_{1.1}$ to $\pi_{3.1}$ is shown in Table 6. Axioms for inferring within the glue logic the *Narration* and *Correction* moves in the first two turns are discussed in Section 5 (see also Asher and Lascarides (2003)). A then uses the third turn to accept B's correction $\pi_{2.1}$. Thus according to `AC`, we must add a representation of the undenied parts of π_{1A} (and $\pi_{1.3}$)

moves can happen in monologue (*Everyone talked. So Harry talked too*) and also in dialogue. The move is not entirely redundant here anyway, since the illocutionary effects that are borne from the *Contrast* relation that's rendered by *though* are new commitments, even if the commitment to John going to jail is not.

Turn	A's SDRS	B's SDRS
1	π_{1A} : $\mathit{Narration}(\pi_{1.1}, \pi_{1.2}) \wedge \mathit{Narration}(\pi_{1.2}, \pi_{1.3})$	$\emptyset$
2	π_{1A} : $\mathit{Narration}(\pi_{1.1}, \pi_{1.2}) \wedge \mathit{Narration}(\pi_{1.2}, \pi_{1.3})$	π_{2B} : $\mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge$ $\mathit{Correction}(\pi_{1.3}, \pi_{2.1})$
3	π_{3A} : $\mathit{Narration}(\pi_{1.1}, \pi_{1.2}) \wedge$ $\mathit{Background}(\pi_{2B}, \pi_{1.2}) \wedge$ $\mathit{Acceptance}(\pi_{2.1}, \pi_{3.1})$	π_{2B} : $\mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge$ $\mathit{Correction}(\pi_{1.3}, \pi_{2.1})$

Table 6: A representation of the dialogue $\pi_{1.1}$ to $\pi_{3.1}$ in (18).

Turn	PianoCraft's SDRS	Wildwind's SDRS
1	π_{1P} : $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$	$\emptyset$
2	π_{1P} : $\mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$	π_{1M}: $\mathit{Commentary}(\pi_{1.2}, \pi_{2.1}) \wedge \mathit{Contrast}(\pi_{2.1}, \pi)$ $\wedge \mathit{Explanation*}(\pi, \pi_{2.3})$ π: $\mathit{Correction}(\pi_{1P}, \pi_{2.2})$

Table 7: A representation of dialogue (9).

to A's SDRS as well. The background part $\pi^b_{1.3}$ of $\pi_{1.3}$ is $\top$. So by the compactness principle mentioned earlier, $\pi^b_{1.3}$ is omitted from the representation and hence so is $V(\pi_{1.2}, \pi^b_{1.3})$. This makes the undenied content of π_{1A} simply $\mathit{Narration}(\pi_{1.1}, \pi_{1.2})$. Thus $\pi_{1.2}$ is its last label, and so by AC part (c) it is linked to π_{2B} with *Background*. Finally, AC assigns $\mathit{Narration}(\pi_{1.1}, \pi_{1.2})$ and $\mathit{Background}(\pi_{2B}, \pi_{1.2})$ the root label, as shown in Table 6. This logical form correctly predicts that A is (currently) committed to the proposition that John came in the pub and that he (then) sat on an old coat. But $\pi_{1.2}$ is unavailable for subsequent rhetorical connections: it is not the first argument to a subordinating relation where the second argument is available; nor does it outscope any available label. This correctly predicts that the indefinite noun phrase *a coat* is not an available antecedent to the pronoun *it* in the anomalous continuation $\pi_{3.2}$ of this dialogue. Similarly, $\pi_{1.1}$ is not available. And so had this first utterance been *John came in with a jacket over his arm*, we would correctly predict that *it* in $\pi_{3.2}$ cannot co-refer with the jacket. Thus the principle AC preserves both commitments and anaphoric dependencies appropriately in such examples. Discourse subordination is not an option for $\pi_{3.2}$, for in contrast to the utterances in (11) and (17), it lacks the specific prosodic and linguistic cues that are required for an attachment off the right frontier.

We finish this section with a logical form for (9), given in Table 7 (the semantic representations of the clauses are omitted). Roughly, this logical form commits PianoCraft to the album's theme of violence explaining why people like it, while Wildwind's (complex) utterance $\pi_{2.2}$ commits him to people liking the album and to its theme being violence, but he denies that one explains the other. He is also committed to a justification for performing this denial—the anecdote in $\pi_{2.3}$ that he knows at least one person who hates the album because of its theme.

4 The Logical Form of Dialogue

Section 3 describes a theory of agreement and disputes in SDRT that meets the four general criteria from Section 2. We extended logical forms from being a single SDRS to a tuple of them, each one representing the public commitments of an individual agent. We also argued in favour of extending the *glue logic* to include principles that stipulate when a prior commitment persists, given the particular speech acts that the speaker performed in the current turn. We saw that the scope possibilities of acceptances and corrections interact with the persistence of commitments and also with the interpretation of anaphora. Our task now is to formalise this fully. We start by focussing on the formal language in which the logical form of dialogue is expressed, defining its syntax and dynamic semantic model theory. In Section 5 we will extend the glue logic with axioms that formalise `Undenied Commitments` and `AC`—the principles from Section 3 that identify which prior commitments persist.

4.1 Syntax

We now define the syntax of SDRSs and then use this to define the syntax of Dialogue SDRSs (DSDRSs). Definitions 1 and 2 are from Asher and Lascarides (2003).

Definition 1 The Syntax of SDRS-Formulae

SDRS-formulae are constructed from the following vocabulary:

vocab-1. microstructure: A classical first order vocabulary, consisting of predicates, terms, boolean connectives and quantifiers, augmented with the modal operator $\Box$, the modal operator δ that turns formulae into action terms ($\delta\phi$ is the action of bringing it about that ϕ), a modal operator ! that turns an action term into a formula ($!\delta\phi$ is used to represent imperatives); and the operators '?' and λ-terms for representing questions as $?\lambda x_1 \ldots \lambda x_n \phi$, each x_i corresponding to a *wh*-element.

vocab-2. labels: π, π_1, π_2, etc.

vocab-3. a set of relation symbols for discourse relations: R, R_1, R_2, etc.

The set $\mathcal{L}$ of well-formed SDRS-formulae is defined as follows:

1. Let $\mathcal{L}_{basic}$ be the set of well-formed formulae that are derived from vocab-1 using the usual syntax rules for first order modal languages with action terms. Then $\mathcal{L}_{basic} \subseteq \mathcal{L}$.
2. If R is an n-ary discourse relation symbol and $\pi_1, \ldots, \pi_n$ are labels, then $R(\pi_1, \cdots, \pi_n) \in \mathcal{L}$.
3. For $\phi, \phi' \in \mathcal{L}$, $(\phi \wedge \phi'), \neg\phi \in \mathcal{L}$.

Definition 2 An SDRS

Let $\mathcal{L}$ be the set of SDRS-formulae. Then an SDRS is a triple $\langle \Pi, F, \mathit{last} \rangle$, where:

- Π is a set of labels; i.e. $\Pi \subseteq$ vocab-2.
- *last* is a label in Π (intuitively, this labels the last clause); and
- F is a function that assigns each member of Π a member of $\mathcal{L}$.
- We say that π *immediately outscopes* π' iff $F(\pi)$ contains π' as a literal. The relation $\succ$ that is its transitive closure satisfies the following two constraints: it forms a well-founded partial order over Π; and it has a unique root (that is, there is a unique $\pi_0 \in \Pi$ such that $\forall \pi \in \Pi$, $\pi_0 \succeq \pi$).

When there is no confusion, we may write $\langle \Pi, F \rangle$ instead of $\langle \Pi, F, \mathit{last} \rangle$.

In the prior sections, the value of F is shown with colons: $\pi : \phi$ means $F(\pi) = \phi$. Definition 3 now formally defines DSDRSs, as illustrated in Tables 1 to 7.

Definition 3 A Dialogue SDRS (DSDRS)

Let D be a set of agents. Then a DSDRS is a tuple $\langle n, T, \Pi, F, \mathit{last} \rangle$, where:

- n is a natural number (intuitively, $j \leq n$ is the j^{th} turn in the dialogue);
- Π is a set of labels;
- F is a function that assigns each member of Π a member of $\mathcal{L}$;
- T is a mapping from $[1, n]$ to a function from D into SDRSs, such that each SDRS is drawn from Π and F. That is, if $T(j)(d_i) = \langle \Pi_j^{d_i}, F_j^{d_i}, \mathit{last}_j^{d_i} \rangle$ where $j \in [1, n]$ and $d_i \in D$, then $\Pi_j^{d_i} \subseteq \Pi$, $F_j^{d_i} =_{def} F \restriction \Pi_j^{d_i}$ (that is, $F_j^{d_i}$ is F restricted to $\Pi_j^{d_i}$), and $\mathit{last}_j^{d_i} \in \Pi_j^{d_i}$.
- $\mathit{last} =_{def} \mathit{last}_n^d$, where d is the unique speaker of the last turn n (each turn has a unique speaker, because turn boundaries occur whenever the speaker changes).

We will sometimes write $T(j)(d_i)$ as $T^{d_i}(j)$.

Intuitively, T will map each turn and dialogue participant to an SDRS that represents everything he is currently publicly committed to. While one would expect a contribution in monologue to be coherent (especially if it is edited written text), the same is not true of dialogue. The above definitions allow for this: the formula that's associated with the root label of a turn may include parts that aren't rhetorically connected to any other part of the dialogue (consider in particular the range of well-formed SDRS-formulae from Definition 1). However, Definition 3 restricts each turn to having a unique root label. This makes an individual turn part of a single dialogue (addressed to a unique group of people). As Polanyi (1985) shows, this isn't always the case (e.g., a 'self-interruption' such as *Stop that you kids!*). We would need to analyse such a dialogue turn with multiple SDRSs, to reflect the idea that more than one conversation is going on simultaneously. But we ignore these complexities here.

Definition 3 allows label sharing across speakers and across turns. However, each label is associated with unique content in whatever turn $j \in [1, n]$ it appears in. That is, $\forall \pi \in$

$\Pi_l^{d_1} \cap \Pi_j^{d_2}$, $l, j \in [1, n]$, $d_1, d_2 \in D$, $F_l^{d_1}(\pi) = F_j^{d_2}(\pi)$. As we explained earlier, a situation where d_1 and d_2 interpret π differently won't correspond to a situation where π is assigned distinct contents in distinct SDRSs within the *same* DSDRS. Rather, it corresponds to a situation where d_1 and d_2 have each built different DSDRSs (although we won't explore misunderstandings further here).

There are several notational variants for DSDRSs. For instance, Table 1 and (19) are notational variants of the DSDRS (1′), the logical form for (1).

(1) $\pi_{1.1}$. Mark (to Karen and Sharon): Karen 'n' I're having a fight,
$\pi_{1.2}$. Mark (to Karen and Sharon): after she went out with Keith and not me.
$\pi_{2.1}$. Karen (to Mark and Sharon): Wul Mark, you never asked me out.

(1′) $\langle 2, T, \{\pi_{1M}, \pi_{2K}, \pi_{1.1}, \pi_{1.2}, \pi_{2.1}\}, F, \pi_{2.1}\rangle$, where:

- $F(\pi_{1.1}) = K_{\pi_{1.1}}$, $F(\pi_{1.2}) = K_{\pi_{1.2}}$, $F(\pi_{2.1}) = K_{\pi_{2.1}}$
  $F(\pi_{1M}) = \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$
  $F(\pi_{2K}) = \mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$
- $T(1) = \{(M, \langle\{\pi_{1M}, \pi_{1.1}, \pi_{1.2}\}, F_1, \pi_{1.2}\rangle), (K, \emptyset), (S, \emptyset)\}$,
  where $F_1 = F \restriction \{\pi_{1A}, \pi_{1.1}, \pi_{1.2}\}$
- $T(2) = \{(M, \langle\{\pi_{1M}, \pi_{1.1}, \pi_{1.2}\}, F_1\, \pi_{1.2}\rangle), (K, \langle\{\pi_{2K}, \pi_{1.1}, \pi_{1.2}, \pi_{2.1}\}, F_2, \pi_{2.1}\rangle), (S, \emptyset)\}$,
  where $F_2 = F \restriction \{\pi_{2B}, \pi_{1.1}, \pi_{1.2}, \pi_{2.1}\}$

(19) $\langle 2, T, \{\pi_{1M}, \pi_{2K}, \pi_{1.1}, \pi_{1.2}, \pi_{2.1}\}, F, \pi_{2.1}\rangle$, where:

- $F(\pi_{1.1}) = K_{\pi_{1.1}}$, $F(\pi_{1.2}) = K_{\pi_{1.2}}$, $F(\pi_{2.1}) = K_{\pi_{2.1}}$
  $F(\pi_{1M}) = \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$
  $F(\pi_{2K}) = \mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$
- $T^M(1) = \{\pi_{1A}, \pi_{1.1}, \pi_{1.2}\}$, $T^K(1) = T^S(1) = \emptyset$
- $T^M(2) = \{\pi_{1A}, \pi_{1.1}, \pi_{1.2}\}$, $T^K(2) = \{\pi_{2B}, \pi_{1.1}, \pi_{1.2}, \pi_{2.1}\}$, $T^S(2) = \emptyset$

We will usually represent DSDRSs as tables.

4.2 Semantics for DSDRSs

Asher and Lascarides (2003) offer a dynamic semantics of SDRSs, with contexts being world variable assignment pairs following Fernando (1994), van Eijk and Kamp (1997) and Groenendijk and Stokhof (1991). So the semantics of an SDRS defines how an input pair (w, f) changes to a different output one (w', g), where w and w' are possible worlds, and f and g are partial variable assignment functions.[8] While this semantics isn't ideal, it will suit

[8] As usual, the quantifier $\exists x$ *extends* the input assignment function f to a new one g that is like f, save that it is defined for x (van Eijk and Kamp, 1997). This ensures that assignments to x when interpreting the subformula $\psi(x)$ in $(\exists x \phi) \wedge \psi(x)$ match those that are used to satisfy the body ϕ of the quantified formula. In terms of natural language, this captures anaphoric dependencies across sentence boundaries.

our purposes here—to give an idea of how to interpret DSDRSs and predict agreement and disputes.[9]

One crucial task is to specify the content of rhetorical relations. Unlike other atomic formulae, these *update* the input context rather than acting as a test on it. Veridical discourse relations, such as *Explanation*, *Acceptance* and *Background* receive the following interpretation (where as before m in $[\![.]\!]_m$ stands for monologue):

- `Veridical Schema:`
  $(w,f)[\![R(\pi_1,\pi_2)]\!]_m(w',g)$ iff $(w,f)[\![K_{\pi_1} \wedge K_{\pi_2} \wedge \varphi_{R(\pi_1,\pi_2)}]\!]_m(w',g)$

Meaning postulates impose conditions on when $\varphi_{R(\pi_1,\pi_2)}$ is true for various relations R. This forms a major component of SDRT, since it constrains the illocutionary effects of speech acts. For instance, $\varphi_{\mathit{Explanation}(\pi_1,\pi_2)}$ entails that K_{π_2} is an answer to *why* K_{π_1}?, where the semantics of *why*-questions follows that given in Bromberger (1962) and Achinstein (1980). A divergent discourse relation such as *Correction* entails the negation of its first argument:

- `Semantics of Correction:`
  $(w,f)[\![\mathit{Correction}(\pi_1,\pi_2)]\!]_m(w',g)$ iff $(w,f)[\![(\neg K_{\pi_1}) \wedge K_{\pi_2} \wedge \varphi_{\mathit{Correction}(\pi_1,\pi_2)}]\!]_m(w',g)$

The meaning postulates for $\varphi_{\mathit{Correction}(\pi_1,\pi_2)}$ entail that K_{π_1} and K_{π_2} are mutually inconsistent. The semantics of an SDRS is then unpacked recursively, starting with the SDRS-formula that is assigned to its unique root label.

As we explained in Section 3 the interpretation of a DSDRS is the *product* of the interpretation of its component SDRSs. Definition 4 formalises this, with the context of evaluation being one dynamic proposition per agent:

Definition 4 Dynamic Semantics of DSDRSs

> Let K be a DSDRS $\langle n, T, \Pi, F, \mathit{last}\rangle$ with dialogue participants $D = \{d_1, \ldots, d_k\}$ and $j \in [1,n]$. Let σ_1 and σ_2 each be a set of k pairs of world assignment pairs and let ρ_i, $i \in [1,k]$ be a projection function onto the i^{th} element of σ_1 and σ_2. Then:
>
> $$\begin{array}{rcl} \sigma_1[\![K]\!]_d\sigma_2 & \text{iff} & \sigma_1[\![T(n)]\!]_d\sigma_2 \\ \sigma_1[\![T(j)]\!]_d\sigma_2 & \text{iff} & \forall d_i \in D, \rho_i(\sigma_2) = \rho_i(\sigma_1) \circ [\![T^{d_i}(j)]\!]_m \end{array}$$

In words, the context change potential (CCP) of the DSDRS is that of its last turn, which in turn is computed in terms of $[\![.]\!]_m$. The CCP of a dialogue turn updates the commitments each agent held in the dialogue initial state to include the (dynamic) content of his SDRS for that turn. Thus the CCPs of the turns in a DSDRS reflect the evolving commitments of each dialogue agent.

It is standard when defining truth in a product of models to say that $\mathfrak{A} \times \mathfrak{B} \models \phi$ iff $\mathfrak{A} \models \phi$ and $\mathfrak{B} \models \phi$. This natural definition readily transfers to our dynamic setting, providing a

[9] The problem is that it doesn't produce a suitable dynamic interpretation of questions. We fix this in Lascarides and Asher (forthcoming) by building on Groenendijk's (2003) semantics of questions.

definition of entailment for DSDRSs that is constant whatever the number of participants. Let $\mathcal{K} = \langle n, T, \Pi, F, \mathit{last} \rangle$ be a DSDRS for dialogue participants D, and let $\models_m$ be the dynamic semantic entailment relationship afforded by $[\![.]\!]_m$.[10]

Definition 5 Grounding

- $\mathcal{K} \models_d \phi$ iff for all $d_i \in D$, $T^{d_i}(n) \models_m \phi$, where n is the last turn in the conversation.

This notion of entailment for dialogue matches exactly the definition of agreement, or the grounding of a proposition.

The illocutionary contributions of speech acts are encoded in the semantics of DSDRSs. And thus our definition of agreement as a joint entailment on each agent's commitments models implicit agreement. For example the SDRSs for the last turn of dialogue (1), shown in Table 1, have the following dynamic implications:

$$\begin{array}{ccl}
\mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) & \text{iff} & K_{\pi_{1.1}} \wedge K_{\pi_{1.2}} \wedge \varphi_{\mathit{Explanation}(\pi_{1.1},\pi_{1.2})} \\
\begin{array}{l}\mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \\ \mathit{Explanation}(\pi_{1.2}, \pi_{2.1})\end{array} & \text{iff} & \begin{array}{l}K_{\pi_{1.1}} \wedge K_{\pi_{1.2}} \wedge \varphi_{\mathit{Explanation}(\pi_{1.1},\pi_{1.2})} \wedge \\ K_{\pi_{1.2}} \wedge K_{\pi_{2.1}} \wedge \varphi_{\mathit{Explanation}(\pi_{1.2},\pi_{2.1})}\end{array} \\
 & \text{only if} & K_{\pi_{1.1}} \wedge K_{\pi_{1.2}} \wedge \varphi_{\mathit{Explanation}(\pi_{1.1},\pi_{1.2})} \wedge K_{\pi_{2.1}}
\end{array}$$

Thus $\varphi_{\mathit{Explanation}(\pi_{1.1},\pi_{1.2})}$—i.e., the illocutionary effects that stem from $\pi_{1.2}$ explaining $\pi_{1.1}$—is agreed upon, even though the compositional semantics of neither Mark's nor Karen's utterances entail this.

Now let's examine some dialogues involving *Correction*. Consider first the dynamic semantic interpretation of dialogue (11), whose logical form is shown in Table 5. The entailments for the SDRSs of the last turn unpack as follows:[11]

$$\begin{array}{lcl}
K_{\pi_{3A}} & \text{iff} & V(\pi_{1.1}, \pi^b_{1.2}) \wedge \mathit{Background}(\pi_{2B}, \pi^b_{1.2}) \wedge \mathit{Acceptance}(\pi_{2B}, \pi_{3.1}) \\
 & \text{only if} & K_{\pi_{1.1}} \wedge K_{\pi^b_{1.2}} \wedge K_{\pi_{2B}} \wedge \varphi_{\mathit{Background}(\pi_{2B},\pi^b_{1.2})} \wedge K_{\pi_{3.1}} \wedge \varphi_{\mathit{Acc}(\pi_{2B},\pi_{3.1})} \\
 & \text{only if} & K_{\pi_{1.1}} \wedge \exists x(\mathit{embezzle}(e', x, y)) \wedge \mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge \mathit{Correction}(\pi_{1.2}, \pi_{2.1}) \wedge \\
 & & \quad \mathit{Explanation^*}(\pi_{2.1}, \pi_{2.2}) \wedge K_{\pi_{3.1}} \\
 & \text{only if} & K_{\pi_{1.1}} \wedge \exists x(\mathit{embezzle}(e', x, y)) \wedge \neg \mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \neg K_{\pi_{1.2}} \wedge K_{\pi_{2.2}} \\
K_{\pi_{2B}} & \text{iff} & \mathit{Acceptance}(\pi_{1.1}, \pi_{4.1}) \wedge \mathit{Contrast}(\pi_{2B}, \pi_{4.1}) \\
 & \text{only if} & K_{\pi_{1.1}} \wedge \mathit{Correction}(\pi_{1A}, \pi_{2.1}) \wedge \mathit{Correction}(\pi_{1.2}, \pi_{2.1}) \wedge \\
 & & \quad \mathit{Explanation^*}(\pi_{2.1}, \pi_{2.2}) \\
 & \text{only if} & K_{\pi_{1.1}} \wedge \neg \mathit{Explanation}(\pi_{1.1}, \pi_{1.2}) \wedge \neg K_{\pi_{1.2}} \wedge K_{\pi_{2.2}}
\end{array}$$

Thus the following are all grounded: $\neg K_{\pi_{1.2}}$ (i.e., that John did not embezzle the funds), $K_{\pi_{2.1}}$ (i.e., that Bill stole the funds), $K_{\pi_{2.2}}$ (i.e., that B was at the trial), $\neg \mathit{Explanation}(\pi_1, \pi_2)$ (i.e,. that John embezzling the funds did not cause him to go to jail) and $K_{\pi_{1.1}}$ (i.e., that John

[10] That is, $\phi \models_m \psi$ iff for all intensional structures $\mathfrak{A}$ and for all world assignment pairs (w, f), if there is a pair (w', f') such that $(w, f)[\![\phi]\!]^{\mathfrak{A}}_m(w', f')$, then there is a pair (w'', f'') such that $(w', f')[\![\psi]\!]^{\mathfrak{A}}_m(w'', f'')$.

[11] Roughly, $\varphi_{\mathit{Acc}(\alpha,\beta)}$ constrains K_β to be in a non-monotonic equivalence with $K_\alpha \wedge K_\beta$.

Turn	A's SDRS	B's SDRS
1	$\emptyset$	$\pi_{1.1} : K_{\pi_{1.1}}$
2	$\pi_{2A} : \mathit{IQAP}(\pi_{1.1}, \pi_{2.1})$	$\pi_{1.1} : K_{\pi_{1.1}}$
3	$\pi_{2A} : \mathit{IQAP}(\pi_{1.1}, \pi_{2.1})$	$\pi_{3B} :\ \mathit{Correction}(\pi_{2A}, \pi_{3.1})\wedge$ $\mathit{Correction}(\pi_{2.1}, \pi_{3.1})\wedge$ $\mathit{Explanation}*(\pi_{3.1}, \pi_{3.2})$
4	$\pi_{4A} :\ \mathit{IQAP}(\pi_{1.1}, \pi_{2.1}) \wedge V(\pi_{2.1}, \pi_{4.1})\wedge$ $\mathit{Correction}(\pi_{3B}, \pi_{4.1})\wedge$ $\mathit{Correction}(\pi_{3.1}, \pi_{4.1})\wedge$ $\mathit{Explanation}*(\pi_{4.1}, \pi)\wedge$ $\mathit{Elaboration}(\pi_{2.1}, \pi)$ $\pi : \mathit{Narration}(\pi_{4.2}, \pi_{4.3})$	$\pi_{3B} :\ \mathit{Correction}(\pi_{2A}, \pi_{3.1})\wedge$ $\mathit{Correction}(\pi_{2.1}, \pi_{3.1})\wedge$ $\mathit{Explanation}*(\pi_{3.1}, \pi_{3.2})$

Table 8: The DSDRS for dialogue (20).

went to jail). Further, A remains committed to the stealing being an embezzlement, but B is neutral about it.

In the real dialogue (20) (from personal communication) a correction is corrected:

(20) $\pi_{1.1}$. B: Hey, what happened?
$\pi_{2.1}$. A: I got the climb.
$\pi_{3.1}$. B: No you didn't.
$\pi_{3.2}$. B: I saw you fall off.
$\pi_{4.1}$. A: No.
$\pi_{4.2}$. A: First time I fell off.
$\pi_{4.3}$. A: Next time I redpointed it.

The DSDRS for (20) is shown in Table 8. Let's motivate it in detail. In the second turn, A publicly commits to a particular answer to the question π_1 being true; B then corrects this in the third turn, and he is also publicly committed to the fact that B saw A fall off explains why he asserts the correction $\pi_{3.1}$.

Now let's consider the fourth turn, in which A corrects B's correction. `AC` doesn't apply; intuitively A wishes to preserve his commitments to *all* the content prior to the dispute (and not just the parts of that content that B remained neutral about), and also preserve what's available within it. More formally, if the glue logic validates $\mathit{Correction}(\alpha, \beta)$, where the representation of the discourse context includes $\mathit{Correction}(\gamma, \alpha)$, then one needs an axiom ensuring that all of γ's content K_γ is part of the undenied commitments of α in this context. Moreover, the axiom should make available all the labels within γ that were available before γ was corrected. `Undenied Commitments for Denying Corrections` (`DC`) captures this, and it's a default just in case A's current speech act conflicts with preserving all prior commitments:

- `Undenied Commitments for Denying Corrections (DC):`
  If A's SDRS (for a given turn) contains $\lambda_1 : \mathit{Correction}(\alpha, \beta)$, and B's SDRS (for that

turn) contains $\lambda_2 : \mathit{Correction}(\gamma, \alpha)$, then normally the undenied commitments for α, which are assigned the label λ_1, include:

- $V(\gamma, \beta)$ if $K_\gamma \in \mathcal{L}_{basic}$.
- $K_\gamma \wedge V(\gamma', \beta)$, where γ' is the last label of K_γ, if $K_\gamma \notin \mathcal{L}_{basic}$.

DC determines the undenied commitments of a type of speech act in a particular context. The role of the relation V is to guarantee the correct effects both in semantics and in what's available. In dialogue (20), A's SDRS at the point where the dialogue is updated with $\pi_{4.1}$ includes $\mathit{Correction}(\pi_{3.1}, \pi_{4.1})$. Since $\mathit{Correction}(\pi_{2.1}, \pi_{3.1})$ and $\mathit{Correction}(\pi_{2A}, \pi_{3.1})$ are in B's SDRS, DC and the Persistence Principle yield that A's SDRS must include $\mathit{IQAP}(\pi_{1.1}, \pi_{2.1})$ and $V(\pi_{2.1}, \pi_{4.1})$ labelled with the root label (see Table 8).

Intuitively, $\pi_{4.2}$ and $\pi_{4.3}$ together form a narrative (which we've labelled π), which in turn elaborates A's original answer $\pi_{2.1}$ to the question (i.e., it elaborates how A got the climb), and also explains why he performs the correction $\pi_{4.1}$, making A's final commitments as shown in Table 8. At this point, the labels $\pi_{3.2}$ and $\pi_{4.2}$ are not available. This seems to concur with intuitions: the pronoun *it* in a subsequent utterance *It hurt* could not refer back to the fall.

The entailments of the DSDRS in Table 8 ensure that B is committed to A not getting the climb ($\neg K_{\pi_{2.1}}$) and that he fell off ($K_{\pi_{3.2}}$ entails this so long as seeing someone falling is interpreted evidentially). A, on the other hand, is committed to A getting the climb ($K_{\pi_{2.1}}$) and to first falling off ($K_{\pi_{4.2}}$) and then redpointing it ($K_{\pi_{4.3}}$). So at this point, the only content that is agreed upon is that A fell off. Neither *A got the climb* nor *A didn't get the climb* are agreed upon; nor is any answer to B's question.

5 Constructing Logical Form

Section 4 detailed the syntax and interpretation of the language in which logical forms for dialogue are expressed. This language expresses logical forms for a range of dialogues in a way that captures intuitions about agreement and disputes. Our task now is to describe how those logical forms are constructed during dialogue interpretation. This involves extending SDRT's glue logic to model the principles of dialogue interpretation from Section 3. So we start with a brief description of the glue logic (for details see Asher and Lascarides (2003)).

Roughly put, the glue logic exploits the underspecified semantics derived from linguistic form (e.g., Egg et al. (2001)). Underspecified logical forms (ULFs) are partial descriptions of (complete) logical forms, in our case DSDRSs. The glue logic incorporates default axioms for inferring (in a decidable manner) a more specific ULF: in other words, the logic identifies the pragmatically preferred way of resolving underspecified aspects of compositional content.

Rhetorical connections are inferred on the basis of default axioms of the form shown in Glue Logic Schema, where the symbols α and β are metavariables ranging over the labels in the DSDRS, and ? is a variable in the glue-logic language that indicates that the value of some constructor in the fully-specific logical form (in this case the value of a rhetorical-relation predicate symbol) is currently unknown:

- `Glue Logic Schema:` $\lambda :?(\alpha,\beta) \wedge \mathit{Info}(\alpha,\beta,\lambda)) > \lambda : R(\alpha,\beta,\lambda)$

In words, if β is to be connected to α with a rhetorical relation whose value we don't know yet, and the result is to appear in the scopal position of the DSDRS that's labelled λ, and moreover $\mathit{Info}(\alpha,\beta,\lambda)$ holds of the content labelled by λ, α and β, then normally the rhetorical relation is R. The conjunct $\mathit{Info}(\alpha,\beta,\lambda)$ is cashed out in terms of the ULFs that α, β and λ label, and the rules are justified either on the basis of underlying linguistic knowledge, world knowledge, or knowledge of the cognitive states of the dialogue participants. Thus glue logic axioms encapsulate default inferences about which types of speech act were performed, on the basis of the content and context of the utterances.

One default that we mentioned in Section 3 is that the necessary consequences of a speech act being performed are normally sufficient for inferring that it was performed. For instance, if α and β are rhetorically connected and the glue logic evaluates that they are semantically incompatible (glossed as $\mathit{CorrS}(\alpha,\beta)$), then normally they are connected with *Correction*:

- `Correction:` $(\lambda :?(\alpha,\beta) \wedge \mathit{CorrS}(\alpha,\beta)) > \lambda : \mathit{Correction}(\alpha,\beta)$

Assuming a standard semantics of the *it*-cleft and an alternative semantics to pitch accents (Rooth, 1992), the compositional semantics of $\pi_{2.1}$ in (11) conveys that Bill *as opposed to anyone else mentioned in the context* stole the funds.

(11) $\pi_{1.1}$. A: John went to jail.
$\pi_{1.2}$. A: He embezzled the pension funds.
$\pi_{2.1}$. B: No, it was BILL who stole the pension funds.
$\pi_{2.2}$. B: I was at the trial.
$\pi_{3.1}$. A: Oh, OK.
$\pi_{4.1}$. B: John did go to jail though.

If $\pi :?(\pi_{1.2},\pi_{2.1})$ holds, then SDRT's constraints on anaphoric interpretation (see definitions given shortly) mean that the funds in $\pi_{1.2}$ and $\pi_{2.1}$ co-refer. And so long as the incompatibility between this and the compositional (and lexical) semantics of $\pi_{1.2}$ and $\pi_{2.1}$ is transferred into the glue language, the antecedent to `Correction` will be satisfied, yielding $\pi : \mathit{Correction}(\pi_{1.2},\pi_{2.1})$. Further axioms for inferring *Correction* exploit explicit cue phrases such as *No* and *you're wrong*.

$\mathit{Explanation}(\alpha,\beta)$ can be inferred when there's evidence in the discourse that β causes α (written $\mathit{cause}_D(\beta,\alpha)$). Evidence of a causal relation does not entail an actual causal relation, but they are nonmonotonically linked thanks to the default rule `Explanation` below and the semantics of *Explanation* given earlier.

- `Explanation`: $(\lambda :?(\alpha,\beta) \wedge \mathit{cause}_D(\beta,\alpha)) > \lambda : \mathit{Explanation}(\alpha,\beta)$

Glue-logic axioms for inferring $\mathit{cause}_D(\beta,\alpha)$ are monotonic, for either the discourse contains evidence of a causal connection or it doesn't. For example, the axiom `Fight` stipulates that if x and y have a fight z and x didn't go out with y, then there is evidence in the discourse of the latter causing the former. As we'll shortly see, this axiom contributes to the construction of the DSDRS for dialogue (1):

- `Fight`:
  $(\alpha : (\mathit{have}(e_\alpha, X, z) \wedge \mathit{and\_c}(X, x, y) \wedge \mathit{fight}(z)) \wedge \beta : \mathit{neg}(\gamma) \wedge \gamma : \mathit{go\text{-}out}(e_\beta, x, y)) \rightarrow$
  $\mathit{cause}_D(\beta, \alpha)$

In SDRT the inferences for constructing logical form can flow in one of several directions. If the premises of glue logic axioms are satisfied by the ULFs derived from the grammar, and one can thus infer via the glue logic's consequence relation $\mid\!\sim$ a particular rhetorical relation, then this particular rhetorical connection becomes a part of the (updated) logical form. Further, the semantic consequences of this rhetorical relation may lead to inferences about how other underspecified conditions are resolved (e.g., identifying antecedents to pronouns). Alternatively, there are cases where compositional semantics is insufficient for satisfying the premises to any glue logic axioms. In this case, one can resolve the underspecified compositional semantics to specific values so as to satisfy antecedents to glue-logic axioms, leading in turn to a rhetorical relation being inferred. If one adopts this strategy, and moreover there is a choice of which way to resolve the underspecified content so as to infer a rhetorical relation from it, then one chooses an interpretation which *maximises the coherence* of the logical form (see Asher and Lascarides (2003) for details).

Definition 6, taken from Asher and Lascarides (2003), stipulates this general principle that one interprets discourse in a way that maximises its coherence, and illustrates the conservative assumptions that SDRT makes about what factors influence the degree of coherence. While the original version of `MDC` applied to SDRSs, it also applies now to DSDRSs.[12]

Definition 6 **Maximising Discourse Coherence** (`MDC`)

> Discourse is interpreted so as to maximise discourse coherence, where the (partial) ranking among interpretations is determined by the following principles:
>
> 1. All else being equal, the more rhetorical connections there are between two items in a discourse, the more coherent the interpretation.
> 2. All else being equal, the more semantically underspecified elements are resolved to specific values, the more coherent the interpretation. Moreover, resolutions that lead to $\mid\!\sim$-consequences for a particular rhetorical relation are preferred over resolutions that are logically unrelated to any rhetorical connection.
> 3. Some rhetorical relations are inherently scalar. For example, the quality of a *Narration* is dependent on the specificity of its common topic. All else

[12] SDRT's more formal definition of `MDC` (Asher and Lascarides, 2003, p233), which ranks SDRSs into a partial order, easily extends to rank DSDRSs: roughly put, one DSDRS K_1 is more coherent than another K_2 if (a) they are comparable (i.e., they consist of the same number of turns and the same dialogue participants); and (b) each SDRS in K_1 is at least as coherent as the SDRS for the same dialogue participant and turn in K_2.

being equal, an interpretation which maximises the quality of its rhetorical relations is more coherent than one that doesn't.

4. All else being equal, the number of labels in the semantic representation is minimal, so long as minimising the number of labels does not create semantic anomalies among the rhetorical relations in the representation.

The glue logic for DSDRSs involves constructing an SDRS for each turn and each participant. The axioms of SDRS-construction just described still apply, serving to provide values for the function F in a DSDRS for each label π. In addition, given the definition of DSDRSs, the glue logic must stipulate for every $\pi \in \Pi$ which SDRSs it is a member of. So we extend the glue logic axiom with a 3-place predicate symbol T: where $d \in D$ and $j \in [1, n]$, $T(d, j, \pi)$ means that the label π is a part of the SDRS $T^d(j)$. Thus the glue-language predicate symbol T is used to express statements about the function T in the DSDRSs it describes.

We need to add axioms to the glue logic that formally specify the principles of dialogue interpretation that we proposed (e.g., `Undenied Commitments`, `AC`, `DC`). But before doing this, we define discourse update and availability for DSDRSs. As in original SDRT, updating a representation of the discourse context with new content involves adding all the $\mid\!\sim$-consequences of the old and new content to the logical form. If there is underspecified information about which of the available labels the new content attaches to, then update is conservative, and generalises over all the possibilities (see the second part of Definition 7).

Definition 7 Discourse Update for DSDRSs

Simple Update. We first define how to update a context with new information β, given a particular available attachment site α.

The ULF-formula $\lambda : ?(\alpha, \beta) \wedge T(d, j, \lambda)$ specifies that the new information $\mathcal{K}_\beta$ is to be attached to the DSDRS as a part of the SDRS $T^d(j)$. Let σ be a set of (fully-specified) DSDRSs, and let $\mathit{Th}(\sigma)$ be the set of all ULFs that partially describe the DSDRSs in σ. Let ψ be either (a) a ULF $\mathcal{K}_\beta$, or (b) a formula $\lambda : ?(\alpha, \beta) \wedge T(d, j, \lambda)$ about attachment, where $\mathit{Th}(\sigma) \vdash_{\mathit{ulf}} \mathcal{K}_\beta$. Then $\sigma + \psi$ is a set of DSDRSs defined as follows:

1. $\sigma + \psi = \{\tau : \text{ if } \mathit{Th}(\sigma), \psi \mid\!\sim \phi \text{ then } \tau \vdash_{\mathit{ulf}} \phi\}$, provided the result is not $\emptyset$;
2. $\sigma + \psi = \sigma$ otherwise.

Discourse Update. Suppose that A is the set of available attachment points in the old information σ for the new information β. Then the power set $\mathcal{P}(A)$ represents all possible choices for what labels in σ the new label β is actually attached to. $\mathit{update}_{\text{SDRT}}$ is neutral about which member of $\mathcal{P}(A)$ is the 'right' choice, for $\mathit{update}_{\text{SDRT}}(\sigma, \mathcal{K}_\beta)$ is the *union* of DSDRSs that result from a sequence of +-operations for each member of $\mathcal{P}(A)$.

The updated ULF may not identify a *unique* logical form (i.e., $|\mathit{update}_{\text{SDRT}}(\sigma, \mathcal{K}_\beta)| > 1$). The Principle `MDC` then ranks the alternative, fully specific logical forms. However, in contrast to the analysis of disputes from Asher and Lascarides (2003), constructing a logical form always involves *extending* the logical form from the context, and never revising it. That is, $\mathit{Th}(\mathit{update}_{\text{SDRT}}(\sigma, \mathcal{K}_\beta)) \subseteq \mathit{Th}(\sigma)$. In essence, as a dialogue proceeds, we learn strictly more information about the content of (prior) utterances, never revising those prior interpretations but rather refining them. Of course, this monotonicity is feasible only because Discourse Update does not restrict the new information β to being the label of a single clause. It could label the content of an entire turn or more, and discourse update abstracts over all these possibilities. In short, we maintain monotonicity only by relaxing incremental interpretation. Any implementation of SDRT in a practical dialogue system would need to restrict the massive search space that ensues from non-incrementality, and the restricted 'beam search' may mean that revision processes in implementation become inevitable. But approximating this account of update in a dialogue system is a matter for future research; here we focus only a competence model of dialogue understanding.

Definition 8 stipulates that the available labels of an SDRS are the last label and all labels that are connected to it by a sequence of outscopes relations and/or subordinating rhetorical relations (ignoring the complexities we discussed earlier concerning discourse subordination).

Definition 8 The Original Definition of Availability for an SDRS

> Let $\langle \Pi, F, \mathit{last}\rangle$ be an SDRS. Furthermore, where $\pi_1, \pi_2 \in \Pi$, we say that $\pi_1 > \pi_2$ iff either: (i) $R(\pi_1, \pi_2)$ is within the range of F, where R is a subordinating relation (e.g., *Q-Elab*, *Plan-Elab*, *IQAP*, *Correction*, *Background*, *Elaboration*, *Explanation*); or (ii) π_1 immediately outscopes π_2 (i.e., $F(\pi_1)$ contains the literal π_2). Let $> *$ be the transitive closure of the relation $>$. Then the available labels $A \subseteq \Pi$ of the SDRS is:
>
> $$A = \{\pi \in \Pi : \pi \geq *\mathit{last}\}$$

The definition of availability for DSDRSs is defined in terms of that for SDRSs: they are all the available labels of its SDRSs for the last turn.

Definition 9 Definition of Availability for DSDRSs

> Let D be a set of discourse participants, and let $\langle n, T, \Pi, F, \mathit{last}\rangle$ be a DSDRS for D. Furthermore, where $d_i \in D$ and $j \in [1, n]$, let $A^{d_i}_j \subseteq \Pi^{d_i}_j$ be the set of the available labels for the SDRS $T^{d_i}(j)$, as defined in Definition 8. Then the set $A \subseteq \Pi$ of available labels for the DSDRS is defined as:
>
> $$A = \bigcup_{d_i \in D} A^{d_i}_n$$
>
> In other words, A is the union of all available labels from all the SDRSs for the last turn n (and so by Definitions 3 and 8 $\mathit{last} \in A$).

In the particular examples we have analysed so far, we have always assumed that the content of the current turn attaches to an available label from the SDRS of the unique speaker of the last turn. But Definition 9 allows a speaker to completely *ignore* what the last speaker said, and instead address content that was conveyed in a prior turn to the last one. While it might be rare to ignore someone in a two-person dialogue, it is more frequent in a multi-party conversation, especially if the participants have unequal power. A competence model of dialogue should reflect 'ignoring' moves in a transparent way. We have achieved this here: the hallmark that someone has ignored the prior speaker's turn is that his SDRS features no labels that were introduced in that turn. However, all else being equal, an interpretation where a speaker addresses the prior turn is arguably more coherent than one where he ignores it. So for the sake of simplicity, we will from now on focus on interpretations where the current speaker does not ignore the last speaker.

Let's now examine how to express axioms in the glue logic that capture the principles from Section 3 for computing which commitments persist from prior turns. The axiom `Non-Speakers` stipulates that normally you change your commitments only if you speak, where $speaker(d, j)$ means that dialogue participant d is the (unique) speaker of turn j:

- `Non-Speakers:` $\neg speaker(d, j) \rightarrow (T(d, j-1, \alpha) > T(d, j, \alpha))$

The axiom is a default because silence can be a meaningful act in sufficiently specific contexts (Grice, 1975). The labels of utterances that are spoken by a dialogue participant d in a given turn j must, on the other hand, be a part of $T^d(j)$. So, if $partof(\alpha, j)$ means that α labels the content of an *individual clause* that was said in turn j, then the following axiom holds:

- `Speakers:` $(partof(\alpha, j) \wedge speaker(d, j)) \rightarrow T(d, j, \alpha)$

The `Persistence Principle` and `Undenied Commitments` for simple-left veridical relations from Section 3 have a straightforward formalisation in the glue logic:

- `The Persistence Principle:` $\lambda : R(\alpha, \beta) \rightarrow \lambda : \textit{Undenied-Commitments}(\alpha)$
- `Undenied Commitments` for Simple Left Veridical Relations:
  $(\lambda : R(\alpha, \beta) \wedge T(d_1, j, \lambda) \wedge \textit{simple-left-veridical}(R) \wedge \lambda' : R'(\gamma, \alpha) \wedge T(d_2, j-1, \lambda')) >$
  $(\lambda : \textit{Undenied-Commitments}(\alpha) \rightarrow \lambda : R'(\gamma, \alpha))$

There is a similar rule to `Undenied Commitments` for the case where $\lambda' : R'(\alpha, \gamma)$.

We can now use the glue logic to derive the DSDRS of dialogue (1), shown in Table 1 (where M is Mark, K is Karen and S is Sharon). The axiom `Non-Speaker` makes $T^K(1)$ and $T^S(1)$ empty. But the axiom `Speaker` means $\{\pi_{1.1}, \pi_{1.2}\} \subseteq T^M(1)$. By `MDC`, we prefer to minimise labels and maximise rhetorical connections. Given the definition of availability, this means that an SDRS $T^M(1)$ that satisfies $\pi_{1M} :?(\pi_{1.1}, \pi_{1.2})$ is preferred to any SDRS that does not satisfy this. This assumption about attachment resolves the pronoun *she* in $\pi_{1.2}$ to Karen, the only accessible antecedent within $\pi_{1.1}$. Therefore, given the ULFs of $\pi_{1.1}$ and $\pi_{1.2}$ derived from the grammar and this resolution of *she*, the antecedent `Fight` is satisfied, yielding $cause_D(\pi_{1.2}, \pi_{1.1})$ (i.e., there is evidence in the discourse that $\pi_{1.2}$ caused $\pi_{1.1}$). So the antecedent to `Explanation` is satisfied, yielding $\pi_{1M} : \textit{Explanation}(\pi_{1.1}, \pi_{1.2})$.

Now consider the interpretation of the second turn. `Non-Speaker` makes $T^S(2) = T^S(1)$ and $T^M(2) = T^M(1)$. But `Speaker` means that $\pi_{2.1} \in T^K(2)$. K's second turn can be interpreted as ignoring the first, but by MDC relating it to M's turn is preferred. M's SDRS has three available labels: $\pi_{1.1}$, $\pi_{2.1}$ and π_{1M}. And so we consider updates with all combinations of these labels; MDC will tell us which of these is preferred. Let's suppose that $\pi_{2.1}$ attaches to π_{1M}, but not to $\pi_{1.1}$ or $\pi_{1.2}$. Then the glue logic fails to validate any rhetorical connection (e.g., we argued in Section 2 that an *Explanation* connection would be implausible) or at the very least, the rhetorical relation would be of inferior quality to the one that's inferable between $\pi_{2.1}$ and $\pi_{1.2}$, and hence dispreferred by MDC. A connection between $\pi_{2.1}$ and $\pi_{1.1}$ but not $\pi_{1.2}$ suffers a similar fate. On the other hand, a glue logic axiom that is similar in style to Fight and that encapsulates the world knowledge about Mark not asking Karen out and Karen not going out with Mark should validate an inference that there is evidence in the discourse that the former caused the latter. And so via Explanation one can infer $?_\pi : \mathit{Explanation}(\pi_{1.2}, \pi_{2.1})$ is a part of $T^K(2)$ for some label $?_\pi$ that is also a part of $T^K(2)$.

This together with $T^M(1)$ validates the antecedent to Undenied Commitments, and so its non-monotonic consequence is inferred (since it is consistent with the premises), leading by Modus Ponens on the Persistence Principle to the conclusion $?_\pi : \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$. Finally, MDC makes the SDRS minimal, resolving $?_\pi$ to the root label π_{2K}. And so K's SDRS for the third turn is as shown in Table 1.

Now let's consider the principles from Section 3 for interpreting explicit endorsements and challenges. We suggested in Section 3 that whenever $\lambda : \mathit{Correction}(\alpha, \beta)$ forms a part of an SDRS, then $\lambda : \mathit{Correction}(\gamma, \beta)$ should be a part of it too for all labels γ that outscope α. In fact, this principle follows from the semantics of *Correction*, Correction and MDC. That's because the (dynamic) interpretation of $\mathit{Correction}(\alpha, \beta)$ entails $\mathit{Correction}(\gamma, \beta)$. Assuming this entailment is transferred (in shallow form) to the glue logic, then this together with the assumption $\lambda :?(\gamma, \beta)$ would satisfy the antecedent to Correction, yielding $\lambda : \mathit{Correction}(\gamma, \beta)$. And according to MDC, an interpretation where one assumes $\lambda :?(\gamma, \beta)$ is more coherent than one where γ and β aren't related at all, for it yields more rhetorical connections (note that γ is available because γ outscopes α).

Let's return to the analysis of (11), starting with the SDRSs for the first turn. B's SDRS, according to Non-Speakers is $\emptyset$. As before, availability, Speakers and MDC means that $\pi_{1A} : ?(\pi_{1.1}, \pi_{1.2})$ holds. This makes John the only possible antecedent to *He* in $\pi_{1.2}$. So by MDC, the pronoun is resolved this way whatever the rhetorical relation. Using rules that encapsulate relevant world knowledge, one will infer from these premises that $\mathit{cause}_D(\pi_2, \pi_1)$ and hence $\pi_{1A} : \mathit{Explanation}(\pi_{1.1}, \pi_{1.2})$. By Non-Speakers, $T^A(2) = T^A(1)$. By Speakers, $\{\pi_{2.1}, \pi_{2.2}\} \subseteq T^B(2)$. There are many options for how B's SDRS may be updated with $\pi_{2.1}$ and $\pi_{2.2}$, and according to Definition 7 they must all be considered and then ranked by MDC. First, one could form a segment out of $\pi_{2.1}$ and $\pi_{2.2}$ and attach the result to an available label; by MDC attaching to a label from A's prior SDRS is preferred, and so the choices are to attach to one or more of π_{1A}, $\pi_{1.1}$ and $\pi_{1.2}$. Alternatively, one could attach $\pi_{2.1}$ to one of these labels first, and then attach $\pi_{2.2}$ to an available label in the result.

If we were to form a segment out of $\pi_{2.1}$ and $\pi_{2.2}$ first, and then attach the (root) label of that segment to the context, then as we'll see shortly this would create an extra label as compared with the strategy for updating 'clause by clause', and so it's dispreferred by MDC. So let's consider now the option of attaching $\pi_{2.1}$ to an available label within $T^A(1)$. We saw earlier that if we assume that $\pi_{2.1}$ attaches to $\pi_{1.2}$, then Correction validates an inference to $?_\pi$: *Correction*$(\pi_{1.2}, \pi_{2.1})$. This also yields from an assumption that $\pi_{2.1}$ attaches to π_{1A} an inference to $?_\pi$: *Correction*$(\pi_{1A}, \pi_{2.1})$. So attaching $\pi_{2.1}$ to $\pi_{1.2}$ and π_{1A} is preferred by `MDC` to not attaching to them, since this maximises the number of rhetorical connections. Furthermore, if $\pi_{2.1}$ attaches to $\pi_{1.1}$, then no glue logic axioms validate an inference about the identity of the rhetorical connection between them. So, by clause 2 of `MDC`—one prefers interpretations with fewer unknown values—an interpretation where $\pi_{2.1}$ attaches to $\pi_{1.1}$ is dispreferred. Given that $\pi_{2.1}$ is interpreted as a correction, relevant glue logic axioms will validate that $\pi_{2.2}$ explains this corrective move. Finally, by `MDC` preferring a minimum number of labels, $?_\pi$ resolves to the root label π_{2B}. And so the final SDRS $T^B(2)$ is as shown in Table 5.

A's SDRS for the third turn of (11) has a linguistic form that entails that it is either an *Acceptance* of some available prior content, or an acknowledgement of understanding (which is represented in SDRT with the metatalk relation *Acknowledgement**). We assume a glue-logic axiom that makes *OK* default to the more specific form of grounding, namely *Acceptance*; this means that although the metatalk relation *Acknowledgement** is then inferable from this (on the basis that the necessary semantic consequences of a speech act are normally sufficient for inferring it has been performed), the relation isn't added to the logical form because it is semantically and structurally redundant to do so.

There is, however, an ambiguity in A's utterance *OK*: its highly underspecified compositional semantics doesn't determine the first argument to the *Acceptance* relation. Once again using the principle that we prefer interpretations where the current turn relates to the previous one, we prefer the first argument of *Acceptance* to be one or more of $\pi_{2.1}$, $\pi_{2.2}$ and π_{2B}. We argued in Section 3 that there appears to be a tendency to interpret explicit endorsements with highly underdetermined compositional semantic content—like *OK*—so that they have the widest scope that is consistent with the premises. This can easily be expressed in the glue logic as a default axiom about attachment:

- `Wide Scope OK:`
  $(?_\lambda : \mathit{Acceptance}(?_\alpha, \beta) \wedge \beta : \mathit{OK} \wedge T(d_1, i-1, ?_\alpha) \wedge T(d_2, i, \beta)) >$
  $(T(d_1, i-1, \gamma) \rightarrow ?_\alpha \succeq \gamma))$

In words, if β has the form *OK* and is interpreted as an acceptance act of some part of a different agent's prior turn, then normally, that acceptance act is of the root label of that turn. So in (11), π_{3A} : *Acceptance*$(\pi_{2B}, \pi_{3.1})$ is inferred.

This analysis is not yet complete, however, because we need to apply `AC` from Section 3.1 (page 18). `AC` involves computing the undenied content of the corrected material. So we start by adding glue logic axioms that recursively compute undenied content. We define a function *sc* (standing for Subordinated Correction), which for any label π that is corrected by π' yields the set of labels outscoped by π that (a) are also corrected by π', and (b) of those labels that satisfy (a) they also have 'highest' scopal position within π. More formally:

$$\begin{array}{ll} & \text{Where } \mathit{Correction}(\pi, \pi'), \\ \mathit{sc}_{\pi'}(\pi) = & \{\pi'' \text{ st } \mathit{Correction}(\pi'', \pi'), \pi \succ \pi'' \text{ and} \\ & \quad \forall \pi''' \text{ st } \pi \succ \pi''' \succ \pi'', \neg \mathit{Correction}(\pi''', \pi')\} \end{array}$$

We also introduce a function *Last* from labels to labels: *Last*(π) is the (unique) label π' that is the last label in the SDRS-formula K_π if there is such a label, otherwise, if $K_\pi \in \mathcal{L}_{basic}$, then *Last*$(\pi) = \pi$.

We then introduce into the glue logic a function *Undenied* from labels to formulae, that computes for a label π that is corrected by π' the undenied part of π (relative to that correction by π'). $\mathit{Undenied}_{\pi'}(\pi)$ is constructed via substitution and computing it is decidable (unlike downdating and revision in first order logics). We write $K[\phi/\phi']$ to mean that all occurrences of ϕ within the formula K are substituted with ϕ'. We also 'overload' this notation: where $\Pi = \{\pi_1, \ldots, \pi_n\}$, $K[K_\Pi/K'_\Pi]$ means that each occurrence of K_{π_i} in K is substituted with K'_{π_i}, for $1 \leq i \leq n$. The function *Undenied* is defined as follows, and matches the informal recursive definition we described in Section 3.1.

- `Undenied:`
  If $\mathit{Correction}(\pi, \pi')$, then:

  1. If $sc_{\pi'}(\pi) \neq \emptyset$, then $\mathit{Undenied}_{\pi'}(\pi) = K_\pi[R(\pi''', sc_{\pi'}(\pi))/V(\pi''', sc_{\pi'}(\pi)^b)]$ where for each $\pi'' \in sc_{\pi'}(\pi)$, $F(\pi''^b) = \mathit{Undenied}_{\pi'}(\pi'')$.
  2. If $sc_{\pi'}(\pi) = \emptyset$, then $\mathit{Undenied}_{\pi'}(\pi) = \exists\text{-}\mathit{closure\ of}\ \zeta(\pi'^{b,\lambda})$.

The glue-logic axiom which expresses `AC` is given below.

- `Undenied Commitments when Accepting Corrections (AC)`
  $(\lambda : R(\beta, \gamma) \wedge \mathit{left\text{-}veridical}(R) \wedge T(d_1, j, \lambda) \wedge$
  $\quad \lambda' : \mathit{Correction}(\alpha, \beta) \wedge T(d_2, j-1, \lambda') \wedge T(d_1, j-1, \alpha) \wedge (d_1 \neq d_2)) >$
  $\qquad (\lambda : \mathit{Undenied\text{-}Commitments}(\beta) \rightarrow \lambda : (\mathit{Undenied}_\beta(\alpha) \wedge \mathit{Background}(\beta, \mathit{Last}(\alpha))))$

This affects the construction of the logical form for (11) in the desired way, yielding the SDRSs shown in Table 5.

`Denying Corrections` (`DC`) ensures that the content and the availability of the commitments prior to the dispute are normally preserved when the correction is itself in dispute:

- `Undenied Commitments for Denying Corrections (DC):`
  $(\lambda_1 : \mathit{Correction}(\gamma, \alpha) \wedge T(d_1, j, \lambda_1) \wedge \lambda_2 : \mathit{Correction}(\alpha, \beta) \wedge T(d_2, j, \lambda_2) \wedge d_1 \neq d_2) >$
  $\qquad (\lambda_2 : \mathit{Undenied\text{-}Commitments}(\alpha) \rightarrow \lambda_2 : (K_\gamma \wedge V(\mathit{Last}(\gamma), \beta)))$

The axiom `DC` applies when constructing A's SDRS for the fourth turn in (20), and ensures that $\mathit{QAP}(\pi_{1.1}, \pi_{2.1}) \wedge V(\pi_{2.1}, \pi_{3.1})$ is added to the root label, as shown in Table 8.

6 Conclusion

We have presented a novel treatment of agreement and disputes, which captures facts about implicit agreement and also about what's agreed upon when a dispute has taken place. We argued that any adequate account of agreement and disputes must rest on a logical form for dialogue that tracks the public commitments of each agent, including their commitments to rhetorical relations. This ensures that an agent's commitments are not only to the compositional semantics of his utterances, but also to their illocutionary effects. By committing agents to this illocutionary content, a definition of agreement as shared public commitment captures how implicit agreement is dependent on the particular relational speech acts that each agent performs and the semantic relationships among these speech acts. The analysis of disputes also benefits from a logical form that distinguishes among each agent's commitments. It ensures that if one agent commits to the negation of another agent's commitments, then since the interpretation of logical form is a product of the interpretation of each agent's commitments, the overall dialogue remains consistent.

We argued that any adequate theory of agreement and disputes needs to include axioms that determine which prior commitments persist. We provided logically precise persistence axioms within the glue logic of SDRT, and demonstrated by example that they capture facts about agreement and disputes. These axioms reflect the logical co-dependence among the interpretations of corrective moves vs. endorsing moves that are performed in dialogue. Overall, the persistence axioms followed a general principle that dialogue interpretation maximise each dialogue participant's commitments to prior commitments (even if those commitments belonged to another agent), proviso this is consistent with the illocutionary effect of his current utterance. This is similar to the effects of performing downdating and revision on prior commitments when adding new commitments to them (although downdating and revision don't typically effect a transition of prior commitments to another agent). But while downdating and revision are an unsolved problem for first order languages and certainly uncomputable, our model of dialogue interpretation is computable and precise.

The relationship between what is agreed upon (or grounded) and the interpretation of the dialogue is entirely transparent, since it is defined in terms of the model theory of the logical forms. This, together with the logic for constructing logical form, provides a logical basis for exploring Clark's (1992) notion of *positive evidence* for grounding, endowing some of his claims with predictive power through logical reasoning.

This paper presents just some first steps towards a dynamic theory of grounding. It did not provide a detailed analysis of how questions and imperatives affect commitments, agreements and disputes. This will be addressed in Lascarides and Asher (forthcoming). We also wish to rationally reconstruct SDRT's logic of cognitive modelling, to take into account the model of dialogue interpretation presented here. This involves linking public commitments to inferences about other attitudes, such as beliefs, preferences, and intentions. For instance, we have ignored in our analysis of (1) the fact that Mark might be using his utterances in an accusatory manner, to attribute the blame for the fight to Karen, while Karen attempts to re-assign the blame to Mark. This is a matter of ongoing research, with some initial results reported in Asher and Lascarides (2008). Finally, we need to extend this account to a model of grounding at the lower levels (e.g., grounding an understanding of what was said). In future work we intend to incorporate insights from prior models of grounding (e.g., Ginzburg (2008)) into the SDRT model of agreement presented here.

Authors' Addresses

Alex Lascarides,	Nicholas Asher,
School of Informatics,	IRIT,
University of Edinburgh,	Université Paul Sabatier,
10, Crichton Street,	118, Route de Narbonne,
Edinburgh, EH8 9AB,	F-31062 Toulouse,
Scotland, UK.	France.
alex@inf.ed.ac.uk	asher@irit.fr

Acknowledgements. This work has benefited from comments and feedback from Julia Hirschberg, David Israel, Paul Portner, Colin Matheson, Ron Petrick, Mark Steedman, Matthew Stone, David Traum, Lyn Walker, from the participants of Sigdial 2008, Londial 2008 and Sub13 and from two anonymous reviewers for this journal. Any errors that remain are our own.

References

P. Achinstein. *The Nature of Explanation.* Oxford University Press, 1980.

N. Asher. *Reference to Abstract Objects in Discourse.* Kluwer Academic Publishers, 1993.

N. Asher. From discourse micro-structure to macro-structure and back again: The interpretation of focus. In H. Kamp and B. Partee, editors, *Context-Dependence in the Analysis of Linguistic Meaning.* Elsevier, 2004.

N. Asher and A. Lascarides. *Logics of Conversation.* Cambridge University Press, 2003.

N. Asher and A. Lascarides. Commitments, beliefs and intentions in dialogue. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (Londial)*, pages 35–42, London, 2008.

S. Bromberger. An approach to explanation. In R. Butler, editor, *Analytical Philsophy*, pages 72–105. Oxford University Press, 1962.

H. Clark. *Arenas of Language Use.* University of Chicago Press, Chicago, 1992.

H. Clark. *Using Language.* Cambridge University Press, Cambridge, England, 1996.

H. Clark and E.F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.

M. Egg, A. Koller, and J. Niehren. The constraint language for lambda structures. *Journal of Logic, Language, and Information*, 10:457–485, 2001.

T. Fernando. Bisimulations and predicate logic. *Journal of Symbolic Logic*, 59(3):924–944, 1994.

B. Gaudou, A. Herzig, and D. Longin. Grounding and the expression of belief. In *Proceedings of the 10th International conference on Principles of Knowledge Represetnation and Reasoning (KR'06)*, pages 221–229, Riva de Garda, Italy, 2006.

J. Ginzburg. *Semantics and Conversation.* CSLI Publications, 2008.

H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Synax and Semantics Volume 3: Speech Acts*, pages 41–58. Academic Press, 1975.

J. Groenendijk. Questions and answers: Semantics and logic. In *Proceedings of the 2nd CologNET-ElsET Symposium. Questions and Answers: Theoretical and Applied Perspectives*, pages 16–23, 2003.

J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14: 39–100, 1991.

C. Hamblin. *Fallacies.* Metheun, 1970.

J. Hirschberg. *A Theory of Scalar Implicature.* PhD thesis, Computer and Information Science, University of Pennsylvania, 1985.

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

M. Krifka. A compositional semantics for multiple focus constructions. In Steven Moore and Adam Zachary Wyner, editors, *Proceedings from Semantics and Linguistic Theory I*, pages 127–158, Ithaca, New York, 1991. Cornell University. Available from CLC Publications, Department of Linguistics, Morrill Hall, Cornell University, Ithaca, NY 14853-4701.

S. Larsson and D. Traum. Information state and dialogue management in the trindi dialogue move engine toolkit. *Natural Language Engineering*, 6(3–4):323–340, 2000.

A. Lascarides and N. Asher. Agreement and disputes in dialogue. In *Proceedings of the 9th SigDial Workshop on Discourse and Dialogue (SIGDIAL)*, pages 29–36, 2008.

A. Lascarides and N. Asher. The interpretation of questions in dialogue. In A. Riester and T. Solstad, editors, *Proceedings of Sinn Und Bedeutung 13*, forthcoming.

C. Matheson, M. Poesio, and D. Traum. Modelling grounding and discourse obligations using update rules. In *Proceedings of the first Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 2–9, 2000.

M. Poesio and D. Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3), 1997.

M. Poesio and D. Traum. Towards an axiomatisation of dialogue acts. In J. Hulstijn and A. Nijholt, editors, *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogue.* 1998.

L. Polanyi. A theory of discourse structure and discourse coherence. In P. D. Kroeber W. H. Eilfort and K. L. Peterson, editors, *Papers from the General Session at the 21st Regional Meeting of the Chicago Linguistics Society.* 1985.

M. Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116, 1992.

H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4):696–735, 1974.

M. Steedman. *The Syntactic Process.* MIT Press, 2000.

D. Traum. *A Computational Theory of Grounding in Natural Language Conversation.* PhD thesis, Computer Science Department, University of Rochester, 1994.

R. van der Sandt. Presuppositional denials. In *Proceedings of the 5th International Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog)*, pages 107–128, Bielefeld, 2001.

J. van Eijk and H. Kamp. Representing discourse in context. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Linguistics*, pages 179–237. Elsevier, 1997.

N. van Leusen. The interpretation of correction. In P. Bosch and R. van der Sandt, editors, *Focus and Natural Language Processing.* Cambridge University Press, 1994.

W. Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation.* Springer, 2000.

M. Walker. Inferring acceptance and rejection in dialogue by default rules of inference. *Language and Speech*, 39(2), 1996.