On the Role of Relations and Structure in Discourse Interpretation

Julie Hunter and Kate Thompson

LINAGORA Labs

December 15, 2020

We communicate for a variety of reasons, be it to exchange information, to persuade someone of a certain point of view, or simply to entertain each other. In each case, achieving our goals requires linking together the contents of multiple *discourse units*, which include the contents of individual speech acts and, for conversations situated in a shared visual environment, the contents contributed by physical gestures and other nonlinguistic events. A fundamental insight that guides work on discourse and dialogue interpretation is that *the way in which* discourse units are related to one another within the context of a conversation is essential to the conversation's meaning.[1]

To develop an intuition of what we mean when we talk about relations between discourse units, we will start with a simple example. Suppose a friend says to you:

(1)   I need my hat back. I'm leaving for São Paulo in two days.

Your friend's utterance contains two sentences, each of which contributes a single *discourse unit*, understood here roughly as a single proposition.[2] Although there is no lexical or syntactic indication that the two discourse units are related, given that they are uttered together you will automatically try to find a connection between them that explains what your friend's needing their hat back has to do with their leaving in two days. Perhaps the most reasonable explanation is that they want to wear the hat in São Paulo, in which case your interpretation will be roughly the same as if they had used an explicit discourse marker as in (2):

(2)   I need my hat back **because** I'm leaving for São Paulo in two days.

Interpreting (1) along the lines of (2) places the discourse units in a semantic relation of explanation, but there are other types of relations you could infer between the same discourse units given different contextual factors. For instance, a change in intonation between the two moves might signal that (1) merely conveys a list of your friend's thoughts, linked by a conjunction relation, and would be roughly equivalent to:

(3)   I need my hat back. **Also**, I'm leaving for São Paulo in two days.

Semantic relations can also be inferred between the contents of discourse units made by different speakers. Questions and answers are primary examples of this.

---

[2] In this paper, we treat discourse units as roughly clause level contents. However, we note that some clauses might contain more than one discourse unit, as illustrated by Hobbs (1985)'s example *A car hit a jogger in Palo Alto last night*, in which the deverbal noun supplies an event description and we infer a temporal sequence between the event of jogging and the event of being hit by a car.

(4)   a.   What are you going to do downtown?
          b.   I'm going to the bookstore.

It is a basic assumption of conversational exchange that if someone is asked a question they will answer it promptly, if not directly, in the next discourse move, as seen in (4). However, just as with the discourse units in (1), nothing about the content or surface form of (4-b) when considered in isolation indicates that it is an answer to any question, let alone (4-a). The question-answer relation between the two units is inferred from their content plus the assumption that people tend to promptly answer questions when asked.

The central role of inference in interpreting question-answer relations between discourse units becomes more apparent if we imagine more moves intervening between a question and its answer:

(5)   a.   What are you going to do downtown?
          b.   Ugh, I'm so mad! My brother lost my copy of *The Watchmen*, and I need to reread it for class. I'm going to the bookstore.

The answer to the question asked in (5-a) is the same as the answer provided in (4-b), but in (5), the speaker first provides unsolicited background information before giving the answer. Despite this detour, the speaker who asked (5-a), or someone just listening in on the conversation, would be able to identify this answer by reasoning about the content of each intervening discourse unit expressed after the question.[3]

In order to provide a systematic account of the inferences needed for the interpretation of discourse, such as those seen in the foregoing examples, formal approaches to discourse structure and interpretation such as Rhetorical Structure Theory (RST; Mann and Thompson, 1987) and Segmented Discourse Representation Theory (SDRT; Asher, 1993; Asher and Lascarides, 2003), building on the work of Hobbs (1979, 1985), incorporate semantic relations, called *discourse relations*, into their models of discourse.[4] They focus on determining the variety of relations that can connect discourse units, e.g. Explanation or Question-Answer Pair, and the kinds of information – semantic, discursive or otherwise – that speakers use to determine them. Discourse relations are considered to contribute truth-conditional content above and beyond that conveyed by the collection of discourse moves alone, which has an effect on the logical form of a discourse. Moreover, going beyond single relation instances, theories of discourse structure seek to identify the structural constraints on discourse development which limit how discourse structures built from multiple discourse units can evolve as a discourse proceeds, and to describe the nature and interpretation of full discourse structures.

In the discussion that follows, we start by taking a look in Section 1 at why discourse relations are important for philosophy and linguistics and by situating theories of discourse structure in the larger field of dynamic semantics. In Section 2, we show how recent work on discourse relations and structure has been used to model phenomena along the semantics-pragmatics interface, to analyze multi-modal discourse, and to provide an account of discourse goals and the interaction of bias and discourse interpretation. We conclude in Section 3 with a set of open questions for theories of discourse structure and their role in semantics and pragmatics debates as well as a discussion of what kinds of tools will be most helpful for answering these questions.

---

[3]The observation that we make sense of natural language discourse by making inferences about how discrete utterances might cohere is foundational to the study of discourse relations. Notably, Jerry Hobbs suggests that abduction, or inference to the best explanation, underlies the process of speaker interpretation (Hobbs, 1979).

[4]Other terms for discourse relations include *coherence relations* or *rhetorical relations*.

1 The semantic effects of discourse structure

A simple sentence consisting of a single clause is the minimal tool for conveying a description of the world.[5] In modern philosophy and linguistics, specifically in truth-conditional semantics, the meaning of a clause is modelled as a proposition, which is often defined as the set of possible worlds in which the state of affairs, or eventuality, described by the sentence holds. Propositions have held the interest of philosophers and linguists because they are the minimal bearers of truth or falsity, allowing us to exchange information and learn new things about the world. From Plato and Aristotle up through modern day model-theoretic accounts of linguistic meaning stemming from the work of Frege and Russell, simple sentences, and the propositions they express, have been the primary units of study in semantics and philosophy of language; likewise they are the starting point for discourse-based language modelling.

It is clear that a discourse or conversation proffers a more complex representation of the world than does a simple sentence. Discourse units combine recursively to create more and more complex semantic structures, giving rise to both "bottom up" and "top down" effects. This section introduces some of these effects, starting with those that arise with individual discourse relations and finishing with effects related to complex discourse structures built from more than two discourse units.[6]

1.1 Discourse relations

A discourse will generally support inferences which would not be entailed by the set of its constituent propositions alone,[7] allowing us to communicate, and infer, more complicated messages. Let's consider again our introductory examples (1)-(3), about the hat and leaving for São Paulo in two days. Inferring the relation Explanation between the two discourse units of (1), as made explicit in (2), will entail that the speaker needs the hat back in less than two days. This inference, however, cannot be attributed solely to the discourse units involved, for it is not supported if we infer a different relation between them: if (1) were interpreted along the lines of (3), the inference would not be supported and the speaker could reasonably continue (3) with "I can pick up the hat once I get back from Brazil."

To some extent, we can say the same thing about complex propositions formed from Boolean operators: while $p \wedge q$ and $p \vee q$ are built up from the same constituent propositions, they do not support the same set of entailments because they do not involve the same relation. Only $p \wedge q$ entails the set $\{p,q\}$, for example. But it is important to note that, whereas the truth value of a complex formula formed by a Boolean operator is determined entirely by the truth values of its arguments, the truth value of an instance of a discourse relation cannot be so reduced: if two propositions, p and q, are both true, this automatically entails the truth of $p \wedge q$, but not that of Explanation(p,q). The Explanation relation adds additional semantic content that is itself truth-evaluable. Another crucial difference is that discourse relations can add substantial semantic content to discourse even in the absence of an explicit relation marker, as illustrated by (1). In this case, content is added to the discourse during composition as the result of a reasoning process, not by a logical operation determined by the semantics of a particular operator. In this way, the bottom up effects of discourse relations go beyond those observed in standard truth conditional semantics.

---

[5]We leave the question of sentence fragments to the side in this paper.

[6]Other articles in this volume highlight the importance of discourse relations and structure in contexts not covered in this article. See Márta Abrusán's entry on presupposition, Pranav Anand and Maziar Toosarvandani's entry on narrative discourse, Carlotta Pavese's entry on arguments, and Julian Schlöder's entry on rejection.

[7]While we will often equate a discourse unit with a proposition for simplicity, it is more precise to say that a discourse unit is an *instance* of a proposition relativized to a particular discourse context (Asher, 1993). If we took a discourse relation to hold between two propositions, then this discourse relation would hold between the set of possible worlds that each proposition represents, leading to overly strong claims. If a speaker says "John fell because he tripped on a rock," for example, we do not want the speaker to be committed to the claim that in every world in which John trips on a rock, he falls as a result of tripping on the rock. Rather, the speaker most likely just means that John fell at this particular place and time (in the discourse context) as a result of his having tripped on a rock at this particular place and time.

When discourse is modeled using discourse relations, a variety of "top-down" effects are also observed. That is, information the interpreter has about how discourse units are related – which might come from explicit discourse markers or other lexical information from the discourse units, or even non-lexical contextual information – can be used to interpret the content of *an individual discourse unit*. In particular, discourse relations have significant effects on various anaphoric phenomena, including the resolution of anaphoric pronouns and the temporal interpretation of individual clauses. This is part of what makes discourse relations so important to philosophy, particularly the philosophy of language, and linguistics: not only do they provide a framework which allows a more complete representation of the complex semantic structures that mediate nearly all human information exchange and knowledge acquisition, but they do so by offering a new perspective on problems, such as anaphora resolution, that have been discussed in the literature for a long time.

As an illustration of top down effects, consider the minimal pair in (6):

(6) a. Andy's bike broke down this morning. He showed up late for work.
     b. Andy showed up late for work. His bike broke down this morning

The past tense employed in (6-a) indicates that the events of Andy's bike breaking down and his showing up late for work happened in the past of the utterance time, but the example as a whole suggests more information about the timing of the individual events. In particular, we infer that the time at which Andy's bike broke down was in the past of the time at which he showed up late for work. Surely the order in which the events are described plays a role in this interpretation, but, as noted already by Ancient rhetoricians (e.g., Quintilian (1963)), this cannot be the whole story. If we reverse the arguments, as in (6-b), our tendency to understand Andy's bike troubles as the cause of his tardiness—an interpretation motivated by world knowledge—leads us to understand the event described second as actually having occurred first (Lascarides and Asher (1993)). From the perspective of SDRT, these observations are explained by noting that we infer different discourse relations in (6-a) and (6-b), namely Result and Explanation, respectively. The semantics of these relations then entail the differing temporal interpretations: Result requires that the event described by its second argument occur after that described by its first, whereas Explanation requires the opposite structure.[8]

Similarly, there are situations in which pronoun resolution is most effectively explained by appealing to discourse relations and world knowledge, as illustrated by (7), taken from Kehler et al. (2008) and adapted from Winograd (1972) (cf. also Hobbs, 1979; Kehler, 2002).

(7) The city council denied the demonstrators a permit because...
     a. ... they feared violence.
     b. ... they advocated violence.

The pronoun *they* is understood as referring to the city council in (7-a), but to the demonstrators in (7-b). Arguably, this is because world knowledge suggests that fearing violence is a good reason for an agent to reject a permit, while advocating violence is a good reason to have one's request rejected.

The role of world knowledge and reasoning comes out perhaps even more clearly if we consider an ambiguous discourse marker, such as *and*, which in (8) could support either a Parallel relation or a Result relation, leading to different interpretations of *they*:

---

[8]That the temporal interpretation of a discourse depends on conversational purposes and world knowledge echoes claims in Grice (1981), as does the fact that discourse relations that are not explicitly marked are often cancellable and reinforceable. Theories of discourse structure like SDRT go beyond Grice in numerous ways, however. First, they attempt to explain how tense and aspect, lexical choice, and update order influence discourse interpretation—the point is not that such information is unimportant, but that its role in the interpretation of anaphora passes via the determination of discourse relations. Second, they do not treat unambiguous discourse markers, such as *therefore*, as implicatures of any sort. Finally, more recent versions of SDRT do not presuppose cooperativity in conversation (see Section 2.2).

(8)   The city council denied the demonstrators a permit and they advocated violence.

The context of (8) and the world view of the speaker will be what tips the balance in favor of one inter-pretation or the other. If it is understood that demonstrators would likely react violently to authoritarian obstacles, the Result relation would best support this reading of (8), where *they* are the demonstrators. But in a context in which one accepts that a body of government might advocate violence against a group of people who are wanting to protest a cause, (8) could equally express a Parallel relation which would make an interpretation of *they* as picking out the city council more accessible than one in which *they* picks out the demonstrators. The important point here is that the sentence (8) itself supports two different inter-pretations of the pronoun *they*, and the choice of interpretation is accounted for by the type of semantic relation inferred.

The foregoing analysis of anaphora resolution and temporal interpretation generalizes insights from dynamic semantics. In dynamic semantics, models of pronominal anaphora take into account the order in which two clauses are added to the discourse context in order to capture the fact that, for instance, revers-ing the sentences in (6-a) would lead to a less felicitous discourse (Kamp and Reyle, 2013). Temporal interpretation can likewise be sensitive to update order and also to tense and aspect: were we to change the aspect in the second sentence of (6-a) to the past perfect, this would change the inferred temporal relation between the clauses (Kamp (1988)). Work on discourse relations, in particular Hobbs (1979); Kehler (2002); Asher (1993) and Asher and Lascarides (2003), incorporates the idea that a model of anaphoric phenomena must take into account the way in which the utterance content is linked to other contents in the incoming information state and the way in which discourse units are described.[9] In these accounts, however, update order and tense and aspect influence interpretation only *indirectly* by helping an interpreter determine what discourse relation is at work. Anaphora and temporal interpretation are thus understood as byproducts of reasoning about discourse relations (and, as we will see in the next subsection, discourse structure).[10]

1.2   Discourse structures

So far we have discussed the role of discourse relations in the computation of temporal structure and the resolution of anaphora using examples that contain pairs of discourse units. However, there are some other important anaphoric facts, such as propositional anaphora, which cannot be explained by considering pairs alone. Consider this example from Asher (1993):

(9)   a.   One plaintiff was passed over for promotion three times.
       b.   Another didn't get a raise for five years.
       c.   A third plaintiff was given a lower wage compared to males who were doing the same work.
       d.   But the jury didn't believe this.

What is the antecedent for the pronoun *this*? For most speakers, the only possible antecedents are either the proposition expressed by the combination of (9-a)-(9-c), a *complex discourse unit* that we will denote as [(9-a)-(9-c)], or the proposition expressed by the discourse unit (9-c) alone.

Note that how we resolve the pronoun *this* in (9) goes hand in hand with how we understand the scope of the discourse marker *but* in (9-d): if *this* picks out the proposition expressed by (9-c), then (9-c) is also understood as the first argument of *but*. However, if *this* is understood as picking up on the complex proposition formed from (9-a)-(9-c), then it is the complex discourse unit [(9-a)-(9-c)] that provides the first argument to *but*. Observations about propositional anaphora take insights from Section

---

[9]A dynamic semantic account explicitly underlies *Segmented Discourse Representation Theory* as developed in Asher and Lascarides (2003).

[10]For further discussion of why other dynamic frameworks which don't assume discourse relations fail to capture temporal anaphora, see Altshuler (2016); Lascarides and Asher (1993); Webber (1988).

1.1 concerning the relation between discourse relations and anaphora to a new level. Now, it's not just a question of how anaphoric relations between two consecutive clauses are interpreted; in SDRT, RST, and the theory of Polanyi (1985), *discourse attachment itself becomes anaphoric*.

When a new discourse unit is introduced into an ongoing discourse, we must consider which discourse units already present in the discourse will be able to connect with it via a semantic relation. In a coherent discourse, each new unit of discourse content must attach and bear some semantic relation to some other constituent in the discourse structure; each discourse unit becomes, in effect, a "zero-anaphor" looking for an antecedent discourse unit or complex discourse unit. And as illustrated by (9), it might be that only a subset of the constituents in a discourse representation are salient and available as attachment points when updating the discourse context with new information.[11]

To define the set of salient constituents that are accessible for attachment, commonly called *The Right Frontier* (RF; Asher (1993); Polanyi (1985)), we need to represent discourse structure as a graph whose nodes are discourse units and whose edges are instances of discourse relations between constituents.[12] We thus introduce a few fundamental features of the SDRT language here. The vocabulary contains a countable set of discourse unit labels $DU = \{\pi, \pi_1, \pi_2, \dots\}$ for *elementary discourse units* (EDUs), which are discourse units that cannot be decomposed into further discourse units, and complex discourse units (CDUs), which group together multiple DUs (EDUs or CDUs). It further includes a finite set of discourse relation symbols $Rel = \{R, R_1, \dots R_n\}$, which we add to the vocabulary of a language L, such as the language of dynamic predicate logic, for describing the contents of EDUs. Formulas in the SDRT language are of the form $\pi\colon \phi$, where ϕ, which describes the content of π, is: a formula of L; a formula of the form $R(\pi_1, \pi_2)$, which says that π_2 stands in coherence relation R to π_1; or a conjunction of SDRT formulas. Following Asher and Lascarides (2003), each discourse relation comes with constraints as to when it can be coherently used in context and when it cannot.[13]

A discourse structure for a text can be represented as a graph $(V, E_1, E_2, Last, \ell)$, where $V \subseteq DU$ is a set of vertices each representing a discourse unit; $E_1 \subseteq V^2$ a set of directed edges representing links between discourse units that are labelled by $\ell : E_1 \to Rel$ with discourse relations; $E_2 \subseteq V^2$ describes the membership relation between the set of DUs figuring in CDUs and the CDUs in which they figure; and $Last \in V$ is the last EDU in the linear, textual ordering of EDUs in d. An SDRS is *spanning* in that all elements of V other than the root have at least (and possibly more than) one incoming edge: $\forall \pi_x \in V.(\pi_x \neq \text{ROOT} \to \exists \pi_v \in V.((\pi_v, \pi_x) \in E_1))$. Note that when discourse units are grouped together in a CDU, they will be related in such a way as to determine a sub-graph respecting the foregoing conditions.
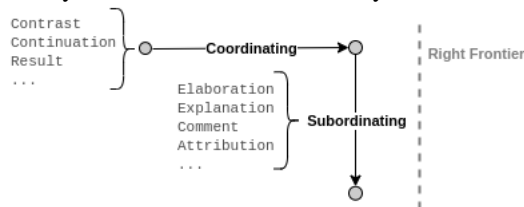
The *Right Frontier Constraint* (RFC) requires that given a discourse graph G, a new EDU to be attached to G must be attached to a node along the (RF) of G. (Nodes that are not on the RF can be accessed, but only through what Asher (1993) calls *discourse subordination*.) The RF evolves dynamically

---

[11]Note that discourse attachment is a more complicated anaphoric process than pronominal or temporal anaphora, because the semantic relations involved in discourse attachment are semantically varied. Furthermore, discourse theories provide complex, recursive structures for the discourse context, so there are typically several possible antecedents for discourse attachment. In addition, as Asher (1993) and Lascarides and Asher (1993) show, different attachment sites will affect temporal structure of the discourse and the possible interpretation of anaphoric pronouns.

[12]In some theories, such as RST, discourse structures are represented more specifically as trees, but this constraint is not central to our points here.

[13]Discourse structure inferences are generally nonmonotonic or probabilistic; still, implemented SDRT models (Muller et al., 2012; Afantenos et al., 2015; Perret et al., 2016) have proven predictive over large corpora of discourse annotated, extended texts.

as a discourse proceeds and is sensitive to whether a new DU is attached via a *subordinating relation* or a *coordinating relation*. A subordinating relation—including Explanation, Elaboration, and Background— is one in which the second argument seems to provide further information about the first (Asher and Lascarides, 2003). Crucially, the addition of the second argument does not render the first argument less salient or inaccessible for anaphora, which means that both discourse units will be on the RF. Let's return to our first example, repeated here as (10):

(10)   I need my hat back. I'm leaving for São Paulo in two days.

The speaker could easily continue with (11):

(11)   I'm sorry. I know you enjoy wearing it.

In apologizing, the speaker expresses the idea that she feels bad about asking for her hat back; the apology *I'm sorry* is thus related via the relation Comment to the discourse unit *I need my hat back*. The pronoun *it* in (11) likewise depends on the first discourse unit of (10), referring to the hat introduced in that unit. These attachments are possible despite the fact that the second sentence of (10) is uttered in the interim.

The left graph below represents the structure of (10). The vertical arrow connecting the discourse units indicates that Explanation is a subordinating relation, and the dashed line (RF1) represents the RF just before the speaker utters (11). The right graph shows how the discourse structure changes when we update with (11). *I'm sorry* is attached to the top node via Comment, a subordinating relation, and *I know you enjoy wearing it*, which explains why the speaker is sorry, is attached via Explanation. This graph also shows how updating with (11) changes the RF: the unit *I'm leaving for São Paulo in two days* is pushed to the left and is no longer on the RF (RF2), though the remaining three units are.
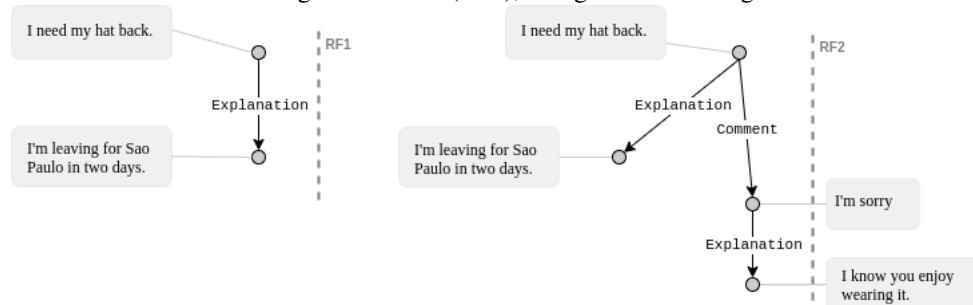
Coordinating relations shut off the accessibility of their first arguments and advance the discourse to a new topic instead of providing further information on the current topic. The discourse units contributed by (9-a)-(9-b) and (9-b)-(9-c), for instance, are related by the coordinating relation Continuation in SDRT (as are those in (3)), whose semantics roughly correspond to Boolean conjunction. This means that (9-b) is predicted to be inaccessible for attachment once (9-c) is introduced, which is what we observed above. Coordinating relations, such as Continuation, Narration, and Result, are represented with horizontal arrows to show that they push the RF forward, or to the "right" as shown in case 1 of the figure below. These assumptions imply that should we insert material in (9) such as:

(12)   These people were really badly treated.

before (9-d), the available antecedents for the pronoun *they* should shift again. And indeed as SDRT predicts, our intuitions change in this new example: (9-c) is no longer available as an antecedent.
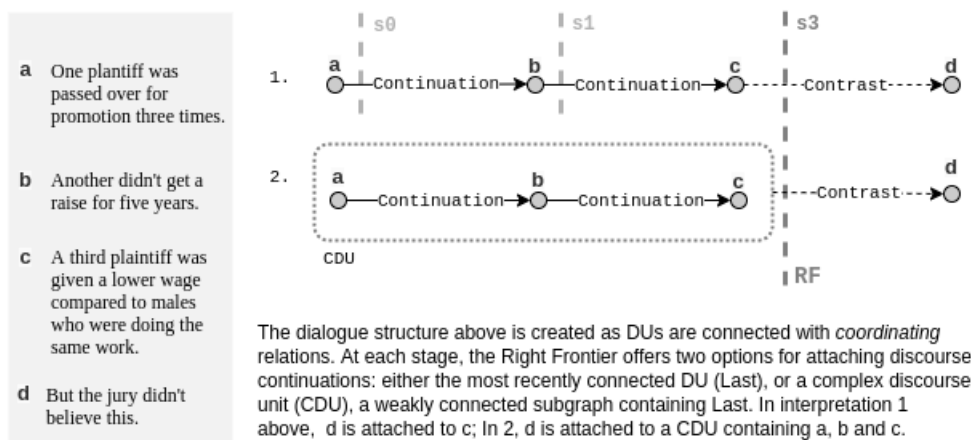
Note that the RF, as it can contain numerous discourse units, does not determine where a new discourse unit *will* attach to the discourse graph to date, but only where it *can* attach. We can now formally define the RF in the style of SDRT. Let $e(\pi_x, \pi_y)$ mean that edge e has initial point π_x and endpoint π_y. A node π_x is on the RF of a graph G, i.e. $\text{RF}_G(\pi_x)$, just in case π_x is *Last*, π_x is related to *Last* via a series of

subordinating (*Sub*) edges, or π_x is a CDU that includes a node in RF_G:

Definition 1 *Let $G = (V, E_1, E_2, Last, \ell)$ be a discourse graph.* $\forall \pi_x, \pi_y, \pi_z \in V$, $\mathrm{RF}_G(\pi_x)$ *iff*

   (i) $\pi_x = Last$,

   (ii) $\mathrm{RF}_G(\pi_y)$ & $\exists e \in E_1$, $e(\pi_x, \pi_y)$ & *Sub(e), or*

   (iii) $\mathrm{RF}_G(\pi_y)$ & $\exists e \in E_2$, $e(\pi_x, \pi_y)$

Note that the RF is updated dynamically each time a new EDU is processed; the RF for (attachment of) an EDU π_n will be determined by the graph $G_{\pi_0 - \pi_{n-1}}$. The RF for a CDU $\pi_m \ldots \pi_n$, $m < n$, is the RF for π_m.[14] This predicts that (9-c) is available for attachment in (9) because it is *Last*, but (9-a) and (9-b) are both inaccessible because neither satisfies any of the conditions (i)-(iii). The complex unit [(9-a)-(9-c)] is correctly predicted to be available, however, because it includes *Last*, which is a member of the RF, thus satisfying condition (iii).

The dialogue structure above is created as DUs are connected with *coordinating* relations. At each stage, the Right Frontier offers two options for attaching discourse continuations: either the most recently connected DU (Last), or a complex discourse unit (CDU), a weakly connected subgraph containing Last. In interpretation 1 above, d is attached to c; In 2, d is attached to a CDU containing a, b and c.

    The picture that emerges from an account of discourse structure is one in which the attachment point of a new discourse unit to an existing discourse graph is itself an anaphoric process guided by a combination of reasoning about world knowledge and linguistic cues. When a speaker makes a new utterance, determining to what part of the conversation their new utterance is relevant is in fact a complex process. Given the hypothesis developed in Section 1.1, that pronominal anaphora resolution and temporal interpretation are byproducts of inferring discourse relations and structure, it follows that the former are guided by the same complex reasoning processes as the latter. While the RF cannot on its own determine where a discourse unit will attach—and thus in what unit a pronoun must find its antecedent or a temporal expression must be interpreted—it helps to greatly restrict the possibilities and facilitate discourse comprehension while offering a more comprehensive mechanism for interpretation.

---

[14]This definition leaves open the possibility that a set of discourse units is first grouped into a CDU and that the graph determined by the CDU is only subsequently attached to the incoming graph. This happens frequently with utterances of sentences with a complex syntactic structure: a sentence of the form *q, if p*, for example, will determine a CDU that contains the DUs π_p and π_q linked to each other via the relation Conditional, i.e., Conditional(π_p, π_q). Sentences containing appositive relative clauses, which we discuss in Section 2.1.1, likewise contribute CDUs containing, at least, the DUs contributed by the main clause and the appositive.

2 Complex discourse structures

If the nature of dynamically evolving, complex discourse structures can influence the interpretation of semantic phenomena such as pronominal anaphora resolution and temporal interpretation, the question arises as to what other semantic phenomena might be efficiently modelled by exploiting the full machinery of a theory of discourse structure. In this section, we examine three other types of phenomena that we feel are best analyzed through the lens of a discourse theory, namely (certain types of) at-issue and not-at-issue content, discourse goals, and multi-modal interactions.

2.1 At-issue and not-at-issue content

Theories of discourse structure in the tradition of SDRT and RST have focused largely on defining the function of a discourse unit in terms of the kind of discourse relation to which it contributes: whether it serves to explain something, to answer a question, to continue a narrative, and so on. But this is not the only way to understand discourse function: fueled by the observations and theory presented in Potts (2005), there is an ongoing and lively debate in linguistics and philosophy of language about how to classify discourse content in terms of how central it is to discourse development and, often, to discourse goals or purposes. In current terminology, the challenge is to determine the conditions under which content is *at-issue* (AI), and thus central to discourse development and/or discourse goals, or *not at-issue* (NAI), and thus relevant to a discourse in some more indirect way.

In this subsection, we take a look at recent work that brings theories of discourse structure and interpretation to bear on the AI/NAI discussion by focusing on two phenomena that have been said to involve NAI content: appositive relative clauses and discourse parenthetical reports. Before addressing these topics in turn, however, we need to clarify what is meant by AI and NAI content. Efforts to define these concepts more precisely have led to a variety of diagnostic tests, and because these tests do not always yield the same judgments, the result is that there is more than one way of carving up the AI/NAI distinction (Koev, 2018). Here we will focus on two ways of categorizing AI and NAI content: as *backward-looking* (N)AI or as *forward-looking* (N)AI.[15]

To determine if content is backward-looking (N)AI, we look at how it interacts with the *preceding* discourse. Consider (13):

(13)   Marie, the chemistry teacher at our old highschool, is joining our volleyball team.

We say that the main clause of (13) is backward-looking AI while the appositive relative clause is backward-looking NAI because the former must be relevant to the preceding discourse in a way that the content of the latter need not be, as shown by the contrast between (14) and (15):

(14)   a.   Who is joining your team this year?
        b.   Marie, the chemistry teacher at our old highschool, is joining.

(15)   a.   Who is Marie?
        b.   ?? Marie, the chemistry teacher at our old highschool, is joining our volleyball team.

The infelicity of (15-a) arguably shows that the main clause of a sentence containing an appositive relative clause must convey main point content, i.e. be backward-looking AI, while the acceptability of (14-b) shows that the an appositive relative clause can be backward-looking NAI.

---

[15]The concepts of backwards looking at-issueness and forward-looking at-issueness correspond roughly to Koev (2018)'s notions of Q-at-issueness and C-at-issueness, respectively, which we find very helpful. We change the terminology for backwards-looking at-issueness to take a more agnostic stance as to the nature of the incoming discourse context; we opt for the notion of forward looking at-issueness to emphasize that Koev's notion of C-at-issueness pertains to those units in a discourse graph that can support anaphoric continuations, namely, the nodes along the RF.

Forward-looking AI status is diagnosed by looking at possibilities for *subsequent* discourse continuations, like those in (16-a) and (16-b):

(16)　　Marie, the chemistry teacher at our old highschool, is joining our volleyball team.

　　　　a.　　That's not true! (=*It's not true that Marie is joining the team.*)
　　　　b.　　Wait, I thought she was the physics teacher.

The main clause content of (16) is forward-looking AI because it is treated as more salient or discourse central by subsequent discourse moves, as shown by the fact that the pronoun *that* in (16-a) seems to automatically target this content, while ignoring that of the appositive relative clause. Correcting the latter requires more effort, as shown by (16-b); here, the speaker must employ explicit descriptive content to show that she is taking issue with the appositive, suggesting that the appositive content is forward-looking NAI (cf. Von Fintel (2004)).

With these notions of forward and backward-looking N/AI content in place, we now turn to a discussion of how discourse structure has been exploited to model the behavior of two types of content that sometimes exhibit unexpected AI behavior: appositive relative clauses and the embedded clauses of speech reports. While the foregoing discussion might lead us to conclude that appositive relative clauses are by their very nature vehicles for backward-looking and forward-looking NAI content, Section 2.1.1 introduce data that show they can be both backward and forward-looking AI in certain cases. Section 2.1.2 then focuses on data that show that the embedded content of a speech report can be backward-looking AI even while syntactically embedded under content that appears to be backward-looking NAI. In both cases, we show how a theory of discourse structure can be brought to bear on these phenomena in a way that accounts for their nuanced behavior. While a discourse-based account of appositive relative clauses emerges naturally from the existing tools such as the RF, an account of speech reports requires some supplemental assumptions.

2.1.1　Appositive relative clauses

As pointed out by numerous authors, appositive relative clauses pass diagnostic tests for forward-looking AI content when they appear in sentence-final position. This is illustrated by the fact that the direct rejection in (17-b) targets the content of the appositive as easily as (17-a) targets the content of the main clause (AnderBois et al., 2015; Syrett and Koev, 2015):

(17)　　This year, we'll be joined by Marie, (who was) the chemistry teacher at our old highschool.

　　　　a.　　That's not true. She's moving to Germany now.
　　　　b.　　That's not true. She was the physics teacher.

In fact, even appositive relative clauses in sentence-medial position can arguably convey forward-looking AI content in certain cases. Compare (18) and (19), from Hunter and Asher (2016).

(18)　　a.　　Marie, the best volleyball player in the district, is joining our team.
　　　　b.　　We're going to be invincible!

(19)　　a.　　Marie, the worst volleyball player in the district, is joining our team.
　　　　b.　　?We're going to be invincible!

While an appositive relative clause cannot be targeted by a direct rejection such as *That's not true*, (18) and (19) show that such a clause can nevertheless play a central role in the acceptability of discourse continuations. And crucially, it can play this role even if a speaker makes no particular effort to raise this content to salience. In contrast to the appositive in (16), which must be explicitly targeted by a move like (16-b) in order to be made salient, the appositive in (18) is automatically understood to be a part of the speaker's main point—that they're going to be invincible because the best player in the district is joining

their team. Arguably, then, sentence-medial appositives can sometimes be forward-looking AI, even if direct rejection tests fail to diagnose them as such.

Examples similar to (18) suggest that sentence-medial appositive relative clauses can be backward-looking AI as well (Syrett and Koev, 2015):

(20)   a.   Our team is so much stronger this year.
   b.   Marie, the best player in the district, joined our team in March.

Without the appositive relative clause, it might have been possible to infer from (20) that the team is stronger *because* Marie joined, but for an audience who does not know Marie or how good of a player she is, this interpretation is greatly aided by making explicit why Marie's presence would strengthen the team. As with (18), the appositive content in (20-b) plays a central role in conveying the speaker's main point, this time by directly contributing to the explanation of (20-a).

In a discourse theory, the above observations fall out naturally by appealing to the nature of subordinating relations and the RF (Hunter and Asher, 2016; Jasinskaja, 2016, cf. Asher (2000)). Let's begin with (17). Ignoring the frame adverbial *This year*, which would introduce complexities irrelevant to the current discussion, (17) can be decomposed into two discourse units, π_1: *we'll be joined by Marie* and π_2: *(Marie was) the chemistry teacher at our old high school*. In SDRT, these units will be related by the subordinating relation Elaboration, i.e., Elaboration(π_1, π_2), because the content of π_2 elaborates on the entity Marie, introduced in π_1. Recall that the set of accessible nodes on the RF include: (i) *Last*, (ii) any unit x directly superordinate to a node y on the RF, and (iii) any CDU z that includes a node y on the RF. The unit π_2 satisfies condition (i), as it is the most recently uttered discourse unit, while π_1 satisfies condition (ii) because it is superordinate to π_2 (e.g. the source of a subordinating relation connecting π_2 to the graph). We thus predict that both π_1 and π_2 are on the RF and available for discourse continuations, as desired.[16]
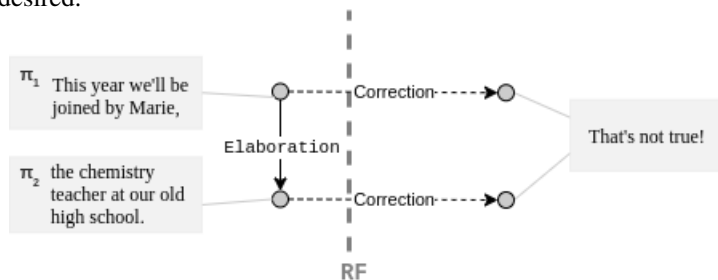
The definition of the RF can also be used to predict that the medial appositive relative clause in (18) cannot be targeted by a direct rejection although it *can* be relevant for discourse continuations like that in (18-b). (18) can be decomposed into two discourse units, π_1: *the best volleyball player in the district* and π_2: *Marie is joining our team*. In this case, the main clause, whose content is π_2, is the last completed unit in (18-a), and so it follows from condition (i) of the RF that it can support discourse continuations. The unit π_1, by contrast, fails to satisfy (ii), because is it actually *sub*ordinate to π_2, not *super*ordinate to it. Thus we predict, correctly, that π_1 *alone* cannot be targeted by a discourse continuation like *That's not true*. However, if π_1 contributes to a complex discourse unit that contains another unit on the RF, then by condition (iii), we predict that the entire complex unit can support discourse continuations. And this is exactly what we observe: the complex unit $[\pi_2, \pi_1]$ in (18) supports the continuation in (18-b). Parallel remarks can be made for the discourse centrality of the appositive relative clause in (20-b): the complex discourse unit as a whole provides the explanans, making the appositive discourse central.

---

[16]We relate the rejection "That's not true!" via the relation Correction. Corrections and rejections are in fact complicated discourse moves. For more, see Julian Schlöder's entry on Rejection in this volume.
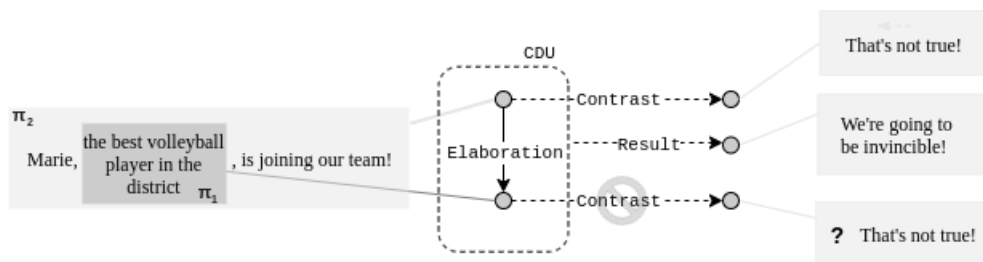
Recasting the AI/NAI distinction as a byproduct of constraints that guide discourse attachment and the construction of complex discourse structures provides an independently motivated and flexible notion of (N)AI content that accounts for the variable AI status of appositive relative clauses. Within such a framework, there is no need to posit that appositive content is by its very nature NAI or that it gives rise to a special interpretation procedure (cf. AnderBois et al., 2015); nor do we need to posit new syntactic constraints as in Koev (2013). In the next section, we consider another phenomenon that raises questions about the AI/NAI distinction, namely discourse parenthetical interpretations of indirect speech reports. Like the behavior of appositive relative clauses, the behavior of the embedded clauses of speech reports appears to motivate a discourse-level explanation. Unlike the former, however, a discourse-based analysis of discourse parenthetical reports requires us to adopt some new assumptions.

2.1.2 Discourse parenthetical reports

In certain cases, the embedded clause of an indirect speech report seems to convey backward-looking AI content despite being syntactically embedded under content that is less discourse central. Consider the contrast between (21) and (22).

(21)   a.   Rose is grumpy.
        b.   Nicholas said her chocolate cake is dry and bland.

(22)   a.   Rose is bringing dessert to the party.
        b.   Nicholas said she is making a chocolate cake.

In (21), the report in (21-b) as a whole explains why Rose is grumpy—regardless of whether or not Rose's cake actually *is* dry and bland, Rose is upset simply because Nicholas *said* it was. Intuitively, we could represent (21) using the first graph below. A parallel analysis for (22), shown in the second graph, is unsatisfactory, however; the speaker is not suggesting that the event of Rose bringing a dessert to the party is going to furthermore be *an event of Nicholas saying* that she is making a chocolate cake. The speaker rather seems to be committed to something closer to the elaboration captured by the third graph below, and the fact that Nicholas said what he did somehow provides evidential support for this elaboration. Following Hunter (2016), we will call speech reports like (22-b) in which the embedded content appears to be backward-looking AI while the report clause plays a supportive, evidential role, *discourse parenthetical*.
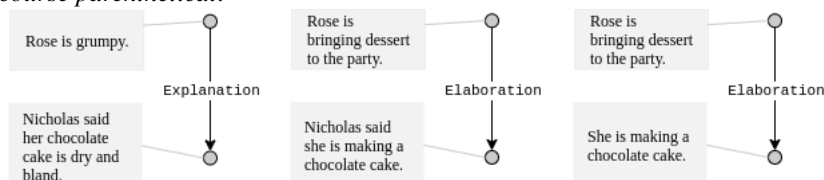

In an attempt to provide more intuitive annotations for discourse parenthetical reports that more accurately represent the inferences that one can draw from them, numerous discourse theories have proposed

that speech reports generate two discourse units, one for the main report clause and one for the embedded clause (Buch-Kromann and Korzen, 2010; Dinesh et al., 2005; Hunter et al., 2006, see also Carlson and Marcu (2001)). Such an approach has been further supported by experimental work in Simons (2019). As illustrated in (24), for example, the report in (22-b) can be decomposed roughly as follows: [Nicholas said [she is making chocolate cake.]$_{\pi'_3}$]$_{\pi'_2}$, so that the report as a whole introduces a discourse unit π'_2, and the embedded clause introduces a separate discourse unit π'_3: *she is making a chocolate cake*. Because the different interpretations of (22-b) and (21-b) seem to result from how the reports are used in the discourse, rather than some kind of hidden syntactic difference (Simons, 2007), we can further assume that all speech reports should be decomposed into two units, even when it is the main report clause that conveys discourse central information, as illustrated in (23).

(23)   a.   [Rose is grumpy]$_{\pi_1}$
       b.   [Nicholas said [her chocolate cake is dry and bland]$_{\pi_3}$]$_{\pi_2}$

(24)   a.   [Rose is bringing dessert to the party]$_{\pi'_1}$
       b.   [Nicholas said [she is making chocolate cake]$_{\pi'_3}$]$_{\pi'_2}$

To derive the distinction between non-parenthetical and parenthetical readings, then, one approach is to posit that they involve two different discourse relations, say Attribution in (21-b) and Source in (22-b) (Hunter et al., 2006), as shown in the figure below:
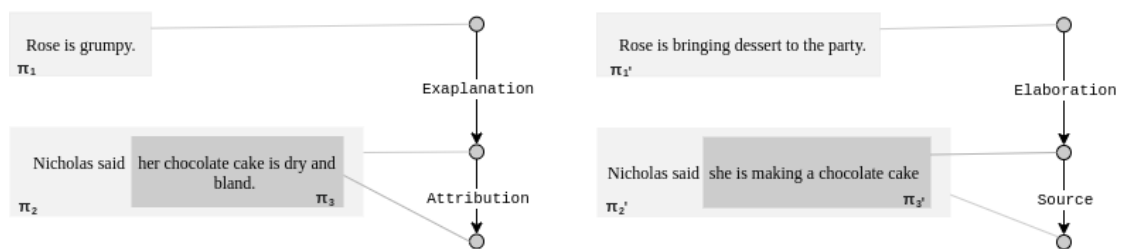
While Attribution mirrors the syntactic structure of a report, keeping the embedded clause subordinate to the main clause, Source reverses the order of its arguments so that the embedded clause can be directly related to the discourse preceding the report. This reflects the intuition that the embedded clause is backward-looking AI and is thus intuitively central for the incoming discourse context. Semantically, Attribution does not entail the truth of its second argument—i.e., the content of the embedded clause— and so is interpreted as expected for a report involving a non-factive verb. When the embedded clause of a report contributes the first argument of Source, however, its truth *is* entailed. Relations such as Elaboration and Explanation are *veridical*, meaning that they entail the truth of both of their arguments; it thus follows that if the embedded content of a report attaches to the incoming discourse via one of these relations, its truth is entailed.

While the Source relation addresses the intuition that π'_3 is backward-looking AI, it creates new problems. First, it fails to capture examples in which both the main clause and the embedded clause of a speech report are backward-looking AI, as illustrated by (25) (for extended examples, in which the clauses are related to very different parts of a discourse, see Hunter, 2016).

(25)   a.   Have you talked to the guests? What are they bringing?
       b.   Nicholas said Rose is bringing a chocolate cake, and he said that he would bring chips and guacamole. Kate is bringing veggie burgers, but I haven't heard from Isabel. Do you think I should call her?

In this example, the second speaker uses consecutive utterances to simultaneously provide suites of answers to both questions in (25-a), telling the first speaker which guests they have talked to, as well as

addressing the question of what each person is bringing. It is thus hard to say for any of the reports in (25-b) which is *the* unit that attaches to the incoming discourse context or licenses discourse continuations, so adopting Source to represent the discourse centrality of π'_3 is unmotivated at best (cf. also two-dimensional accounts such as Maier and Bary, 2015). Moreover, adopting Source fails to account for the fact that a speaker who uses a discourse parenthetical report generally hedges their commitment to the embedded content of that report: in (22-b), the speaker is not fully committed to the claim that Rose is making a chocolate cake for the party; they are committed at most to $\lozenge\text{Elaboration}(\pi'_1,\pi'_3)$, so Elaboration$(\pi'_1,\pi'_3)$ is too strong. This observation would be naturally explained by appealing to the fact that π'_3 is in the scope of a speech report—that is, by making π'_3 subordinate to π'_2 via Attribution.

Now we're back to the drawing board: if we posit Attribution(π'_2,π'_3), how can we represent the intuitive Elaboration relation between π'_1 and π'_3? Directly relating them will lead to a violation of the RF, as can be seen from the figure below.
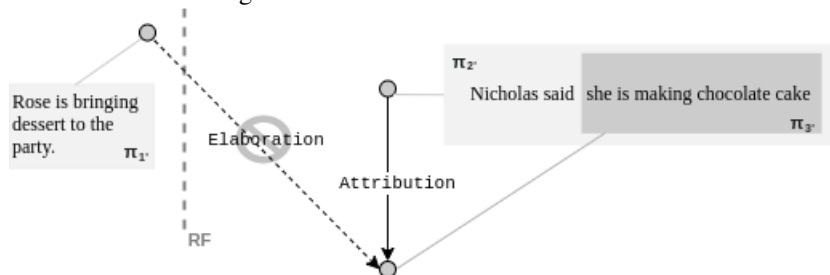
π'_2 and π'_3 form a complex unit that needs to be attached to the incoming discourse, i.e., π'_1. The only way to do this without violating the RF is to attach π'_2 to π'_1, but that would yield the reading of (22-b) that we have rejected, namely a reading in which Nicholas saying what he did elaborates on the event of Rose bringing a dessert to the party. Attaching π'_3 directly to π'_1 is not permitted by the RF as introduced in Section 1.2: π'_1 is not *Last* for π'_3 (π'_2 is), nor is π'_1 superordinate to π'_3 via a chain of subordinating relations, as π'_2 is not attached to π'_1 at all.

As argued in Hunter (2016), however, modifying the RFC to allow for such "violations" is independently motivated in the case of third-party speech reports. As Hunter explains, the constraint that a discourse unit attach to another unit along the RF "was developed as a constraint on how a speaker presents her own commitments in monologue. When a speaker chooses to use someone else's commitments to accomplish this task, she must first set up the other speaker's commitments and then use them".[17] Of course, while speakers can use things that others have said to, say, elaborate on or explain other discourse units, we have seen that when they do so, they weaken their own commitments to the reported content. It follows that they hedge their commitments to the proposed relations as well: if the speaker of (22-b) is not entirely committed to the claim that Rose is making a chocolate cake, they cannot be entirely committed to the claim that the event of Rose's bringing a dessert to the party is going to be an event of her bringing a chocolate cake to the party.

Accordingly, Hunter (2016) posits that anytime a speaker opts for a discourse parenthetical report, the report will contribute an instance of Attribution, just as a non-parenthetical report would, but when we link the embedded content to the discourse context preceding the report, the Attribution will have the effect of weakening the speaker's commitment to the relation. That is, rather than Elaboration(π'_1,π'_3) as in the graph above, linking π'_1 to content inside of an Attribution context weakens the relation to $\lozenge\text{Elaboration}(\pi'_1,\pi'_3)$.[18] This proposal allows us to systematically derive the difference between discourse

---

[17]The RF as defined in Section 1.2 must be adapted for multi-party dialogue as well, where building up a discourse structure becomes a collaborative task (Asher et al., 2016).

[18]In this account, the evidential effect of the report is derived as a byproduct of discourse attachment between multiple discourse units. For a discussion of evidentiality at the propositional and subpropositional level, see Diti Bhadra's entry on Evidentiality (this volume).

parenthetical and non-parenthetical readings with minimal, well-motivated adjustments to a classic discourse theory: in discourse parenthetical readings, the embedded content is backward-looking AI and thus related directly to the incoming discourse context, although the speaker's commitment to this relation is hedged; in non-parenthetical readings, the main clause is backward-looking AI and no speaker commitment is entailed to the embedded content of the report (for third-person reports). Furthermore, this approach predicts that both the main and embedded clauses of a speech report can be backward-looking (and forward-looking) AI, as desired.

In the analysis of at-issue and not-at-issue content that emerges from this section, at-issue content is content that is central to discourse development. A discourse unit π is AI if it attaches directly to the incoming discourse context via a discourse relation or supports anaphoric continuations that need not explicitly evoke the content of π; a discourse unit π' is NAI if it is contributed by a syntactically complex discourse move that contributes more than one discourse unit, and π' does not (or cannot) attach to the incoming context or cannot support anaphoric continuations on its own. The possibilities for attachment of a discourse unit are in turn governed by the RF and rules limiting discourse development. As mentioned at the outset of this section, however, AI content is sometimes presented as content that directly addresses a speaker's discourse goals. In the next section, we take on the topic of modelling discourse goals in a theory of discourse structure and interpretation and show how goals and at-issue content are decoupled in more recent work.

2.2 Discourse goals and subjectivity

Language is a tool for achieving one's ends, even if one's goal is merely to pass the time or to make someone laugh. Understanding how language can be used to bring about certain effects on one's audience has been of interest to the study of language going back to ancient work on rhetoric. The study of discourse goals has also recently become a main topic of interest in discussions of discourse analysis, from SDRT to theories of conversation centered around Questions Under Discussion, in which a discourse goal is understood as a question that a discourse move is expected to address (QUDs; Ginzburg, 2012; Roberts, 2012; Simons et al., 2010). Modelling goals is important for discourse interpretation for multiple reasons. First, because we can expect a speakers' discourse goals to guide the discourse moves they make and the way they put them together, we can expect discourse structure and goals to be very closely related. Moreover, an interpreter's expectations concerning what a speaker aims to achieve with her discourse will affect not only how she chooses to converse with the speaker, but also how she interprets the speaker's moves when there is ambiguity (as is often the case at the discourse level).

In this section, we take a look at how discourse relations and complex discourse structures can be used to model discourse goals. We also consider how the relation between discourse goals and AI content should be understood given the discourse structural perspective on AI content developed in the previous section. We conclude by showing how different perceptions of discourse goals can lead to different interpretations of what is said in discourse.

2.2.1 Goals

Sometimes, speakers converse to get information from an interlocutor or to persuade someone of a position; in other exchanges, the desired outcome might be an action of some sort, as in (26).

(26)   a.   Julie: It's time to go to bed.
        b.   Rose: OK, good night.
        c.   [non linguistic action: Rose goes to bed.]

Intuitively, one might simply say that Julie's discourse goal in (26) is to get Rose to go to bed, as she does in (26-c). If (26) is a conversation that goes well for Julie and meets her discourse goals, however, (27),

which has the same outcome, is a much less satisfactory conversational exchange:

(27)   a.   Julie: It's time to go to bed.
       b.   Rose: OK, but I'm still watching my show.
       c.   [30 mins later] Julie: OK, Rose it's really time now to go to bed.
       d.   Rose: I'm still watching my show. You told me I could!
       e.   Julie: It's no longer the same show! No story! [followed by half an hour of arguing . . .]
       f.   [Rose goes to bed.]

A satisfactory model of the relation between discourse structure and goals cannot focus only on whether a conversation successfully achieves a desired outcome, but also on *how* the outcome is achieved. Speakers are usually trying to satisfy multiple constraints at once.

Because a full understanding of discourse goals usually requires modelling extended discourses and goals can be ranked not only by their final outcomes but by the different paths that the conversation can take to achieve these outcomes, recent work in this area models discourse goals as sets of full discourse structures—the structures in which the conversation "goes well" for a particular conversationalist. Asher et al. (2017) model a conversational goal as a subset of all possible conversations or discourse structures in the sequential game space of all possible discourse moves. The goal of making a conversation coherent, for example, is modelled as the set of all coherent discourse structures or, alternatively, as the set of all conversations or strings of discourse moves that generate such structures. Not angering ones interlocutor might be another goal, denoting a different set of structures. Exogenously given decision problems or conversations aimed at answering a particular question are also instances of such goals. Goals can be complex, formed from combinations of simpler goals. Where Win_i is the set of goals for a player i, the strategies that i adopts in conversation—the discourse moves that i chooses to make and how they are related—will be adopted to steer the conversation into Win_i.

Now if discourse goals are modelled as (sets of) full discourse structures, and AI content is defined in terms of attachment within larger discourse structures, what is the relation between AI content and goals? Consider the following exchange from the film *The Princess Bride*, in which the AI/NAI distinction is exploited to achieve a discourse goal:

(28)   a.   Buttercup: He [Humperdink]... can find a falcon on a cloudy day, he can find you!
       b.   Wesley: You think your dearest love will save you?
       c.   Buttercup: I never said he was my dearest love, and yes, he will save me. That I know.
       d.   Wesley: You admit to me that you do not love your fiancé?
       e.   Buttercup: He knows I do not love him.

As background for those who haven't seen the cult classic, Buttercup and Wesley had previously been in love and had planned to marry, but then Wesley was taken hostage by pirates and Buttercup was told that he was dead. A few years later, Prince Humperdink chose Buttercup to be his bride, though she had no desire to marry him. In this scene, Buttercup has been taken hostage by Wesley whom she believes, given his disguise and behavior, to be the pirate who killed Wesley. In the conversation above, Wesley exploits the fact that Buttercup does not recognize him to try to get her to say, without exogenous influence, whether she still loves him.

Let's now look at how Wesley uses the conversation to achieve this goal. In (28-a), Buttercup tries to intimidate Wesley so that he will release her, and Wesley follows up in (28-b) with a confirmation question that seems to directly address her goal, but in fact, he is merely seizing the opportunity for his own ends. He aims to find out whether Buttercup loves Humperdink, but surmises that a direct question might make Buttercup suspicious or trigger feelings of guilt, leading to a less than fully reliable answer. Wesley thus opts to disguise his question in a presupposition (*your dearest love*), a paradigm NAI con-

struction, in (28-b).[19] Buttercup takes the bait in (28-c) and directly responds to the non-central content of Wesley's question, which allows Wesley to follow up directly on his real question about whether she loved Humperdink in (28-d). Wesley continues to return to the topic in later scenes, and ultimately admits to Buttercup that he disguised himself in order to get an honest answer to his question.

In (28), the presuppositional content in (28-b) is arguably more directly related to Wesley's discourse goal than is the AI content. It's not just that Wesley wants an answer to the question of whether Buttercup still loves him; he wants to get this answer in the most reliable way possible, and opting for a presuppositional expression figures in an optimal strategy for achieving this outcome. From the discourse-driven perspective developed in this paper, then, AI content turns out to be a very local notion in the sense that is understood in terms of how a discourse unit attaches to the incoming discourse or licenses subsequent discourse moves. A discourse goal, on the other hand, will generally be a much larger structure (or set of larger structures) and we do not predict that AI content will reflect a discourse goal in any direct sense.

This understanding of the relation between goals and AI content stands in contrast to that developed in the QUD-based account of Roberts (2012). The latter assumes that conversation is a fundamentally cooperative activity aimed at getting more information about the world and posits that AI content is content that directly addresses a speaker's discourse goal, which is understood as the question (QUD) that the discourse tries to answer. However, while someone who has a cooperative goal of sharing information with an interlocutor might find AI constructions to be the most straightforward means of sharing discourse central information with an interlocutor, people adopt a wide variety of goals that might make use of AI content in less direct ways, and in some cases, hiding one's central concerns behind NAI content might be preferable. It follows that in a discourse based account of goals and AI content, an AI discourse unit may not have a direct relation to a discourse goal, but merely play an important part in how that goal is realized.

2.2.2 Bias and subjective interpretation

In (28), while Wesley and Buttercup seem to agree on what has been said in the conversation, the fact that they come to the conversation with different sets of background beliefs, including their understanding of whom Buttercup is talking to, leads to importantly different perceptions of Wesley's discourse goal. In other situations, discrepancies in background beliefs and expectations can lead to different interpretations of discourse structure. To illustrate this we revisit (8), repeated here as (29):

(29)   The city council denied the demonstrators a permit and they advocated violence.

(29) is ambiguous: it can be interpreted as expressing either an instance of Parallel or of Result depending on the context, the interpreter's background beliefs, and expectations about the speaker's discourse goals. Such small-scale ambiguities both at the level of relation type and attachment point arise somewhat regularly in discourse interpretation, and are a familiar phenomenon to anyone who has tried to annotate texts for discourse structure and had to arbitrate inter-annotator disagreement. In conversation, such interpretive differences might not be exposed unless one interpretation comes into contradiction with some other part of the discourse. Thus, a speaker and interpreter might have conflicting interpretations of a discourse without even realizing it. This is not always problematic; it might be completely irrelevant to an interpreter's goals to settle on one interpretation or another.

In other cases, however, disagreements about discourse interpretation can become central to discourse content and development, and even have legal ramifications. Consider the following exchange, discussed in Asher and Paul (2018), in which a reporter is questioning Sheehan, the spokesperson for the former U.S. senator Norm Coleman:

---

[19]Presuppositions provide another great example of a type of NAI construction whose behavior is arguably best modeled through the lens of a theory of discourse. See Márta Abrusán's entry on presupposition in this volume for more discussion.

(30) a. **Reporter:** On a different subject is there a reason that the Senator won't say whether or not someone else bought some suits for him?
   b. **Sheehan:** Rachel, the Senator has reported every gift he has ever received.
   c. **Reporter:** That wasn't my question, Cullen.
   d. **Sheehan:** (i) The Senator has reported every gift he has ever received. (ii) We are not going to respond to unnamed sources on a blog.
   e. **Reporter:** So Senator Coleman's friend has not bought these suits for him? Is that correct?
   f. **Sheehan:** The Senator has reported every gift he has ever received....

In (30-b), Sheehan responds to the Reporter's question in (30-a). Sheehan acts as though he is answering the question, and an audience biased towards Sheehan or the senator he represents might very well take his response as an answer (and likewise for Sheehan's other responses). The reporter, however, clearly does not interpret Sheehan's move as an answer, leading to a repetitive back and forth exchange, as each tries to push their particular discourse goal.

Asher and Paul (2018) provide a way of modelling competing interpretations of a conversation in an epistemic game theoretic framework, and they show how discourse goals, and interpreters' views on these goals, influence discourse interpretation. They also show how interpreters of conversations such as (30) can become more and more convinced of their interpretation as the dialogue continues. Supporters of the reporter see Sheehan's repetition of *the Senator has reported every gift he has ever received* as confirming more and more that he is evading the reporter's questions, while supporters of Sheehan get more confirmed in their belief or bias that Sheehan has answered the question and that it's time to move on. This phenomenon of *bias-hardening* in interpretation gets replayed at the level of beliefs as well, and can be very hard to control, let alone eliminate. This is a familiar phenomenon from political discussions and even personal relationships, and given the impact that it can have on our ability to use language to exchange ideas or learn about our world, an important topic for philosophers and linguists to grapple with.

2.3 Multimodal interactions

As hinted at in the discussion of (26) and (27), complex discourse structures can also be employed to model multimodal discourse. Let's return to example (26), repeated here as (31):

(31) a. Julie: It's time to go to bed.
   b. Rose: OK, good night.
   c. [Rose goes to bed.]

The exchange in (31) culminates in a nonlinguistic event of Rose going to bed, but with a young child who still needs guidance, successfully getting her to go to bed would likely involve multiple multimodal exchanges along the way. (32) offers one such example:

(32) a. Julie: It's time to get your pyjamas on.
   b. [Rose puts on her pyjamas]
   c. Julie: OK. Now let's go brush your teeth.

In (32), the event of Rose putting on her pyjamas contributes semantic content to discourse in much the same way as (33-b) does in (33):

(33) a. Julie: It's time for snack.
   b. Rose: I'd like some applesauce and cookies.
   c. Julie: OK. Now go wash your hands.

Rose's response in (33-b) contributes a discourse unit that plays a central role in discourse development; were we to take it out, the remaining discourse would be infelicitous in part because there would be no answer for Julie to acknowledge in (33-c) and in part because there would be no concluded event to license the discursive use of *now*, which indicates that the speaker is moving from one eventuality to another in a sequence. In this case, *now* is licensed because the discussion about a snack has been concluded and it is time to move on to the next topic. Similarly, in (32), we need to understand the nonlinguistic event in (32-b) as contributing propositional content that can serve as an argument to a discourse relation. It is this event, and more specifically the event together with semantic content that is understood to describe it—that licenses the Acknowledgment marked by *OK* and makes it possible to close off the pyjama discussion and move on to teeth-brushing via an instance of the relation Sequence, whose second argument is *Now let's go brush your teeth*.

Given the claim laid out in Section 1 that pronoun resolution is guided by reasoning about discourse relations and structure, we should expect reasoning about the discursive role of nonlinguistic eventualities to influence demonstrative reference as well. Suppose that the exchange in (32) continues with (34):

(34)   a.   [Rose starts toward the bathroom]
       b.   Rose: Wait!
       c.   [Rose goes back to her bed, grabs her teddy bear, and then heads back to the bathroom]
       d.   Rose: [looking up at Julie] He needs to brush his teeth too.

In (34-d), *he* will clearly refer to the teddy bear. At first glance, this might not sound so surprising—of course third person pronouns can be used to refer to entities in the nonlinguistic context. But there is a lot more going on here than demonstrative reference. For one thing, Rose need not point to her bear or even look at him to get the demonstrative reference to work. In addition, understanding the relation between (34-d) and the nonlinguistic events described in (34-c), and how this interaction contributes to the interpretation of the larger interaction between Julie and Rose requires more than understanding to whom *he* refers. Rose is not merely saying that her teddy bear needs to brush his teeth; she is *explaining why she is taking him to the bathroom*. This Explanation relation is crucial in the context: we can easily imagine a different scenario in which Rose goes back to get her bear because she suddenly remembered that she forgot to give him dinner and now wants to go feed him. This scenario is likely to get a negative reaction from Julie. In explaining her actions as she does in (34-d), she shows that she understands that it's time for teeth-brushing and that she is cooperating with Julie's discourse goal, making her more likely to get a positive reaction. The fact that she is explaining her previous action also explains why she doesn't need to go to any further trouble to make the bear salient. The entire sequence of events in which she went to get him and then started walking with him in her arms is salient, and he is a central figure in that sequence of events (cf. Stojnic et al., 2013).

The important point is that nonlinguistic eventualities do not only influence the *interpretation of a linguistically expressed discourse unit*, as entities picked out through deixis do; they can actually contribute *entire discourse units*, and they can do so without being picked out by any kind of referential expression (Hunter et al., 2018). This means that contents contributed by nonlinguistic events need to be taken into account in models of discourse structure—a difficult task given that nonlinguistic events are parts of the actual world and not just denotations of speech acts. This also means that they might impact discourse development. In fact, Hunter et al. (2018) argue that non-linguistic eventualities do not contribute to the RF in the way that linguistically expressed contents do and thus have different effects on salience (cf. the concluding discussion of Simons, 2019, comparing implicated content and explicit content).

Moreover, the top-down effects of multimodal discourse go beyond the interpretation of deictic or temporal expressions in a clause: given that in multimodal conversation, a non-linguistic eventuality can contribute an entire discourse unit in the absence of any linguistic description of that event, reasoning about discourse relations and structure can determine an entire event-level content. The event in (32-b) might be conceptualized differently in a different context, for instance; we might rather think of it as an

event in which *Rose changes out of her dirty clothes* or simply, *Rose changes clothes*. But in the context of (32), these other conceptualizations will not do: Julie must understand the event as one in which *Rose changes into her pyjamas* because only that kind of event will satisfy Julie's request in (32-a). A related point is that there are multiple ways of grouping and describing the events that take place in (32-b). While all of the actions involved in (32-b) were grouped together under the description *Rose puts on her pyjamas*, in another context, it might have been more pertinent to focus on some part of this larger event, e.g., *Rose put on her pyjama top*.

This discussion highlights an aspect of multimodal discourse that makes it very difficult to study systematically. The nonlinguistic context consists of a potentially evolving stream of information that must be decomposed into discourse-unit-level segments according to discourse purposes, but there is nothing like grammatical structure or intonation to suggest segment boundaries. And even if we determine such boundaries, individuated eventualities must be assigned semantic contents. The difficulty of assigning content to nonlinguistic eventualities makes studying either direction of information flow—bottom up or top down—a very daunting task.

Work in Lascarides and Stone (2009a) and Lascarides and Stone (2009b) has made some important first steps to understanding how discourse structure and interpretation can guide the conceptualization or description of nonliguistic events by focusing on the interaction of discourse and coverbal gesture. Their research also reveals, however, that coverbal gesture illustrates yet a different kind of discursive interaction from either those observed between purely linguistically expressed discourse units or those observed above in (32) and (34). Like appositive relative clauses and discourse parenthetical reports, utterances accompanied by coverbal gestures express two discourse units that are introduced into the context as a complex update and the gesture exhibits a dependence on linguistically expressed content not observed with contributions like that of (32-b) in (32); however, like the nonlinguistic eventualities discussed above, Lascarides and Stone argue that the salience of coverbal gestures is not controlled by the RF as defined in Section 1, though because of their dependence on accompanying speech acts, is not like the salience constraints on nonlinguistic eventualities like that described in (32-b) either.

3   Looking ahead

The foregoing discussion raises a variety of questions that will be important to future research on discourse structure and interpretation. First, what is the relation between discourse relations and Questions Under Discussion (QUDs)? While our main focus has been on discourse relations, QUDs have become a popular tool among formal semanticists for diagnosing the presence of utterance-level phenomena that semantically depend on the incoming discourse context. Much of this work centers on very specific types of discourse dependencies, such as focus structure, which have not been at the center of attention in work on discourse relations, and so might be seen as complementary. Some linguistics have posited that QUDs actually play a more fundamental role in determining salience and driving discourse development, however, and that discourse relations are derivative on them (Roberts, 2012), while others have argued against such a position (Hunter and Abrusán, 2015). Regardless of how this debate turns out, there are good reasons to think that there is something interesting to be said about the interaction between discourse relations and questions. Let's return to (5), repeated here as (35)

(35)   a.   What are you going to do downtown?
        b.   Ugh I'm so mad! My brother lost my copy of *The Watchmen*, and I need to reread it for my project. I'm going to the bookstore.

In addition to the relations at work, the speaker of (35-b) seems to be answering an implicit question of why a trip to the bookstore is necessary. How does either a discourse structure or QUD account handle this example? And what are the effects for the theory of discourse goals presented in this paper if part of

what drives discourse development is left implicit?

Another ongoing discussion that will continue to be important for future study concerns how to make more precise predictions about discourse attachment. The RF determines a set of nodes available for discourse attachment, but it cannot help predict for a given incoming discourse unit which node on the RF will be the best choice. By adding information about, say, prosody, or different kinds of constructions (such as appositives), or lexical facts, we might be able to say more. Following Hirschberg and Pierrehumbert (1986)'s attempt to link discourse structure and relations to the interpretation of prosody, researchers in SDRT have examined links between questions, prosody and discourse structure (Reese (2007); Asher and Reese (2007); Reese and Asher (2007)), but there is much, much more to explore.

The discussion in Sections 1 and 2, however, highlights three significant features that will complicate efforts to answer these questions and to systematically study discourse structure in general. First, because background beliefs and discourse goals add a highly subjective element to discourse interpretation that can be hard to pin down and eliminate, any efforts to use experimental or survey data to study discourse sensitive phenomena have to be very careful to control contextual elements that could influence interpretation. When it comes to judgments about discourse, the question is not only whether a certain discourse structure is acceptable, but what relations are at work in that structure. If two speakers disagree in their judgments, it could be that they disagree about the acceptability of the very same discourse structure, but it could also be, especially if a lot of context is left implicit, that they imagined different discourse contexts or inferred different discourse goals and thus interpreted the discourse differently.

A second hurdle is that providing an analysis of conversation that takes place in a shared perceptual environment will require modelling relations between linguistically expressed discourse units on the one hand and contents assigned to nonlinguistic actions, events and states, on the other. But as we have seen, individuating nonlinguistic eventualities is not a straightforward task, nor is specifying their semantic contents. The way in which discourse structure and interpretation guides the conceptualization of nonlinguistic eventualities, and vice versa, is still very much an open question.

Finally, developing an analysis of discourse-sensitive phenomena often requires considering extended discourse structures, not just pairs of discourse units, as some phenomena only develop over multiple discourse moves. The relation between the NAI content in Wesley's move in (28-b) (*You think your dearest love will save you?*) and his discourse goal, for example, would not have been apparent had we only considered that move together with (28-a) (*He [Humperdink]... can find a falcon on a cloudy day, he can find you!*). Nor does the fact that both discourse units in a discourse parenthetical report can be backward-looking AI come out if we only consider the report and one preceding discourse unit. Even the effects of the RF are hard to test if we only consider two or three units.

When developing accounts of linguistic phenomena in formal semantics and philosophy, the standard tool of choice is the minimal pair. Approaches to modelling intersentential phenomena extend this to looking at minimal pairs of pairs. Tests for forward-looking at-issueness, for example, tend to apply the "that's not true!" test to a given example and then apply the "wait" (or "hey, wait a minute!" (Von Fintel, 2004)) test to the same example. Similarly, backward-looking at-issueness is often diagnosed with question/answer pairs. Such diagnositic tests are very useful for showing the *existence* of discourse sensitivity and can shed light on some minimal aspects of discourse structure, as in the case of Simons (2019)'s experiments that support the view that indirect speech reports always contribute two discourse units. But the complexity of discourse structure and the subjectivity of discourse interpretation make these tools inapt for developing explanatory accounts of discursive phenomena. For this, we need discourse examples complicated enough to show the full behavior of the phenomenon in question and also to limit the influence of contextual factors. By embedding a target discourse structure inside of a larger discourse structure, we can better control the background context in which the target structure is interpreted and limit the direction in which an interpreter can expect the discourse to develop.

Extended, natural sounding discourses are difficult to invent, however, and much work on discourse structure and interpretation has heavily relied, and will continue to rely, on the annotation of corpora (see

21

also Abrusan's discussion of the need for corpus data in this volume). Deep learning approaches, or other machine learning methods designed to entirely bypass annotation, have not been successful at learning discourse structure—the lack of good training data, the sparsity of positive attachments in any given data set, and the presence of long-distance attachments makes the task particularly difficult. Of course, corpus annotation comes with its own set of well known problems. It is incredibly time consuming, first of all, and requires annotators who are ready to think carefully about how to most reasonably represent the content of a given discourse, which usually requires a level of experience that makes finding reliable annotators difficult. To further complicate matters, discourse structure often contains long-distance dependencies, where a discourse unit π_n is attached to a discourse unit that was produced many steps back rather than to the discourse unit π_{n-1} that was expressed immediately prior to π_n. This means that you cannot simply divide a discourse into chunks of 3 or 4 units and pass them out to different annotators or appeal to crowdsourcing if you want to get good annotations.

For these reasons, future work on discourse structure is going to have to get creative, especially for studying multimodal discourse. Recent attempts to apply distant supervision methods to produce automatic discourse annotations on chat discussion are very promising (Badene et al., 2019a,b). Hopefully, future work will prove the general applicability of these methods to other types of discourse so that we can make the systematic study of discourse structure as accessible as the study of more local semantic phenomena has been.

References

Afantenos, S., E. Kow, N. Asher, and J. Perret (2015). Discourse parsing for multi-party chat dialogues. In *Empirical Methods in Natural Language Processing*, pp. 928–937. Association for Computational Linguistics.

Altshuler, D. (2016). *Events, states and times: An essay on narrative discourse in English*. Walter de Gruyter GmbH & Co KG.

AnderBois, S., A. Brasoveanu, and R. Henderson (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics 32*(1), 93–138.

Asher, N. (1993). *Reference to abstract objects in English*, Volume 50. D. Reidel, Dordrecht.

Asher, N. (2000). Truth conditional discourse semantics for parentheticals. *Journal of Semantics 17*(1), 31–50.

Asher, N., J. Hunter, M. Morey, F. Benamara, and S. Afantenos (2016). Discourse structure and dialogue acts in multiparty dialogue: The STAC corpus. In *Language Resources and Evaluation Conference*, pp. 2721–2727.

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

Asher, N. and S. Paul (2018). Strategic conversation under imperfect information: epistemic Message Exchange games. *Logic, Language and Information 27.4*, 343–385.

Asher, N., S. Paul, and A. Venant (2017). Message exchange games in strategic contexts. *Journal of Philosophical Logic 46*(4), 355–404.

Asher, N. and B. Reese (2007). Intonation and discourse: Biased questions. *Interdisciplinary studies on information structure 8*, 1–38.

Badene, S., K. Thompson, J.-P. Lorré, and N. Asher (2019a). Data programming for learning discourse structure. In *Association for Computational Linguistics (ACL)*.

Badene, S., K. Thompson, J.-P. Lorré, and N. Asher (2019b). Weak supervision for learning discourse structure. In *Empirical Methods in Natural Language Processing*.

Buch-Kromann, M. and I. Korzen (2010). The unified annotation of syntax and discourse in the copenhagen dependency treebanks. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pp. 127–131.

Carlson, L. and D. Marcu (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545 54*, 56.

Dinesh, N., A. Lee, E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber (2005). Attribution and the (non-) alignment of syntactic and discourse arguments of connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pp. 29–36.

Ginzburg, J. (2012). *The Interactive Stance*. Oxford University Press.

Grice, H. P. (1981). Presupposition and conversational implicature. In P. Cole (Ed.), *Radical Pragmatics*, pp. 183–198. New York: Academic Press.

Hirschberg, J. and J. Pierrehumbert (1986). The intonational structuring of discourse. In *24th Annual Meeting of the Association for Computational Linguistics*, pp. 136–144.

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

Hobbs, J. R. (1985). On the coherence and structure of discourse.

Hunter, J. (2016). Reports in discourse. *Dialogue & Discourse 7*(4).

Hunter, J. and M. Abrusán (2015). Rhetorical structure and QUDs. In *JSAI International Symposium on Artificial Intelligence*, pp. 41–57. Springer.

Hunter, J. and N. Asher (2016). Shapes of conversation and at-issue content. In *Semantics and Linguistic Theory*, Volume 26, pp. 1022–1042.

Hunter, J., N. Asher, and A. Lascarides (2018). A formal semantics for situated conversation. *Semantics and Pragmatics 11*.

Hunter, J., N. Asher, B. Reese, and P. Denis (2006). Evidentiality and intensionality: Two uses of reportative constructions in discourse. In *Proceedings of the Workshop on Constraints in Discourse*, pp. 99–106. Citeseer.

Jasinskaja, K. (2016). Not at issue any more. Ms.

Kamp, H. (1988). Discourse representation theory: What it is and where it ought to go. *Natural Language at the computer 320*(1), 84–111.

Kamp, H. and U. Reyle (2013). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, Volume 42. Springer Science & Business Media.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.

Kehler, A., L. Kertz, H. Rohde, and J. L. Elman (2008). Coherence and coreference revisited. *Journal of semantics 25*(1), 1–44.

Koev, T. (2018). Notions of at-issueness. *Language and Linguistics Compass 12*(12), 1–16.

Koev, T. K. (2013). *Apposition and the Structure of Discourse*. Ph. D. thesis, Rutgers University-Graduate School-New Brunswick.

Lascarides, A. and N. Asher (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and philosophy 16*(5), 437–493.

Lascarides, A. and M. Stone (2009a). Discourse coherence and gesture interpretation. *Gesture 9*(2), 147–180.

Lascarides, A. and M. Stone (2009b). A formal semantic analysis of gesture. *Journal of Semantics 26*(4), 393–449.

Maier, E. and C. Bary (2015). Three puzzles about negation in non-canonical speech reports. In *The 20th Amsterdam Colloquium*.

Mann, W. and S. Thompson (1987). Rhetorical structure theory: A framework for the analysis of texts. *International Pragmatics Association Papers in Pragmatics 1*, 79–105.

Muller, P., S. Afantenos, P. Denis, and N. Asher (2012, December). Constrained decoding for text-level discourse parsing. In *Proceedings of COLING 2012*, Mumbai, India, pp. 1883–1900. The COLING 2012 Organizing Committee.

Perret, J., S. Afantenos, N. Asher, and M. Morey (2016, June). Integer linear programming for discourse parsing. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, pp. 99–109. ACL.

Polanyi, L. (1985). A theory of discourse structure and discourse coherence in papers from the general session at the twenty-first regional meeting. In *CLS. Papers from the General Session at the Regional Meeting of the Chicago Linguistics Society*, Volume 21, pp. 306–322.

Potts, C. (2005). *The logic of conventional implicatures*. Number 7. Oxford University Press on Demand.

Quintilian (1963). *Institutio oratoria*, Volume 3 of *Loeb Classical Library*. Cambridge, Massachusetts: Harvard University Press. H. E. Butler, translator.

Reese, B. and N. Asher (2007). Prosody and the interpretation of tag questions. In *Proceedings of Sinn und Bedeutung*, Volume 11, pp. 448–462.

Reese, B. J. (2007). *Bias in questions*. Ph. D. thesis.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics 5*, 6–1.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua 117*(6), 1034–1056.

Simons, M. (2019, September). The status of main point complement clauses.

Simons, M., J. Tonhauser, D. Beaver, and C. Roberts (2010). What projects and why. In *Semantics and linguistic theory*, Volume 20, pp. 309–327.

Stojnic, U., M. Stone, and E. Lepore (2013). Deixis (even without pointing). *Philosophical Perspectives 27*, 502–525.

Syrett, K. and T. Koev (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics 32*(3), 525–577.

Von Fintel, K. (2004). Would you believe it? the king of france is back! (presuppositions and truth-value intuitions). In M. Reimer and A. Bezuidenhout (Eds.), *Descriptions and beyond*. Clarendon Press.

Webber, B. L. (1988). Tense as discourse anaphor. *Technical Reports (CIS)*, 441.

Winograd, T. (1972). *Understanding Natural Language*. New York: Academic Press.