Discourse relations as descriptions of an algorithm for perception, action and theorem-proving

Alistair Knott

Abstract

This paper proposes a new framework for representing the semantics of discourse relations: they are taken to be descriptions of the operation of an algorithm being carried out by an agent interacting with the world. The algorithm interleaves perception, action and theorem proving; it is the kind of algorithm, in fact, which must be assumed to underlie the behaviour of any agent. Discourse relations, and the text units which they comprise, are to be thought of as describing stages (or sequences of stages) in this algorithm. The agent executing the algorithm can be the protagonist in a narrative discourse, or one of the conversational participants in a dialogue. We suggest a number of advantages that emerge from thinking about the semantics of discourse relations in these terms.

1 The presuppositional model of discourse relations

Discourse relations are commonly modelled as dynamic constructs. In some of the earliest dynamic conceptions, inspired by the functional definitions of RST (Mann and Thompson, 1988), relations were characterised as planning operators, and defined in terms of their preconditions and postconditions (see e.g. Hovy, 1988; Moore and Paris, 1989). More recently, a conception of relations as presuppositional constructions has been developed (see among others Lagerwerf, 1998; Jayez and Rossari, 1998; Webber *et al*, 1999), which allows relations to be represented within the framework of dynamic semantics.[1]

Theories of dynamic semantics (e.g. Kamp and Reyle, 1993; Veltman, 1996) see discourse as an ordered collection of sentences, with each sentence being added to the representation of a discourse in turn. The semantics of a sentence can be modelled as a function, from representations prior to its addition to representations after its addition, this function being undefined in cases where the prior representation does not contain the presupposed information. (In the case of Kamp and Reyle's theory, some modifications need to be made before this picture of sentences holds; see Muskens, 1996.) The proposal that discourse relations can also have presuppositions allows a similar characterisation: a portion of text containing a relation can be modelled as a function from the representation of the context prior to its addition to the context after its addition. Thinking about relations as presuppositional departs from the 'standard' framework for dynamic semantics only in that it allows the units which are being added to correspond to units larger than single sentences.

In this paper, I will propose some more far-reaching extensions to the dynamic model, which I argue make it more suitable for representing certain discourse relations. These extensions will be motivated by considering a simple example text, given in Section 2, and they are presented in Section 3.

2 The time-course of events underlying narrative *but*

Consider the two-clause text in Example 1.

(1) Jim's paper had just been rejected, but he didn't look upset.

[1] Asher and Lascarides (in press) also propose a dynamic framework (SDRT) in which discourse relations are represented. But in this framework, presupposition is not used to represent the semantics of the relations themselves.

The relation signalled by *but* in this example is typically analysed as SEMANTIC: i.e. it is taken to hold between the eventualities described by the two clauses, rather than between utterances or intentions associated with the clauses themselves. (See e.g. Sanders *et al*, 1992; Knott, 1999 for a presentation of the relevant distinctions.)

There are a number of things we must explain about this text. Firstly, we have to explain what it is about the connective *but* that allows it to appear in the context of these two clauses. The standard presuppositional story (e.g. Lagerwerf, 1998; Knott, 1996) is that P *but* $\neg Q$ presupposes a context where a defeasible rule $\forall x P' > Q'$ is present, where P and Q are copies of P' and Q' with any occurrences of the variable x replaced by some particular entity. (The rule for Example 1 would be something like $\forall x\ has\_paper\_rejected(x) > upset(x)$.) It can then be taken to assert that, in the situation being described, the rule is defeated.

Secondly, we have to explain why it is that, in RST's terminology, *he didn't look upset* is the **nucleus** of the relation; in other words, why this clause is regarded as being more central to the purposes of the text than the latter. Note that the presuppositional explanation as given above does not address this second question at all.

Thirdly, we have to explain why, given that this is a narrative text, there should be any *need* for default inferences. In a narrative scenario, the hearer is disposed to believe everything the speaker says, and clearly the facts of the case are already known by the speaker; it is therefore not clear *whose* default rule is being referred to in the presuppositional account, and why this rule needs to be invoked. In fact, this problem is related to a perennial problem for SEMANTIC relations. What does it mean for a default rule to hold 'between events in the world'? It seems like a category error to suggest that default rules can hold between anything except propositions.

The account proposed in this paper aims to address the above problems. It rests on a number of assumptions. A first assumption is that, as well as a speaker and hearer, a third agent (who we will call the protagonist) needs to feature in the representation of the clauses in Example 1. It is the protagonist who observes that Jim's paper has been rejected, and that Jim does not look upset: the clauses *Jim's paper had been rejected* and *he didn't look upset* are to be thought of as descriptions of these observations, rather than as descriptions directly of the world. This is a way of expressing the suggestion that the text should be interpreted from the *perspective* of the protgonist. (See Caenepeel, 1989; J. Sanders, 1994 for extended treatments of the notion of perspective in narrative.)

A second assumption is that the defeasible inference implicit in Example 1 is a description of an inference made by the protagonist, from the fact that Jim's paper had been rejected to the fact that Jim is (likely to be) upset. This inference occurs, naturally, after the protagonist has observed the former fact, and before she has observed that Jim does not look upset. This account explains how it is that we can specify a defeasible rule as a presupposition of the text, even in a narrative register of text where neither speaker nor hearer is in a position to question the facts being presented.

A third assumption is that the inference made by the protagonist occurs in pursuit of a *goal* to discover a particular piece of information, and that it is important to represent this goal as well as the inference itself. In the present example, it is natural to suggest that the protagonist's main concern is to determine whether Jim is upset. We can then envisage that she decides to do this *first* by drawing inferences from facts she already knows—perhaps because the opportunity to observe at first-hand has not yet arisen, or perhaps because she happens recently to have entertained a relevant fact (the paper being rejected)—and *then* by some other method, such as direct observation. The suggestion that the agent has an 'epistemic' goal (i.e. a goal to discover a certain piece of information) is independently plausible, given that the possible observations and inferences that an agent can make at any given moment are massively unconstrained in the absence of such a goal. But note that it also has a linguistic justification in the present case, as it allows us to explain why it is that the second clause in the example counts in RST's terms as the nucleus of the passage. We can think of the whole passage as describing the protagonist's pursuit of the goal to discover how Jim is feeling, and the second clause as describing the point at which this goal is eventually achieved. This goal-based conception of nuclearity is in the spirit of both

Mann and Thompson (1988) and Grosz and Sidner (1986).[2]

In summary, the time-course of the mental processes underlying our representation of Example 1 can be represented as follows. (P denotes the protagonist.)

1. P posts a goal to discover how Jim is feeling.
2. P decides to pursue this goal by inference from known facts about Jim.
3. P entertains a known fact, that Jim's paper has just been rejected.
4. P infers (defeasibly) that Jim is upset.
5. P decides to pursue the goal by a second route: direct observation.
6. P discovers that Jim does not look upset.
7. P deletes the defeasibly-reached conclusion.

3 Outline of a framework for representing discourse units

In the previous section, we suggested that some of the problems of representing a complex sentence expressing a narrative *but* relation can be addressed by making reference to a particular sequence of mental processes in a hypothesised protagonist. In order to formulate this claim more precisely, we must specify a general semantic framework within which the sequence of processes is to be understood. We must also say more clearly what it means for a unit of discourse, whether simple or complex, to be denoted by a description of one or more of these mental processes.

3.1 An algorithm underlying discourse representations

The framework we have in mind is based on the assumption that the protagonist is executing a general-purpose *algorithm* which governs her interactions with the world, the stages of which include instructions to post goals, to gather perceptual data, and to perform theorem-proving operations. Given this assumption, the sequence of mental processes described above can be seen as a description of the operations of this algorithm over the course of a certain time period. Note that we are not suggesting that the algorithm itself is being represented in the text. What the text represents is the operation of the algorithm in a particular set of circumstances: it is thus influenced both by what these circumstances are and by what the algorithm is.

The proposal to represent text units as operations in an algorithm is at the heart of the programme of dynamic semantics, and thus that there is a large literature describing the advantages to this conception of text units: see e.g. Muskens *et al* (1997) for a review. We will describe the relationship of the present proposal with an existing dynamic theory in Section 3.2. But since we are primarily concerned with discourse relations here, we should note one particular advantage. An algorithm can be thought of as specifying legal sequences of operations; i.e. what operations can succeed one another. The resulting determinism gives us a useful way to think about discourse coherence. It is axiomatic for discourse theorists that the sentences in a text do not succeed one another randomly. It should be equally clear that the mental processes described above do not succeed each other randomly, if we assume a rational protagonist. Suggesting that the protagonist is running an algorithm is just a concrete way of expressing this latter point: the algorithm is a specification of the various possible ways for moving from one mental operation to another. If we assume that successive text units describe successive stages in this algorithm, then the algorithm can effectively function to describe constraints on what counts as a coherent text.

[2]Note that the second method for obtaining the goal need not be direct observation; it might be a second attempt to reach a conclusion by reasoning. Note also that if an alternative conclusion is reached by the second method, there is no requirement that the second conclusion is preferred: to explicitly signal a situation where neither method is preferred, an equivocal connective like *then again* would probably be preferred.

This conception of coherence bears an interesting relation to a conception which is widespread in theories of discourse, namely that a text is coherent because the mental representation it gives rise to in the hearer is coherent. Our proposal is likewise expressed in cognitive terms. But notice that we introduce a third agent, the protagonist, into the picture, and express coherence in terms of the mental representations of this protagonist, rather than those of the reader/hearer.

It is worth making a methodological comment at this point. The algorithm which we have been discussing can naturally be specified at varying levels of abstraction: for instance, to actually build an agent that executes operations 1–7 in the above sequence would require the implementation of a defeasible theorem-prover, a visual perception module and a meta-level planner to invoke these two systems. But to begin to describe the linguistic phenomena in Example1, we don't need this level of detail; in fact, we are not envisaging anything beyond the relatively simple folk-psychological model we have already given, in which goals are posted, and theorem-proving is performed and perceptually tested. The details of the algorithm only need to be given insofar as they play a part in explaining the linguistic structures to be represented.

3.2 Modelling the algorithm using dynamic semantics

As should be clear, the model outlined above falls into the class of dynamic theories of semantics. The purpose of the present section is to outline how the present proposal differs from existing dynamic models.

In the dynamic semantics of Veltman (1996), the meaning of a sentence is defined as a function from one **information state** of an agent to another. An information state σ is a representation of the set of situations the agent regards as possible. The **update value** of a sentence representing an atomic proposition p is defined as a function which takes an information state σ_1, and returns another information state σ_2 in which some of the possible situations in σ_1—namely, those which do not contain p—have been eliminated. The standard logical operations of negation, implication etc can all be defined in terms of set-theoretic operations on the set σ_1.

This position needs to be modified in several ways. Most importantly, an operation in our algorithm is not—at least in the general case—the action of interpreting a sentence, as it is in the above model. In fact, our operations are not (in the general case) linguistic actions at all; rather, they are 'cognitive' actions. The operations in Veltman's theory can be imagined as specified by an algorithm whose function is to interpret a sequence of sentences. The operations envisaged in our model are much more general-purpose, and include things like the posting of a goal, the execution of a piece of theorem-proving, the asking of and getting aswers to questions, or even the entertaining of a known fact. But note that we can still define them as functions from one state of the agent to another. Consequently, we must ensure that the update function which characterises an operation in our model allows the representation of a much wider range of operations than simply the interpretation of sentences. Given that an operation is defined as a function that transforms one state into another, this means that the states in our model must specify not just what the agent knows or expects, but a much wider range of things, including (i) the agent's goals (in particular her epistemic goals), (ii) what goal the agent regards as immediately to be pursued, and (iii) what the outcome of an action in pursuit of some goal is. We will elaborate a little on these requirements in the next section.

Secondly, we must allow that the cognitive states between which operations move us are not necessarily those of the hearer, as in Veltman, but (as in our treatment of Example 1) can be those of an agent being described in the text. This emphasises the fact that we are not modelling the interpreter of a text, but the interpretation itself. Where the agent is the protagonist in a narrative, an interpreter's job will be to infer what the relevant stages of the agent's algorithm are—a task in which plan recognition is likely to feature prominently.

Thirdly, we must envisage a more complex relationship between operations in the algorithm and units in the text than in Veltman's model. For Veltman, each update operation 'consumes' a single sentence. In our model, as the analysis of Example 1 demonstrates, there are more operations than there are clauses. We can describe the mapping we envisage between operations and textual elements as follows:

- Some operations in the algorithm are represented by the discourse relation in which the clauses feature, rather than by the clauses themselves. In our analysis of Example 1, for instance, stages 4 and 7 do not feature in the representation of either of the atomic clauses.
- A single clause in the text can be represented by more than one operation. For instance, stages 2 and 3 are both involved in representing the first clause, and stages 5 and 6 are both involved in representing the second.
- A clause can be represented by *non-consecutive* operations in the algorithm. For instance, stage 1 should feature in the representation of the second clause, as well as in that of the first.
- Finally, a single operation in the algorithm can feature in the representation of more than one discourse element. (In fact, this requirement is necessary for the algorithm to describe how one discourse element links to another.)

The complexity of this mapping between clauses and operations means it is not yet clear how to specify a *compositional* dynamic account of the text in Example 1, of the kind advocated by Muskens (1996). However, the individual operations in the algorithm certainly seem promising ingredients for such a compositional account, as they would be useful in the representation of a number of discourse phenomena. For instance, it is specified in our analysis of Example 1 that the information in the first clause is obtained by the agent by consulting known facts and that the information in the second clause is obtained by direct observation: these statements could feature in an account of the different aspectual types (past perfect and simple past) of the two clauses. If we assume that the relevant stages of the algorithm occur at the Reichenbachian reference time (c.f. Reichenbach, 1947), then it follows automatically that the event time of a remembered event is prior to the reference time, while that of a perceived event is the same as the reference time, which is the characterisation we want. And, given that an extended piece of text is represented by a sequence of *consecutive* stages in the agent's algorithm, the same conception of reference time allows an explanation of why a sentence's reference time can be determined anaphorically from the preceding context. To take another discourse phenomenon: the fact that an indicative sentence is represented in relation to an agent's goal to discover some fact provides a framework for an account of its topic-focus structure. For instance, the representation of the second clause in Example 1 (*[he]$_T$ [didn't look upset]$_F$*) involves representing the posting of an epistemic goal and its satisfaction, stages which closely mirror the question-answer structures which underlie topic-focus structure in Rooth's alternative semantics (Rooth, 1985; 1995).

3.3 Representing epistemic actions in the dynamic model

A key difference between Veltman's scenario and ours is that the agent in our model is active, while Veltman's agent is passive. (Or at least, the goal of Veltman's agent to interpret the next sentence is hardwired into the model.) Veltman's sentence-interpreting agent is not in a position to choose which fact to take on board next: this is determined by the order of sentences in the text, which is not under her control. In our account of Example 1, the agent's epistemic goal—or question—plays a crucial role; we must therefore allow questions and answers to be represented in the model. We will not attempt to give a formal definition of the operations needed in our model. But looking at the representation of epistemic actions is particularly important, and should give an idea of the kind of definitions we have in mind.

As Higginbotham (1995) notes, there are two important things to capture in the representation of a question. Firstly, we have to capture the dynamic nature of questions, as actions which transform a state where a goal to discover something is 'in force' into a state where the thing has been discovered. Secondly, we need to ensure that the subsequent state contains not only the new information, but also the relationship between this new information and the original goal. Generalising from questions to epistemic actions, we can add a third requirement to these two, namely that we are able to distinguish between epistemic actions which are operations in memory to retrieve known facts (as in clause 1 of Example 1), and those which are perceptions of the world (as in clause 2). The representation we sketch here aims to reconcile these three goals.

To begin with the dynamic characterisation: we must be able to specify that at a certain point during the execution of the algorithm, some epistemic goal G becomes the current goal. We can represent this by specifying that the state contains a variable *current_goal*, which at different points in the algorithm's operation has different values. We must likewise be able to represent that the system is in a state where a question has just been answered; we therefore provide another program variable *current_content*, which can be filled by either ? (denoting that no answer has been provided) or a particular value (denoting that this value is the answer). We want the epistemic action operation to take us from a state where $current\_goal = G$ and $current\_content =?$, to a state where *current_content* has a particular value, and $current\_goal \neq G$. In general, an operation will be defined as an axiom specifying the type of state in which it can be executed, and the type of state which results its execution. The concept of a current goal allows us to match each operation to a specific goal. If we assert for each operation O_n that it can only be executed in a state which includes the fact $current\_goal = G_n$, we ensure a one-to-one correspondence between goal states and operations.

There are two problems with the solution so far outlined. Firstly, since the question represented by *current_goal* has to be changed in the state resulting from the operation, this latter state does not actually represent the fact itself: we will have successfully represented the fact that the asking and answering of a question is a dynamic process, but failed to capture the other aspect of an epistemic action, namely that it adds a new fact to the knowledge base. Secondly, is is surely wrong to characterise the epistemic action as an *axiom*: were it an axiom, the result would already be known.

To answer both these questions, we first distinguish two types of proposition in the system: *accessible* facts and *inaccessible* ones. The former type of facts represent what the agent knows; the second type represent how she acts. We then envisage that, as a second effect of the question-answer operation, a *trace* of the operation itself is created, and stored as an axiom in a database, indexed with the situation S in which the operation was executed. We then specify that only these database axioms are *accessible*, all other facts being *inaccessible*. Finally, we specify a third effect of the operation g_1, which is that an (*inaccessible*) axiom representing a new epistemic operation g' to *access* the database is created, whose goal specifies the question G_1 as well as the situation S. The distinction between the operation g itself and the newly created operation g' reflects the need to distinguish between perceptual and memory-accessing epistemic operations.

4 Conclusion

The main aim of this paper is to argue for a framework in which units of text are represented as descriptions of an algorithm being executed by an agent, in which operations representing the posting of goals and the execution of sensorimotor and cognitive actions are specified at a finer level of granularity than is commonly assumed. The final section gives a very preliminary indication of what some of the ingredients of this account might be, just for the sake of concreteness, though we should stress that the specification of the algorithm is a major task in its own right. Our only claim at present is that the hypothesis that discourse units should be represented as descriptions of a trace of the algorithm suggests an interesting new approach to the problem.

References

Asher, N. and Lascarides, A. (in press). The semantics and pragmatics of presupposition. *Journal of Semantics*.

Caenepeel, M. (1989). *Aspect, Temporal Ordering and Perspective in Narrative Fiction*. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, pages 175–203.

Higginbotham, J. (1995). The semantics of questions. In S. Lappin, editor, *The handbook of contemporary semantic theory*, pages 361–383. Blackwell, Oxford.

Hovy, E. H. (1988). Planning coherent multisentential text. In *Proceedings of the 26th ACL Conference*, pages 163–169, Buffalo.

Jayez, J. and Rossari, C. (1998). Discourse relations versus discourse marker relations. In *Proceedings of the ACL Workshop on Discourse Relations and Discourse Markers*, pages 72–78, Montreal, Canada.

Kamp, H. and Reyle, U. (1993). *From discourse to logic*. Kluwer Academic Publishers, Dordrecht.

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

Knott, A. (1999). Semantic and pragmatic relations and their intended effects. In T. Sanders, J. Schilperoord, and W. Spooren, editors, *Text representation: linguistic and psycholinguistic aspects*. Benjamins, Amsterdam.

Lagerwerf, L. (1998). *Causal connectives have presuppositions*. Ph.D. thesis, Katholieke Universiteit Brabant, The Hague.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: A theory of text organization. *Text*, **8(3)**, 243–281.

Moore, J. D. and Paris, C. L. (1989). Planning text for advisory dialogues. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 203–211.

Muskens, R. (1996). Combining Montague Semantics and discourse representation. *Linguistics and Philosophy*, **19**, 143–186.

Muskens, R., van Benthem, J., and Visser, A. (1997). Dynamics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 587–648. Elsevier, Amsterdam.

Reichenbach, H. (1947). *Elements of symbolic logic*. University of California Press, Berkeley, CA.

Rooth, M. (1985). *Association with Focus*. Ph.D. thesis, University of Massachusets at Amherst.

Rooth, M. (1995). Focus. In S. Lappin, editor, *The handbook of contemporary semantic theory*, pages 271–297. Blackwell, Oxford.

Sanders, J. (1994). *Perspective in Narrative Discourse*. Ph.D. thesis, Tilburg University.

Sanders, T. J. M., Spooren, W. P. M., and Noordman, L. G. M. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, **15**, 1–35.

Veltman, F. (1996). Defaults in update semantics. *Journal of philosophical logic*, **25**, 221–261.

Webber, B., Knott, A., Stone, M., and Joshi, A. (1999). Anaphoric presupposition and discourse connectivity. In *Proceedings of the 37th Annual meeting of the Association for Computational Linguistics*, College Park, Maryland. (To appear.).