Extending a Formal and Computational Model of Rhetorical Structure Theory with Intentional Structures à la Grosz and Sidner

Daniel Marcu

Information Sciences Institute and Department of Computer Science
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292-6601
marcu@isi.edu

Abstract

In the last decade, members of the computational linguistics community have adopted a perspective on discourse based primarily on either Rhetorical Structure Theory or Grosz and Sidner's Theory. However, only recently, researchers have started to investigate the relationship between the two perspectives. In this paper, we use Moser and Moore's (1996) work as a departure point for extending Marcu's formalization of RST (1996). The result is a first-order axiomatization of the mathematical properties of text structures and of the relationship between the structure of text and intentions. The axiomatization enables one to use intentions for reducing the ambiguity of discourse and the structure of discourse for deriving intentional inferences.

1 Motivation

In the last decade, members of the computational linguistics community have adopted a perspective on discourse based primarily on either Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) or Grosz and Sidner's Theory (GST) (Grosz and Sidner, 1986).

In GST, the linguistic constituents are called *discourse segments (DS)* and the linguistic discourse structure is explicitly stipulated to be a tree of recursively embedded discourse segments. Each discourse segment is characterized by a primary intention, which is called *discourse segment purpose (DSP)*. GST identifies only two kinds of intention-based relations that hold between the *DSPs* of two discourse segments: *dominance* and *satisfaction precedence*. When a discourse segment purpose DSP_1 that characterizes discourse segment DS_1 provides part of the satisfaction of a discourse segment purpose DSP_2 that characterizes discourse segment DS_2, with DS_1 being embedded in DS_2, it is said that there exists a dominance relation between DSP_2 and DSP_1, i.e., DSP_2 *dominates* DSP_1. If the satisfaction of DSP_1 is a condition of the satisfaction of DSP_2, it is said that DSP_1 *satisfaction-precedes* DSP_2.

RST has a richer ontology of relations than GST: intentional and semantic rhetorical relations are considered to hold between non-overlapping textual spans. Most of these relations are asymmetric, i.e., they distinguish between their associated nuclei, which express what is most essential to the writer's purpose, and their satellites, which support the nuclei. In RST, the linguistic discourse structure is modeled recursively as a tree of related segments. Hence, unlike GST, where relations are considered to hold between the *DSPs* associated with *embedded segments*, relations in RST hold between *adjacent, non-overlapping segments*.

Because RST has traditionally been applied to build discourse trees of finer granularity than GST, we will use it here as the starting point of our discussion. Assume, for example, that we are given the following text (in which the elementary textual units are labelled for reference).

(1) [No matter how much one wants to stay a non-smoker,[A1]] [the truth is that the pressure to smoke in junior high is greater than it will be any other time of one's life.[B1]] [We know that 3,000 teens start smoking each day,[C1]] [although it is a fact that 90% of them once thought that smoking was something that they'd never do.[D1]]

Assume for the moment that we do not analyze this text as a whole, but rather, we determine what rhetorical relations could hold between every pair of elementary units. When we apply, for example, the definitions proposed by Mann and Thompson (1988), we obtain the set given below.[1]

$$
(2) \quad \left\{
\begin{array}{l}
rhet\_rel(\text{JUSTIFICATION}, A_1, B_1) \\
rhet\_rel(\text{JUSTIFICATION}, D_1, B_1) \\
rhet\_rel(\text{EVIDENCE}, C_1, B_1) \\
rhet\_rel(\text{CONCESSION}, D_1, C_1) \\
rhet\_rel(\text{RESTATEMENT}, D_1, A_1)
\end{array}
\right.
$$

These relations hold because the understanding of both A_1 (teens want to stay non-smokers) and D_1 (90% of the teens think that smoking is something that they would never do) will increase the reader's readiness to accept the writer's right to present B_1 (the pressure on teens to start smoking is greater than it will be any other time of their lives); the understanding of C_1 (3000 teens start smoking each day) will increase the reader's belief of B_1; the recognition of D_1 as something compatible with

---

[1]Throughout this paper, we use the convention that rhetorical relations are represented as sorted, first-order predicates having the form $rhet\_rel(name, satellite, nucleus)$. Multinuclear relations are represented as predicates having the form $rhet\_rel(name, nucleus_1, nucleus_2)$.
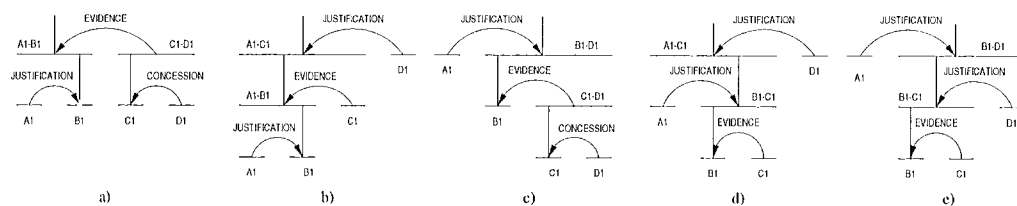
Figure 1: The set of all RS-trees that can be built for text (1).

the situation presented in c_1 will increase the reader's negative regard for the situation presented in c_1; and the situation presented in D_1 is a restatement of the situation presented in A_1.

Marcu (1996) has shown that on the basis of *only* the rhetorical judgments in (2) and without considering intentions, there are five valid RS-trees that one can build for text (1) (see figure 1). What happens though when we consider intentions as well? Moore and Pollack (1992) have already shown that different high-level intentions yield different RS-trees. But how do we formalize the relationship between intentions and rhetorical structures? For example, how can we use the discourse trees in figure 1 in order to determine the primary intention associated with each analysis? And how can we determine what would be the corresponding dominance relations in a GST account of the same text?

Consider also a slightly different problem: assume that besides rhetorical judgments, such as those shown in (2), one can also make intentional judgments. For example, assume that one is interested in an interpretation in which one knows that the *DSP* of segment $[A_1, D_1]$, which contains all units from A_1 to D_1, *dominates* the *DSP* of segment $[c_1, D_1]$. Then what is the primary intention of the text in that case? And how many discourse trees are both valid and consistent with that intentional judgment? Neither RST nor GST can answer these questions *on their own*. However, a unified theory can. In this paper, we provide such a theory.

2 The limits of Moser and Moore's approach

In a recent proposal, Moser and Moore (1996) argued that the primary intentions in a GST representation can be derived from the nuclei of the corresponding RST representation. Although their proposal is consistent with the cases in which each textual span is characterized by an explicit nucleus that encodes the primary intention of that span (as in the case of text (1)), it seems that an adequate account of the correspondence between GST and RST is somewhat more complicated. For example, in the case of text (3) below, whose RST analysis is shown in figure 2, we cannot apply Moser and Moore's approach because we can associate the primary intention of discourse segment $[A_2, B_2]$ neither to unit A_2 nor to unit B_2.
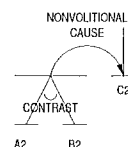
(3) [John wanted to play squash with Janet,[A2]] [but he

Figure 2: A rhetorical analysis of text (3).

also wanted to have dinner with Suzanne.[B2]] [He went crazy.[C2]]

In Grosz and Sidner's terms, we can say that the primary intention of segment $[A_2, B_2]$ is (Intend writer (Believe reader "John wanted to do two things that were incompatible")). But in order to recognize this relation, we need to recognize that the two desires given in units A_2 and B_2 are incompatible, which is captured by the CONTRAST relation that holds between the two units. In other words, the intention associated with segment $[A_2, B_2]$ is a function both of its nuclei, A_2 and B_2, *and* of the rhetorical relation of CONTRAST that holds between them.

In this paper, we generalize this observation by making use of the compositionality criterion proposed in (Marcu, 1996), which stipulates that if a rhetorical relation holds between two textual spans, a similar relation also holds between two salient constructs of those spans.[2] Similarly, we will assume that the primary intention of a discourse segment is not given by the nucleus of the corresponding relation but rather that it depends on the corresponding relation *and* the salient constructs associated with that segment.

3 Melding text structures and intentions

3.1 Formulation of the problem

Formally, the problem that we want to solve is the following. Given a sequence of textual units $U = u_1, u_2, \ldots, u_N$, a set RR of rhetorical relations that hold among these units, and a set of intentional judgments RI that pertain to the same units, find all legal discourse structures (trees) of U, and determine the dominance, satisfaction-precedence relations, and primary intentions of each span of these trees.

Following (Marcu, 1996), we use the predicates $position(u_i, j)$ and $rhet\_rel(name, s, n)$ with the fol-

---

[2]Section 3 discusses in detail how the salient constructs are determined.

lowing semantics: the predicate $position(u_i, j)$ is true for a textual unit u_i in sequence U if and only if u_i is the j-th element in the sequence; the predicate $rhet\_rel(name, u_i, u_j)$ is true for textual units u_i and u_j with respect to rhetorical relation $name$, if and only if the definition provided by RST for rhetorical relation $name$ applies to textual units u_i, in most cases a satellite, and u_j, a nucleus. In order to enable discourse problems to be characterized by rhetorical judgments that hold between large textual spans as well, we use predicate $rhet\_rel\_ext(name, s_s, s_e, n_s, n_e)$. This predicate is true for textual spans $[s_s, s_e]$ and $[n_s, n_e]$ with respect to rhetorical relation $name$ if and only if the definition of rhetorical relation $name$ applies for the textual span that ranges over units s_s–s_e, in most cases a satellite, and textual spans that ranges over units n_s–n_e, a nucleus.[3]

From a rhetorical perspective, text (1) is described at the minimal unit level by the relations given in (2) and (4) below.

$$(4) \quad \left\{ \begin{array}{c} position(A_1, 1), position(B_1, 2), \\ position(C_1, 3), position(D_1, 4) \end{array} \right.$$

The intentional judgments RI are given by the following functions and predicates:

- The predicate $dom(l_1, h_1, l_2, h_2)$ is true whenever the DSP of discourse segment/span $[l_1, h_1]$ dominates the DSP of discourse segment $[l_2, h_2]$. A dominance relation is well-formed if segment $[l_2, h_2]$ is a proper subsegment of segment $[l_1, h_1]$. i.e., $l_1 \leq l_2 \leq h_2 \leq h_1 \wedge (l_1 \neq l_2 \vee h_1 \neq h_2)$.

- The predicate $satprec(l_1, h_1, l_2, h_2)$ is true whenever an intentional satisfaction-precedence relation holds between the DSPs of segments $[l_1, h_1]$ and $[l_2, h_2]$. A satisfaction-precedence relation is well-formed if the segments do not overlap.

- The oracle function $f_I(r, x_1, \ldots, x_n)$ takes as arguments a rhetorical relation r and a set of textual units, and returns the primary intention that pertains to that relation and those units. For example, in the case of segment $[A_2, B_2]$ in text (3), the oracle function $f_I(\text{CONTRAST}, A_2, B_2)$ is assumed to return a first-order object whose meaning can be glossed as "inform the reader that John wanted to do two things that were incompatible". And the oracle function $f_I(\text{EVIDENCE}, B_1)$ associated with segment $[A_1, D_1]$ in text (1) is assumed to return a first-order object whose meaning can be glossed as "increase the reader's belief that the pressure to smoke in junior high is greater than it will be any other time of one's life".

Without restricting the generality of the problem, discourse structures are assumed to be binary trees. In our formalization, each node of a discourse structure is characterized by four features: the *status* (nucleus or satellite), the *type* (the rhetorical relations that hold between

the text spans that that node spans over), the *promotion set* (the set of units that constitute the most "salient" (important) part of the text that is spanned by that node), and the *primary intention*. By convention, for each leaf node, the type is LEAF, the promotion set is the textual unit to which it corresponds, and the primary intention is that of *informing* the content of that unit. For example, a representation of the tree in figure 1.a that makes explicit the features of all spans that play an active role in the final representation is given in figure 3. In general, the salient units are computed using the compositionality criterion proposed in (Marcu, 1996), i.e, they are given by the union of the salient units of the immediate subordinated nuclei. Similarly, the primary intentions are a function of the rhetorical relation (type) and salient units of each span.

The status, type, promotion set, and primary intention that are associated with each node in a discourse tree provide sufficient information for a full description of an instance of a tree structure. Given the linear nature of text and the fact that we cannot predict in advance where the boundaries between various segments will be drawn, we should provide a methodology that permits one to enumerate all possible ways in which a tree could be built on the top of a linear sequence of elementary discourse units. The solution we use relies on the same intuition that constitutes the foundation of chart parsing: just as a chart parser is capable of considering all possible ways in which different words in a sentence could be clustered into higher-order grammatical units, so our formalization is capable of considering all the possible ways in which different segments could be joined into discourse trees.

Let $span_{i,j}$, or simply $[i, j]$, denote a text span that includes all the elementary discourse units between position i and j. Then, if we consider a sequence of discourse units $u_1, u_2, \ldots, u_n$, there are n ways in which spans of length one could be built, $span_{1,1}, span_{2,2}, \ldots, span_{n,n}$; $n - 1$ ways in which spans of length two could be built, $span_{1,2}, span_{2,3}, \ldots, span_{n-1,n}$; $n - 2$ ways in which spans of length three could be built, $span_{1,3}, span_{2,4}, \ldots, span_{n-2,n}$; $\ldots$; and one way in which a span of length n could be built, $span_{1,n}$. Since it is impossible to determine a priori the spans that will be used to make up a discourse tree, we will associate with each span that could possibly become part of a tree a status, a type, promotion, and primary intention relation and let discourse and intentional constraints determine the valid discourse trees. In other words, we want to determine from the set of $n + (n - 1) + (n - 2) + \ldots + 1 = n(n + 1)/2$ potential spans that pertain to a sequence of n discourse units, the subset that adheres to some constraints of rhetorical and intentional well-formedness. For example, for text 1, there are $4 + 3 + 2 + 1 = 10$ potential spans, i.e., $span_{1,1}, span_{2,2}, span_{3,3}, span_{4,4}, span_{1,2}, span_{2,3}, span_{3,4}, span_{1,3}, span_{2,4}$, and $span_{1,4}$, but

---

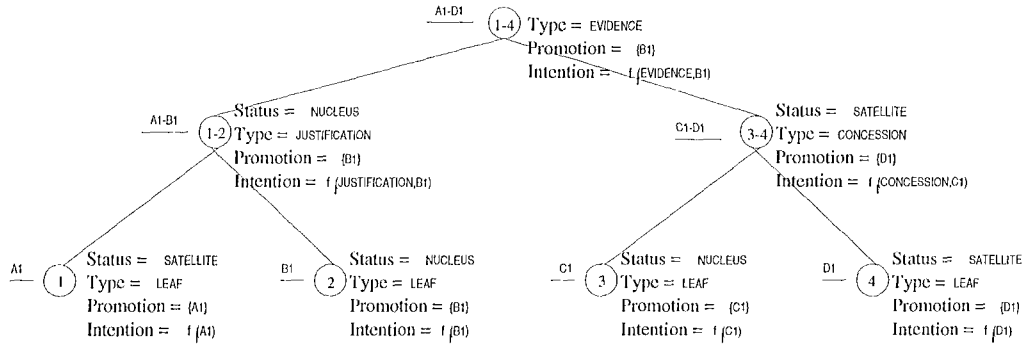[3]The s and e subscripts correspond to starting and ending positions.

Figure 3: A representation of tree 1.a that includes the status, type, promotion, and primary intention features that characterize every node that does not have a NONE status. The numbers associated with each node denote the limits of the text span that that node characterizes.

only seven of them play an active role in the representation given in figure 1.a, i.e., $span_{1,1}, span_{2,2}, span_{3,3}, span_{4,4}, span_{1,2}, span_{3,4}$, and $span_{1,4}$.

To formalize the constraints that pertain both to RST and GST, we thus assume that each potential span $[l, h]$ is characterized by the following predicates:

- $S(l, h, status)$ provides the status of span $[l, h]$, i.e., the text span that contains units l to h; $status$ can take one of the values NUCLEUS, SATELLITE, or NONE according to the role played by that span in the final discourse tree. For example, for the tree depicted in figure 3, some of the relations that hold are: $S(1, 2, \text{NUCLEUS}), S(3, 4, \text{SATELLITE}), S(1, 3, \text{NONE})$.

- $T(l, h, relation\_name)$ provides the name of the rhetorical relation that holds between the text spans that are immediate subordinates of span $[l, h]$ in the discourse tree. If the text span is not used in the construction of the final tree, the type assigned is NONE. For example, for the tree in figure 3, some of the relations that hold are: $T(1, 1, \text{LEAF}), T(1, 2, \text{JUSTIFICATION}), T(3, 4, \text{CONCESSION}), T(1, 3, \text{NONE})$.

- $P(l, h, unit\_name)$ provides one of the set of units that are salient for span $[l, h]$. The collection of units for which the predicate is true provides the promotion set of a span, i.e., all units that are salient for that span. If span $[l, h]$ is not used in the final tree, by convention, the set of salient units is NONE. For example, for the tree in figure 3, some of the relations that hold are: $P(1, 1, \text{A}_1), P(1, 2, \text{B}_1), P(1, 3, \text{NONE}), P(3, 4, \text{D}_1)$.

- $I(l, h, intention)$ provides the primary intention of discourse span $[l, h]$. The term $intention$ is represented using the oracle function f_I. For example, for the tree in figure 3, some of the rela-

tions that hold are: $I(3, 4, f_I(\text{CONCESSION}, \text{C}_1))$, $I(1, 4, f_I(\text{EVIDENCE}, \text{B}_1)), I(1, 3, \text{NONE})$.

3.2 An integrated formalization of RST and GST

Using the ideas that we have discussed in the previous section, we present now a first-order formalization of discourse structures that makes use both of RST- and GST-like constraints. In this formalization, we assume a universe that consists of the set of natural numbers from 1 to N, where N represents the number of textual units in the text that is considered; the set of names that were defined by Mann and Thompson for each rhetorical relation; the set of unit names that are associated with each textual unit; and four extra constants: NUCLEUS, SATELLITE, NONE, and LEAF. The formalization is assumed to provide unique name axioms for all these constants.

The only function symbols that operate over the assumed domain are the traditional $+$ and $-$ functions that are associated with the set of natural numbers and the oracle function f_I. The formalization uses the traditional predicate symbols that pertain to the set of natural numbers $(<, \leq, >, \geq, =, \neq)$ and eight other predicate symbols: S, T, P and I to account for the status, type, salient units, and primary intention that are associated with every text span; $rhet\_rel$ to account for the rhetorical relations that hold between different textual units; $position$ to account for the index of the textual units in the text that one considers; dom to account for dominance relations; and $satprec$ to account for satisfaction-precedence relations.

Throughout the paper, we apply the convention that all unbound variables are universally quantified and that variables are represented in *lower-case italics* and constants in SMALL CAPITALS. We also make use of the two extra relations, $relevant\_unit$ and $relevant\_rel$. For every text span span $[l, h]$, $relevant\_unit(l, h, u)$ describes the set of textual units that are relevant for that text span, i.e., the units whose positions in the initial sequence are numbers in the interval $[l, h]$. It is only these units that can be used to label the pro-

motion set associated with a tree that subsumes all units in the interval $[l, h]$. For every text span $[l, h]$, $relevant\_rel(l, h, name)$ describes the set of rhetorical relations that are relevant to that text span, i.e., the set of rhetorical relations that span over text units in the interval $[l, h]$ and the set of extended rhetorical relations that span over text spans that cover the whole interval $[l, h]$ (see (Marcu, 1996) for the formal definitions of these relations.)

For example, for text (1), which is described formally in (2) and (4), the following is the set of all $relevant\_rel$ and $relevant\_unit$ relations that hold with respect to text segment $[1, 3]$: $\{relevant\_rel(1, 3, \text{JUSTIFICATION}), relevant\_rel(1, 3, \text{EVIDENCE}), relevant\_unit(1, 3, A_1), relevant\_unit(1, 3, B_1), relevant\_unit(1, 3, C_1)\}$.

The constraints that pertain to the discourse trees that we formalize can be partitioned into constraints related to the domain of objects over which each predicate ranges, constraints related to the structure of the tree, and constraints that relate the structural component with the intentional component. The axioms that pertain to the domains over which predicates S, P, and T range and the constraints related to the structure of the tree are the same as those given by Marcu (1996). For the sake of completeness, in this paper we only enumerate them informally. In contrast, the axioms that pertain to intentions and the relation between structure and intentions are discussed in detail.

Constraints that concern the objects over which the predicates that describe every segment $[l, h]$ of a text structure range (Marcu, 1996, pp. 1072–1073).

• **For every span $[l, h]$, the set of objects over which predicate S ranges is the set $\{\text{NUCLEUS}, \text{SATELLITE}, \text{NONE}\}$.**

• **The status of any discourse segment is unique.**

• **For every segment $[l, h]$, the set of objects over which predicate T ranges is the set of rhetorical relations that are relevant to that span.**

• **At most one rhetorical relation can connect two adjacent discourse spans**

• **The primary intention of a discourse segment is either NONE or is a function of the salient units that pertain to that segment and of the rhetorical relation that holds between the immediate subordinated segments.** Since we want to stay within the boundaries of first-order logic, we express this (see formula (5) below) by means of a disjunction of at most N subformulas, which correspond to the cases in which the span has 1, 2, ..., or N salient units.[4]

---

[4]Formula (5) reflects no preference concerning the order in which rhetorical relations and intentions should be computed (Asher and Lascarides, 1998). It only asserts a constraint on the two.

(5)
$$[(1 \leq h \leq \text{N}) \wedge (1 \leq l \leq h)] \rightarrow$$
$$\{I(l, h, intention_{lh}) \rightarrow$$
$$intention_{lh} = \text{NONE}$$
$$(\exists r, x)[T'(l, h, r) \wedge r \neq \text{NONE} \wedge$$
$$P(l, h, x) \wedge (\forall y)(P(l, h, y) \rightarrow x = y) \wedge$$
$$intention_{lh} = f_I(r, x)] \vee$$
$$(\exists r, x_1, x_2)[T'(l, h, r) \wedge r \neq \text{NONE} \wedge$$
$$P(l, h, x_1) \wedge P(l, h, x_2) \wedge x_1 \neq x_2 \wedge$$
$$(\forall y)(P(l, h, y) \rightarrow (y = x_1 \vee y = x_2)) \wedge$$
$$intention_{lh} = f_I(r, x_1, x_2)] \vee$$
$$\vdots$$
$$(\exists r, x_1, x_2, \ldots, x_N)[T'(l, h, r) \wedge r \neq \text{NONE} \wedge$$
$$x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge \ldots \wedge x_1 \neq x_N \wedge$$
$$x_2 \neq x_3 \wedge \ldots \wedge x_2 \neq x_N \wedge$$
$$\vdots$$
$$x_{N-1} \neq x_N \wedge$$
$$P(l, h, x_1) \wedge P(l, h, x_2) \wedge \ldots \wedge P(l, h, x_N) \wedge$$
$$(\forall y)(P(l, h, y) \rightarrow (y = x_1 \vee \ldots \vee y = x_N)) \wedge$$
$$intention_{lh} = f_I(r, x_1, x_2, \ldots, x_N)]\}$$

• **The primary intention of any discourse segment is unique.**

(6)
$$[(1 \leq h \leq \text{N}) \wedge (1 \leq l < h)] \rightarrow$$
$$[(I(l, h, i_1) \wedge I(l, h, i_2)) \rightarrow i_1 = i_2]$$

• **For every segment $[l, h]$, the set of objects over which predicate P ranges is the set of units that make up that segment** .

Constraints that concern the structure of the discourse trees

• **The status, type, and promotion set that are associated with a discourse segment reflect the compositionality criterion.** That is, whenever a rhetorical relation holds between two spans, either a similar relation holds between the most salient units of those spans or an extended rhetorical relation holds between those spans.

• **Discourse segments do not overlap.**

• **A discourse segment with status NONE does not participate in the tree at all.**

• **There exists a discourse segment, the root, that spans over the entire text.**

(7)
$$\neg S(1, \text{N}, \text{NONE}) \wedge \neg P(1, \text{N}, \text{NONE}) \wedge$$
$$\neg T(1, \text{N}, \text{NONE}) \wedge \neg I(1, \text{N}, \text{NONE})$$

• **The dominance relations described by Grosz and Sidner hold between the DSP of a discourse segment and the DSP of its most immediate subordinated satellite.** This constraint is consistent with Moser and Moore's (1996) discussion of RST and GST. In fact, this is not surprising if we examine the definitions of dominance relation given by Grosz and Sidner and satellite given by Mann and Thompson: a discourse segment purpose DSP_2 dominates a discourse segment purpose DSP_1 if DSP_1 contributes to the satisfaction of the DSP_2. But this is exactly the role that satellites play in RST: they do not express what is most essential for the writer's purpose, but rather, provide supporting information that contributes to the understanding of the nucleus.

The relationship between Grosz and Sidner's dominance relations and Mann and Thompson's distinction between nuclei and satellites is formalized by axioms (8) and (9).

$$
\begin{aligned}
&[(1 \leq h_1 \leq \text{N}) \wedge (1 \leq l_1 \leq h_1) \wedge \\
&(1 \leq h_2 \leq \text{N}) \wedge (1 \leq l_2 \leq h_2)] \rightarrow \\
&\quad \{[\neg S(l_1, h_1, \text{NONE}) \wedge S(l_2, h_2, \text{SATELLITE}) \wedge \\
&\quad l_1 \leq l_2 \leq h_2 \leq h_1 \wedge \\
(8) \quad &\quad \neg(\exists l_3, h_3)(l_1 \leq l_3 \leq l_2 \leq h_2 \leq h_3 \leq h_1 \wedge \\
&\quad\quad (l_3 \neq l_2 \vee h_3 \neq h_2) \wedge \\
&\quad\quad S(l_3, h_3, \text{SATELLITE}))] \rightarrow \\
&\quad dom(l_1, h_1, l_2, h_2)\}
\end{aligned}
$$

$$
\begin{aligned}
&[(1 \leq h_1 \leq \text{N}) \wedge (1 \leq l_1 \leq h_1) \wedge (1 \leq h_2 \leq \text{N}) \wedge \\
(9) \quad &(1 \leq l_2 \leq h_2) \wedge dom(l_1, h_1, l_2, h_2)] \rightarrow \\
&[\neg S(l_1, h_1, \text{NONE}) \wedge S(l_2, h_2, \text{SATELLITE})]
\end{aligned}
$$

Axiom (8) specifies that if segment $[l_2, h_2]$ is the immediate satellite of segment $[l_1, h_1]$, then there exists a dominance relation between the *DSP* of segment $[l_1, h_1]$ and the *DSP* of segment $[l_2, h_2]$. Hence, axiom (8) explicates the relationship between the structure of discourse and intentional dominance. In contrast, axiom (9) explicates the relationship between intentional dominance and discourse structure. That is, if we know that the intention associated with span $[l_1, h_1]$ dominates the intention associated with span $[l_2, h_2]$, then both these spans play an active role in the representation and, moreover, the segment $[l_2, h_2]$ plays a SATELLITE role.

● **The satisfaction-precedence relations described by Grosz and Sidner are paratactic relations that hold between arbitrarily large textual spans.** Nevertheless, as we have seen in the examples discussed in this paper, the fact that a paratactic relation holds between spans does not imply that there exists a satisfaction-precedence relation at the intentional level between those spans. Therefore, for satisfaction-precedence relations, we will have only one axiom, that shown in (10), below.

$$
\begin{aligned}
&[(1 \leq h_1 \leq \text{N}) \wedge (1 \leq l_1 \leq h_1) \wedge (1 \leq h_2 \leq \text{N}) \wedge \\
(10) \quad &(1 \leq l_2 \leq h_2) \wedge satprec(l_1, h_1, l_2, h_2)] \rightarrow \\
&[S(l_1, h_1, \text{NUCLEUS}) \wedge S(l_2, h_2, \text{NUCLEUS})]
\end{aligned}
$$

This specifies that the spans that are arguments of a satisfaction-precedence relation have a NUCLEUS status in the final representation.

4 A computational view of the axiomatization

Given the formulation discussed above, finding the discourse trees and the primary intentions for a text such as that given in (1) amounts to finding a model for a first-order theory that consists of formulas (2), (4), and the axioms enumerated in section 3.

There are a number of ways in which one can proceed with an implementation: for example, a straightforward choice is one that applies constraint-satisfaction techniques, an approach that extends that discussed in (Marcu, 1996). Given a sequence U of N textual units, one can take advantage of the structure of the domain and

associate with each of the $\text{N}(\text{N}+1)/2$ possible text spans a status and a type variable whose domains consist in the set of objects over which the corresponding predicates S and T, range. For each of the $\text{N}(\text{N}+1)/2$ possible text spans $[l, h]$, one can also associate $h - l + 1$ promotion variables. These are boolean variables that specify whether units $l, l+1, \ldots, h$ belong to the promotion set of span $[l, h]$. For each of the $\text{N}(\text{N}+1)/2$ possible text spans $[l, h]$, one can also associate $h - l + 2$ intentional variables: one of these variables has as domain the set of rhetorical relations that are relevant for the span $[l, h]$. The rest of the $h - l + 1$ variables are boolean and specify whether unit $l, l+1, \ldots$, or h are arguments of the oracle function f_I that intentionally characterizes that span.

Hence, each text of N units yields a constraint-satisfaction problem with $\text{N}(\text{N}+1)(2\text{N}+13)/6$ variables $(\text{N}(\text{N}+1)(2\text{N}+13)/6 = 2\text{N}(\text{N}+1)/2 + \sum_{l=1}^{l \leq =\text{N}} \sum_{h=l}^{h \leq =\text{N}} (h-l+1) + \sum_{l=1}^{l \leq =\text{N}} \sum_{h=l}^{h \leq =\text{N}} (h-l+2)))$. The constraints associated with these variables are a one-to-one mapping of the axioms in section 3. Finding the set of RS-trees and the intentions that are associated with a given discourse reduces then to finding all the solutions for a traditional constraint-satisfaction problem.

5 Applications

Reasoning from text structures to intentions. Consider again the example text (1), which was used throughout this paper. As we discussed in section 1, if we assume that an analyst (or a program) determines that the rhetorical relations given in (2) hold between the elementary units of the text, there are five valid trees that correspond to text (1) (see figure 1). If we consider now the axioms that describe the relationship between text structures and intentions, we can infer, for example, that, for the tree 1.a, the *DSP* of span $[\text{A}_1, \text{D}_1]$ *dominates* the *DSP* of span $[\text{C}_1, \text{D}_1]$ and that the primary intention of the whole text depends on unit B_1 and on the rhetorical relation of EVIDENCE. In such a case, the axiomatization provides the means for drawing intentional inferences on the basis of the discourse structure. Also, although there are five discourse structures that are consistent with the rhetorical judgments in (1), they yield only three intentional interpretations, i.e., there are only three primary intentions that one can associate to the whole text. One intention is that discussed above, which is associated with analysis 1.a. Another intention depends on unit B_1 and the JUSTIFICATION relation that holds between units A_1 and B_1; this intention is associated with the analyses shown in figure 1.c and 1.e. And another intention depends on unit B_1 and the JUSTIFICATION relation that holds between units D_1 and B_1; this intention is associated with the analyses shown in figure 1.b and 1.d.

Reasoning from text structures to intentions can be also beneficial in a context such as that described by Lochbaum (1998) because the rhetorical constraints can help prune the space of shared plans that would characterize an intentional interpretation of a discourse.

Using intentions for managing rhetorical ambiguities.
Assume now that besides providing judgments concern-
ing the rhetorical relations that hold between various
units, an analyst (or a program) provides judgments of
intentions as well. If, for example, besides the relations
given in (2) a program determines that the *DSP* of span
$[A_1, D_1]$ *dominates* the *DSP* of unit D_1, the theory that
corresponds to these judgments and the axioms given
in section 3 yields only two valid text structures, those
presented in figure 1.b and 1.d. In this case, the axiom-
atization provides the means of using intentional judg-
ments for reducing the ambiguity that characterizes the
discourse parsing process.

**Investigating the relationship between semantic and
intentional relations.** In their seminal paper, Moore
and Pollack (1992) showed that a text may be charac-
terized by intentional and rhetorical analyses that are not
isomorphic. For example, for the text shown in (11) be-
low, which is taken from (Moore and Pollack, 1992), one
may argue from an informational perspective that A_3 is
a CONDITION for B_3. However, from an intentional per-
spective, one may argue that B_3 can be used to MOTI-
VATE A_3. Similar judgments can be made with respect
to units B_3 and C_3. Hence, the set of relations that com-
pletely characterizes text (11) is that shown in (12) be-
low.

(11) [Come home by 5:00.^{A_3}] [Then we can go to the hard-
ware store before it closes.^{B_3}] [That way we can finish
the bookshelves tonight.^{C_3}]

(12) $\begin{cases} rhet\_rel(\text{CONDITION}, A_3, B_3) \\ rhet\_rel(\text{MOTIVATION}, B_3, A_3) \\ rhet\_rel(\text{CONDITION}, B_3, C_3) \\ rhet\_rel(\text{MOTIVATION}, C_3, B_3) \end{cases}$

When given this discourse problem, our imple-
mentation produces the four discourse trees shown
in figure 4, each of them having a different primary
intention ($f_I(\text{CONDITION}, C_3), f_I(\text{MOTIVATION}, A_3),$
$f_I(\text{MOTIVATION}, B_3),$ and $f_I(\text{CONDITION}, B_3)).$
Hence, our approach enables one to *derive automatically*
and *enumerate all* possible rhetorical interpretations of
a text and to study the relationship between structure
and intentions. Our approach does not provide yet the
mechanisms for choosing between different interpreta-
tions, but it provides the foundations for such a study. In
contrast, Moore and Pollack's informal approach could
neither derive nor enumerate all possible interpretations:
in fact, their discussion refers only to the two trees
shown in figure 4.a and .b.

Unlike Moore and Pollack's approach, where it is sug-
gested that a discourse representation should reflect si-
multaneously both its informational and intentional inter-
pretations, the approach presented here is capable of only
enumerating these interpretations. The formal model we
proposed is not rich enough to accommodate concurrent,
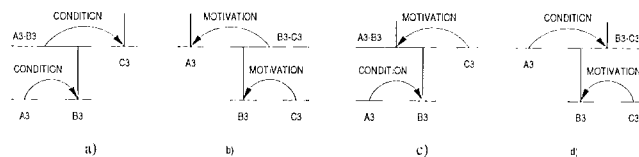non-isomorphic interpretations.

Figure 4: The set of all RS-trees that can be built for
text (11).

6 Conclusion

Crucial to the development of syntactic theories was the
ability to provide mechanisms capable of deriving all
valid syntactic interpretations of a given sentence. Se-
mantic or corpus-specific information was then used to
manage the usually large number of interpretations.

The work described in this paper sets theoretical foun-
dations that enable a similar approach to the study of dis-
course. The way a syntactic theory enables all valid syn-
tactic trees of a sentence be derived, the same way the
axiomatization presented here enables all valid discourse
trees of a text be derived. But the same way a syntac-
tic theory may produce trees that are incorrect from a
semantic perspective for example, the same way the ax-
iomatization described here may produce trees that are
incorrect when, for example, focus and cohesion are fac-
tored in.

A number of researchers have already shown how *in-
dividual* rhetorical and intentional judgments can be de-
rived automatically from linguistic constructs such as
tense and aspect, certain patterns of pronominalization
and anaphoric usages, *it*-clefts, and discourse markers or
cue phrases. But once these judgments are made, we still
need to determine all discourse interpretations that are
not only consistent with these judgments but also valid.
This paper provides mechanisms for deriving and enu-
merating all valid structure of a discourse and enables a
quantitative study of the relation between text structures
and intentions.

References

Nicholas Asher and Alex Lascarides. 1998. Questions in dia-
logue. *Linguistics and Philosophy*, 21(3):237–309.
Barbara J. Grosz and Candace L. Sidner. 1986. Attention, in-
tentions, and the structure of discourse. *Computational Lin-
guistics*, 12(3):175–204.
Karen E. Lochbaum. 1998. A collaborative planning
model of intentional structure. *Computational Linguistics*,
24(4):525–572.
William C. Mann and Sandra A. Thompson. 1988. Rhetorical
structure theory: Toward a functional theory of text organi-
zation. *Text*, 8(3):243–281.
Daniel Marcu. 1996. Building up rhetorical structure trees. In
Proceedings of AAAI-96, pages 1069–1074.
Johanna D. Moore and Martha E. Pollack. 1992. A problem
for RST: The need for multi-level discourse analysis. *Com-
putational Linguistics*, 18(4):537–544.
Megan Moser and Johanna D. Moore. 1996. Toward a synthe-
sis of two accounts of discourse structure. *Computational
Linguistics*, 22(3):409–419.