

## Right Frontier Constraint for discourses in non canonical order

Laurence Danlos, Pierre Hankach

► **To cite this version:**

Laurence Danlos, Pierre Hankach. Right Frontier Constraint for discourses in non canonical order. Proceedings of the Third International Workshop on Constraints in Discourse (CID'2008), 2008, Postdam, Germany. inria-00524740

**HAL Id: inria-00524740**

**<https://hal.inria.fr/inria-00524740>**

Submitted on 8 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Right frontier constraint for discourses in non-canonical order

Laurence DANLOS  
Université Paris 7, ALPAGE, IUF

laurence.danlos@linguist.jussieu.fr

Pierre HANKACH  
France-Télécom

pierre.hankach@orange-ftgroup.com

## 1 Introduction

One of the the main characteristics of a coherent (written) discourse is that its constituents are related to each other in a constrained way. The Right Frontier Constraint (henceforth RFC), which is postulated in Segmented Discourse Representation Theory (SDRT) [Asher and Lascarides2003] after works in other theories including [Grosz and Sidner1986] and [Polanyi1985], stipulates that while a (written) discourse unfolds a newly introduced constituent cannot be related to any of the other constituents in the preceding discourse. Rather, the constituents of the preceding discourse available for the attachment of a new constituent are those on the “right frontier”. Moreover, RFC is also used as an anaphora resolution constraint.

However, SDRT exclusively concentrates on discourses in “canonical order”, namely, using RST terminology [Mann and Thompson1988], discourses in which the Satellite of any nucleus-satellite relation *follows* the Nucleus. This paper aims at examining how RFC should be redefined when taking into account discourses in non-canonical order. We concentrate here only on discourses which include a Satellite *preceding* its Nucleus (to the exclusion of discourses which include a Satellite embedded in its Nucleus).

Section 2 presents how the right frontier is defined and used in SDRT for discourses in canonical order. Section 3 presents how the right frontier should be redefined and used in SDRT for discourses in non-canonical order.

## 2 Right Frontier Constraint in SDRT

### 2.1 Definition and attachment constraint

The right frontier is defined thanks to the hierarchical structure of discourse representations. The hierarchical structure built in SDRT (at the “information packaging” level) is obtained by differentiating discourse relations of two types: subordinating (nucleus-satellite relation in RST) and coordinating (multi-nuclear in RST). A subordinating relation represents a dominance relationship between the two arguments of the relation. It is geometrically represented by a vertical arrow. In RST terminology, the dominant argument corresponds to the Nucleus, the dominated argument to the Satellite. *Elaboration*, *Evidence*, *Explanation*, *Background* and *Purpose* are subordinating relations. A coordinating relations indicates that the two arguments are on equal footing and contribute to the same dominant topic. It is geometrically represented by an horizontal arrow. In RST terminology, both arguments are Nuclei. *Narration* and *Continuation* are coordinating relations<sup>1</sup>.

An abstract illustration of SDRT hierarchical discourse structures is depicted in the graph of Figure 1. This graph includes two kinds of nodes:

- $\pi_i$  atomic nodes that are labels of logical forms for atomic clauses,
- $\pi$  primed nodes ( $\pi'$ ,  $\pi''$ ) that are scope nodes which immediatly outscope atomic nodes, e.g  $i$  –  $outscope(\pi', \pi_2)$ . Outscoping relations are represented with dashed lines.

In [Asher and Lascarides2003], the right frontier is defined as being identical to the right frontier of the graph. Namely, it includes the last atomic node and any node that dominates it via a series of outscoping and/or subordinating relations. In Figure 1, the right frontier includes  $\pi_5$ , the last atomic node, and the nodes  $\pi''$ ,  $\pi_3$ ,  $\pi'$  and  $\pi_1$ .  $\pi_2$  and  $\pi_4$  are not on the right frontier.

<sup>1</sup>*Parallel* and *Contrast* are also coordinating relations, however, they are “structural relations” which are not concerned with RFC as an anaphora resolution constraint. This paper contains no exemple of discourse involving structural relations.

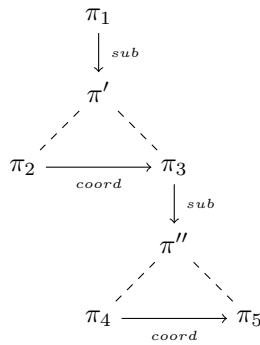


Figure 1: Illustration of SDRT hierarchical discourse structures

RFC stipulates that only nodes on the right frontier are available for the attachment of a new constituent. This correctly predicts that discourse (1) is infelicitous. The structure of the sub-discourse (1a-e) made up of the first five sentences is the one in Figure 1. The last sentence (f) is supposed to be attached to the second one (b) via *Explanation* (Mary did the shopping because the fridge was empty), but this is impossible because  $\pi_2$ , the label of the logical form for (b), is not on the right frontier, therefore it is not available for the attachment of a new constituent.

- (1)a. Mary had a busy day.
- b. First, she did the shopping.
- c. Next, she did some housework.
- d. She cleaned the kitchen
- e. and she vacuumed the dining room.
- f. #The fridge was empty.

## 2.2 Anaphora resolution constraint

The right frontier is also used in SDRT for anaphora resolution. In [Asher and Lascarides2003], it is claimed that an antecedent for **any** anaphoric expression must be DRS-accessible on the right frontier. This claim seems valid for pronominal anaphors<sup>2</sup>. For example, it correctly predicts that discourse (2) is infelicitous. The structure of the sub-discourse made up of the first five sentences (a-e) is the one in Figure 1. The antecedent of the pronoun *he* in (f) can only be John, an entity which is introduced in (b), but this anaphoric link is impossible because  $\pi_2$  is not on the right frontier. On the other hand, (2) becomes acceptable if *he* is replaced by *John*. Therefore, the unacceptability of (2) is only due to the use of the pronoun *he*, whose antecedent is not on the right frontier.

- (2)a. Mary had a busy day.
- b. First, she had lunch with John.
- c. Next, she did some housework.
- d. She cleaned the kitchen
- e. and she vacuumed the dining room.
- f. #He was in a bad mood.
- g. So the lunch was not fun.

However, this constraint does not hold for definite anaphors. Discourse (3) is felicitous<sup>3</sup>, although the antecedent of the definite anaphor *the duck magret* in (f) is in  $\pi_2$  which is not on the right frontier.

- (3)a. John had a great meal.
- b. He ate a duck magret.
- c. Next, he devoured lots of cheese.
- d. He tasted some camembert
- e. and he also savored some roquefort.
- f. The duck magret was especially delicious.

<sup>2</sup>It is also valid for 0 anaphors, see [Asher2008].

<sup>3</sup>The structure of the sub-discourse made up of the first five sentences of (3) is the one in Figure 1.

In [Asher2008], a solution which heavily relies on “topic nodes” is proposed to handle definite anaphors in discourses such as (3). We will not examine this solution, however we want to discuss the following question: discourse (3) is obviously a counter-example to RFC as an anaphora resolution constraint when not taking into account constructed topic nodes, but is it a counter-example to RFC as an attachment constraint? The answer to this question depends on the attachment of  $\pi_6$  labeling the logical form for (f). As explained in [Danlos2008], there are two positions. The first one, defended in [Wolf and Gibson2005], is that  $\pi_6$  should be attached to  $\pi_2$  via *Elaboration* since  $\pi_6$  elaborates an entity introduced in  $\pi_2$ . Using the subtypes of *Elaboration* introduced in [Knott1996], this relation between  $\pi_6$  and  $\pi_2$  is *Elaboration–Object*. The second position, defended by [Egg and Redeker2008] and [Asher2008], is that  $\pi_6$  should be attached to  $\pi_1$  via *Elaboration (Elaboration – Event* in Knott’s terms). The former amounts to systematically representing anaphoric links in discourse structures. We think that the latter approach is the right one for two reasons. First, anaphoric links cannot systematically be represented in discourse structure. This is obvious for a plural pronoun with disjunct antecedents. In (4), the antecedents of *they* are the crab salad and the duck magret, however a discourse structure in which  $\pi_3$  is linked both to  $\pi_1$  and  $\pi_2$  is not tenable.

- (4)a. John ate a crab salad.
- b. He also devoured a duck magret.
- c. They were delicious.

Second, [Asher2008] rightly insists on the rhetorical role of the last sentence of (3). The use of *especially* in (f) explicitly means that the other elements of the meal were not as delicious. In other words, (f) does not comment only on the duck magret but on the meal as a whole. As a consequence,  $\pi_6$  should be attached to  $\pi_1$  via *Elaboration – Event* and not to  $\pi_2$  via *Elaboration – Object*. Moreover, there is no need to have a contrastive marker such as *especially* to get a contrastive effect between an element of a set and the other elements of the set. As an illustration, consider discourse (5a-d) taken from [Wolf and Gibson2005]. The fact that only the price of the basil is described in (d) as being high, implicitly means that the price of the tomatoes was not as high. In other words, (5a-d) and (5a-e) roughly convey the same information.

- (5)a. Susan wanted to buy some tomatoes
- b. and she also tried to find some basil
- c. because her recipe asked for these ingredients.
- d. The basil would be quite expensive at this time of the year.
- e. However, the tomatoes would be less expensive.

In discourse (5a-e), (d) and (e) form a complex constituent, whose topic is the price of the ingredients, which is attached to the complex constituent (a-b), whose topic is the buying of the ingredients, via *Elaboration(–Event)*. This means that the anaphoric links between the definite anaphors in (d) and (e) and their respective antecedents in (a) and (b) are not represented in the discourse structure<sup>4</sup>. Now, since (5a-d) and (5a-e) roughly convey the same information, their discourses must have compatible discourse structures, the one for (5a-d) being obtained from the one for (5a-e) by suppressing the node  $\pi_5$  for (e) and any element involving this node. Therefore, in discourse (5a-d), (d) should be attached to the complex constituent formed by (a) and (b) via *Elaboration(–Event)*<sup>5</sup>.

We can add that discourse (6), which differs from (3) in the last sentence (f), is not a felicitous written text. This is due to the fact that the last sentence, which comments on the color of the duck magret, cannot easily induce a contrastive effect with the (color of the) cheese. Therefore, the reader may wonder why the writer adds this information after finishing the description of the meal. She is forced to infer that the duck magret was in some way or other more memorable than the cheeses.

- (6)a. John had a great meal.
- b. He ate a duck magret.
- c. Next, he devoured lots of cheese.
- d. He tasted some camembert
- e. and he also savored some roquefort.
- f. ??The duck magret was a beautiful red.

<sup>4</sup>If these anaphoric links were represented, then discourse structures for examples such as (5a-e) would exhibit crossing dependencies, as explained in [Egg and Redeker2008] and [Danlos2008].

<sup>5</sup>By the same reasoning, we also claim that in (5abd), where there is no interposition of a sentence between (d) including the definite *the basil* and (b) including its antecedent, (d) should be attached to the complex constituent formed by (a) and (b) via *Elaboration(–Event)*.

Summarizing, the attachment of the last sentence of discourses (3), (5) and (6), which include a definite anaphor whose antecedent is not on the right frontier, doesn't violate RFC as an attachment constraint. On the other hand, these definite anaphors do violate RFC as an anaphora resolution constraint (when not taking into account constructed topic nodes): this constraint seems to hold only for pronominal anaphors (and 0 anaphors).

### 3 Right frontier constraint for discourses in non-canonical order

#### 3.1 Discourses in non-canonical order

In SDRT, the following simplification is made: it is assumed that the dominant argument (Nucleus) of a subordinating relation always precedes the dominated one (Satellite)<sup>6</sup>. As a consequence, a discourse such as (7) with a preposed subordinate clause introducing the Satellite before the Nucleus is not dealt with in SDRT.

- (7) a. As John is sick,  
 b. he didn't go to work today.
- $$\begin{array}{c} \pi_2 \\ \downarrow \textit{Explanation}_l \\ \pi_1 \end{array}$$

Following [Danlos2008], discourses in which the Satellite of a subordinating relation precedes the Nucleus are said to be "in non-canonical order". Given a subordinating relation  $R$ , if its Satellite is on the right of its Nucleus (canonical order), it is noted  $R_r$  and is typed as a "right subordinating" relation, otherwise (non-canonical order), it is noted  $R_l$  and typed as "left subordinating". For example, in (7),  $\textit{Explanation}_l(\pi_2, \pi_1)$  holds and the graph of the discourse structure contains a vertical arrow going from the Nucleus  $\pi_2$  down to the Satellite  $\pi_1$ .

Now we are ready to tackle the following questions: how should the right frontier be defined for discourses in non-canonical order? Can we postulate a right frontier constraint for attachment of a new constituent and/or for pronoun resolution which holds for discourses in canonical and non-canonical order? Since definite anaphora are not concerned with RFC for discourses in canonical order (Section 2.2), they are not studied for discourses in non-canonical order.

Before we continue, let us make a methodological remark. The issue of attachment of new information in discourse update should not be confused with the issue of anaphora resolution. This doesn't mean that discourse update and anaphora resolution must be considered as independent tasks which are performed sequentially. However, even if these two tasks are dependent and performed simultaneously, the issues they raise must be discussed separately. Therefore, Section 3.2 is dedicated only to the former issue. It presents only examples with no anaphora resolution problem: either these examples don't include anaphora or they do so but only in the case where the antecedent of a pronominal anaphor appears in the clause immediately preceding the clause in which the pronoun appears (a case which is known to be unproblematic). Section 3.3 tackles the issue of pronominal anaphora resolution.

We draw a parallel between the discourse notions of Nucleus and Satellite and the syntactic notions of head and modifier. A Nucleus plays the role of a head, a Satellite the role of a modifier. In syntax, the following rules hold.

**Syntactic rule 1** *Two modifiers can share the same head.*

**Syntactic rule 2** *Two heads cannot share the same modifier.*

**Syntactic rule 3** *A given element can play two roles, head and modifier.*

We postulate that these syntactic rules can be extrapolated to discourse through the notions of Nucleus and Satellite.

#### 3.2 Attachment of new information

First, we examine discourses which start with a left subordinating relation  $R_l(\pi_2, \pi_1)$ , in other words discourses starting with a left Satellite  $\pi_1$  (to simplify, the first clause<sup>7</sup>) followed by a Nucleus  $\pi_2$  (to simplify,

<sup>6</sup>This simplification is not made in RST: the Satellite of a subordinating relation follows or precedes the Nucleus.

<sup>7</sup>The left Satellite can be a complex constituent built around a coordinating relation. For example, in (8), we have  $\textit{Background}_l(\pi_3, \textit{Continuation}(\pi_1, \pi_2))$ .

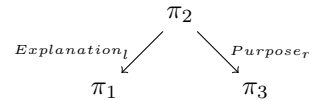
the second clause<sup>8</sup>). Next, we examine left subordinating relations that appear in the middle of a discourse (Section 3.2.2).

### 3.2.1 Left subordinating relations at the beginning of a discourse

For discourses starting with a left subordinating relation  $R_l(\pi_2, \pi_1)$ , the Nucleus  $\pi_2$  is clearly available for attachment of new information, i.e. attachment of a third clause represented in  $\pi_3$ . As shown below,  $\pi_2$  can be linked to  $\pi_3$  via a right and left subordinating relation or via a coordinating relation.

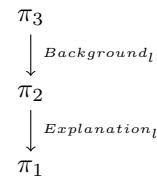
- Discourse (10) exemplifies the case where  $\pi_2$  is linked to  $\pi_3$  via a right subordinating relation,  $Purpose_r$ . (10) illustrates Syntactic rule 1 for discourse: two satellites,  $\pi_1$  and  $\pi_3$ , share the same nucleus  $\pi_2$ .

- (10)a. As it was hot,  
 b. John went walking in the garden.  
 c. to get some fresh air.



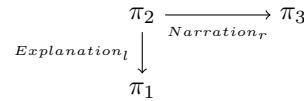
- Discourse (11) exemplifies the case where  $\pi_2$  is linked to  $\pi_3$  via a left subordinating relation,  $Background_l$ ; (11) illustrates Syntactic rule 3 for discourse: a given element,  $\pi_2$ , is both the Nucleus of  $\pi_1$  and the Satellite of  $\pi_3$ .

- (11)a. As the wind had chased the clouds away,  
 b. it was pretty sunny.  
 c. John went walking in the garden.



- Discourse (12) exemplifies the case where  $\pi_2$  is linked to  $\pi_3$  via a coordinating relation,  $Narration$ .

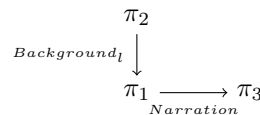
- (12)a. As the fridge was empty  
 b. John did the shopping.  
 c. Next, he went to the movies.



So the question at stake for discourses starting with a left Satellite  $\pi_1$  is the following: is it possible to attach new information  $\pi_3$  to this left Satellite? To answer this question, we must examine three types of relations (coordinating, left and right subordinating) possibly linking  $\pi_1$  and  $\pi_3$ .

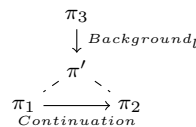
It seems impossible to link  $\pi_1$  and  $\pi_3$  via a coordinating relation: (13) with  $Narration(\pi_1, \pi_3)$  is infelicitous, and we speculate that any discourse with this structure is infelicitous.

- (13)a. While John was shopping,  
 b. he was in a good mood.  
 c. #Next, he went to the movies.



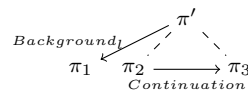
We want to stress the following point. In (13), we can replace the anaphoric link between *John* in (a) and *he* in (b) by a cataphoric link between *he* in (a) and *John* in (b), see *While he was shopping, John was in a good mood. Next, he went to the movies.* This anaphora/cataphora switch doesn't drastically change the (un)acceptability of (13). Similarly, for the following discourses, we claim that such an anaphora/cataphora switch doesn't drastically change their acceptability. Cataphora are discussed in Section 3.4.

- (8) a. It was sunny  
 b. and spring was coming.  
 c. Mary took a long walk in the garden.



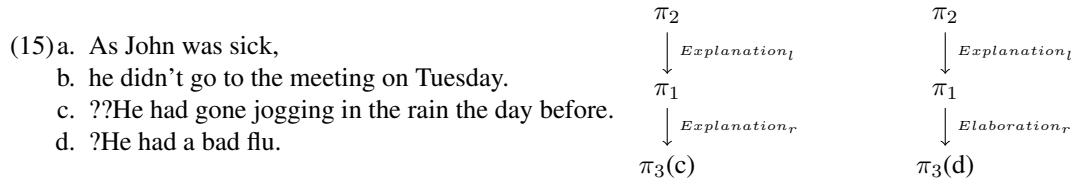
<sup>8</sup>The Nucleus can also be a complex constituent built around a coordinating relation. For example, in (9), we have  $Background_l(Continuation(\pi_2, \pi_3), \pi_1)$ .

- (9) a. It was pretty hot.  
 b. John went into the garden.  
 c. and Mary went into the cave.

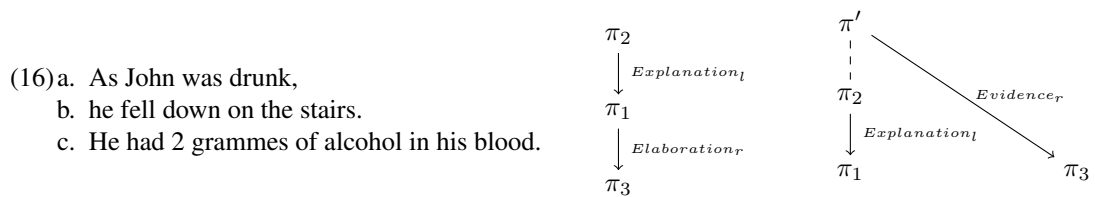


It is also impossible to link  $\pi_1$  and  $\pi_3$  via a left subordinating relation, which would mean that  $\pi_1$  is the left Satellite both for the Nuclei  $\pi_2$  and  $\pi_3$ . This case can be excluded by a general rule stating that two Nuclei cannot share the same Satellite, which extrapolates Syntactic rule 2 for discourse.

We are left with linking  $\pi_1$  and  $\pi_3$  via a right subordinating relation<sup>9</sup>. This case is illustrated with (15abc) which is intended to display  $Explanation_r(\pi_1, \pi_3)$  and which sounds a little unnatural, or with (15abd) which displays  $Elaboration_r(\pi_1, \pi_3)$  and which sounds better without being perfect.



The unpleasant acceptability of discourses such as (15abc) or (15abd) can be explained intuitively. Namely, since the most salient argument of a left subordinating relation – the Nucleus – follows the argument that depends on it – the Satellite –, the Nucleus is likely to eclipse the Satellite, making impossible to attach new information to it. Nevertheless, this intuition fails when considering (16), which *a priori* displays  $Elaboration_r(\pi_1, \pi_3)$  and which sounds perfect.

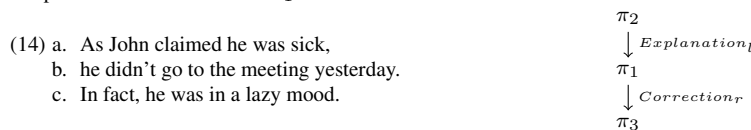


In (16), there is a coreference relation between John's states described in  $\pi_1$  and  $\pi_3$ <sup>10</sup>. If (16) were in canonical order, namely (16bac) switching *John* and *he* in (a) and (b), the analysis with  $Elaboration_r$  reflecting the coreference relation would be undebatable and unproblematic when (16) is uttered in a non-null left context. However, we will show in the next section that analyzing (16) with  $Elaboration_r(\pi_1, \pi_3)$  leads to a crossing dependency when (16) is uttered in a non-null left context. Therefore, with the aim of avoiding crossing dependencies in discourse structures, one can argue that the coreference relation between John's states described in  $\pi_1$  and  $\pi_3$  should have no counter-part in the discourse structure, which would then not include  $Elaboration_r(\pi_1, \pi_3)$ . This position is in the spirit of what was under discussion in Section 2.2, namely some anaphoric (coreference) links must not be represented in discourse structures. If  $\pi_3$  is not attached to  $\pi_1$  via  $Elaboration_r$ , then how and where should  $\pi_3$  be attached?

A first solution consists in considering that  $\pi_3$  provides background information to John's falling or to the causal relation between John's being drunk and his falling. In other words,  $Background_r(\pi_2, \pi_3)$  or  $Background_r(\pi', \pi_3)$  with a complex constituent  $\pi'$  defined by  $i - \textit{outscope}(\pi', \textit{Explanation}_l(\pi_2, \pi_1))$  should hold instead of  $Elaboration_r(\pi_1, \pi_3)$ . However, one must note the incoherence of discourse (17) which differs from (16) in the first sentence<sup>11</sup>. In (17), the first and third sentences describe both a state of John, but without a coreference relation (a background information that doesn't describe a state of John may be possible, see *As John was drunk, he fell down on the stairs. It was pretty hot.*, however, this is not at the heart of the matter).

- (17)a. As John had a high temperature,  
       b. he fell down on the stairs.  
       c. # He had 2 grammes of alcohol in his blood.

<sup>9</sup>We put aside any example displaying  $Correction_r(\pi_1, \pi_3)$  since  $Correction_r$  is, according to [Asher and Lascarides2003], a special subordinating relation whose first argument is not compulsorily on the right frontier. Discourse (14) is perfect, but we don't use it to prove that the left satellite  $\pi_1$  is an available site of attachment.



<sup>10</sup>This coreference relation implies that the discourse relation  $Elaboration$  between  $\pi_1$  and  $\pi_3$  is of a sub-type called *Particularization* in [Danlos2001].

<sup>11</sup>(17) becomes coherent if *also* is added to (c), see *He also had 2 grammes of alcohol in his blood*. Then  $\pi_3$  is another explanation of  $\pi_1$ .

The contrast between (16) and (17) shows that supposedly background information describing a state of John is possible only when there is a coreference relation between John’s states described in  $\pi_1$  and  $\pi_3$ . This constraint on attachment of background information to  $\pi_2$  or to  $Explanation_l(\pi_2, \pi_1)$  looks *ad hoc* since the discourse relation *Background* imposes only an (overlap) temporal constraint on its arguments. Therefore, we think that the solution which consists in attaching  $\pi_3$  via a relation  $Background_r$  is not tenable.

A second solution (advocated by B. Sagot, p.c.) consists in considering that  $\pi_3$  provides some evidence for the causal relation between John’s being drunk and his falling. This analysis amounts in saying that the quantity of alcohol in John’s blood makes that the probability of John falling when going downstairs or upstairs was high: his falling was highly predictable. With this solution, the obligatory coreference between John’s states described in  $\pi_1$  and  $\pi_3$  is explained in the following informal terms (inspired from the semantic constraint imposed by *Evidence* on its arguments that is given in [Asher and Lascarides2003]): if  $\gamma$  provides some evidence for a causal relation  $\alpha \Rightarrow \beta$ , which is an instance of a causal law  $\mathcal{A} \Rightarrow \mathcal{B}$ , then  $\gamma$  must be an instance of  $\mathcal{A}$  such that the probability of  $\gamma \Rightarrow \beta$  is strictly greater than the probability of  $\alpha \Rightarrow \beta$ <sup>12</sup>. Without starting a discussion on the precise definition of  $\mathcal{A}$ , we can assume that stating  $\alpha$  and  $\gamma$  are both instances of  $\mathcal{A}$  is equivalent to stating they are in a coreference relation.

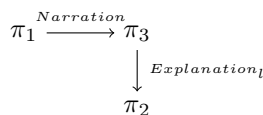
In conclusion, the analysis of (16) as  $Explanation_l(\pi_2, \pi_1) \wedge Evidence_r(\pi', \pi_3)$  – in which  $\pi'$  is defined by  $i - outscope(\pi', Explanation_l(\pi_2, \pi_1))$  – which indirectly explains the obligatory coreference relation between John’s states described in  $\pi_1$  and  $\pi_3$ , is a tenable solution. It competes with the *a priori* analysis  $Explanation_l(\pi_2, \pi_1) \wedge Elaboration_r(\pi_1, \pi_3)$ , which directly reflects the coreference relation between John’s states described in  $\pi_1$  and  $\pi_3$ .

We don’t want to come out in favor of one of these two solutions right now. However, we are going to show that adopting one or the other solution for analyzing (16) has a drastic consequence when this discourse is uttered in a non-null left context: the second solution leads to a crossing dependency.

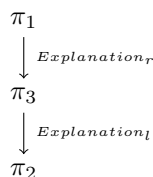
### 3.2.2 Left subordinating relations in the middle of a discourse

We concentrate on left subordinating relations that appear just after the first clause  $\pi_1$ , so left subordinating relations noted  $R_l(\pi_3, \pi_2)$ . For such cases,  $\pi_1$  can be linked to the left Satellite  $\pi_2$  neither via a coordinating relation nor via a right subordinating relation<sup>13</sup>. Therefore, since  $\pi_1$  must be linked to at least another discourse constituent, we assume that it is linked to  $\pi_3$ , via a coordinating or right subordinating relation, see (19a-c) with  $Narration(\pi_1, \pi_3)$  or (20a-c) with  $Explanation_r(\pi_1, \pi_3)$ . For such discourses, the question at stake is the following<sup>14</sup>: is it possible to attach new information labeled  $\pi_4$  to the left Satellite  $\pi_2$ <sup>15</sup>? If the answer is positive, it means that discourse structures can exhibit crossing dependencies, the relation  $R_a(\pi_1, \pi_3)$  crossing  $R_b(\pi_2, \pi_4)$  when  $R_a$  is subordinating.

- (19)a. John entered the living room.  
 b. As he was tired,  
 c. he sat down on the sofa.  
 d. ??He had been running around all day



- (20)a. John broke his leg yesterday.  
 b. As he was drunk,  
 c. he fell down on the stairs.  
 d. He had 2 grammes of alcohol in his blood.



<sup>12</sup>Along the same lines, if  $\gamma$  is some *counter-evidence* for a causal relation  $\alpha \Rightarrow \beta$ , which is an instance of a causal law  $\mathcal{A} \Rightarrow \mathcal{B}$ , then  $\gamma$  must be an instance of  $\mathcal{A}$  such that the probability of  $\gamma \Rightarrow \beta$  is strictly *lower* than the probability of  $\alpha \Rightarrow \beta$ . The difference between evidence and counter-evidence for a causal relation stands in that counter-evidence should be lexically marked, see (18) where (c) is introduced by the contrastive connective *yet*.

- (18) a. As John was drunk,  
 b. he fell down on the stairs.  
 c. Yet, he had (only) 0.5 grammes of alcohol in his blood.

<sup>13</sup>The latter case is excluded by an extrapolation of Syntactic rule 2 to discourse.

<sup>14</sup> Such discourses, especially the ones whose left satellite  $\pi_2$  is not introduced by a connective, see (21) in note 17, also raise a question for the information package level: how can the relations  $Ra_l(\pi_2, \pi_3)$  and  $Rb(\pi_1, \pi_3)$  be computed in a (left-to-right) incremental way? However, this question is left aside in this paper.

<sup>15</sup>Only attachment to  $\pi_2$  via a right subordinating relation is considered, coordinating and left subordinating relations being excluded straight away.



Discourse (19a-d) which is intended to display  $Explanation_r(\pi_2, \pi_4)$  sounds unnatural as (15abc) does. On the other hand, (20a-d) sounds natural as (16) does. Our discussion on the discourse structure of (16) has shown that two solutions can be contemplated. By the same reasoning, we can conclude that two discourse structures can be contemplated for (20). They are depicted in Figure 2 as SDRT graphs and as a dependency graph (DAG) [Danlos2004, Danlos2008] for the discourse structure which exhibits a crossing dependency,  $Explanation_r(\pi_1, \pi_3)$  crossing  $Elaboration_r(\pi_2, \pi_4)$ <sup>16</sup>.

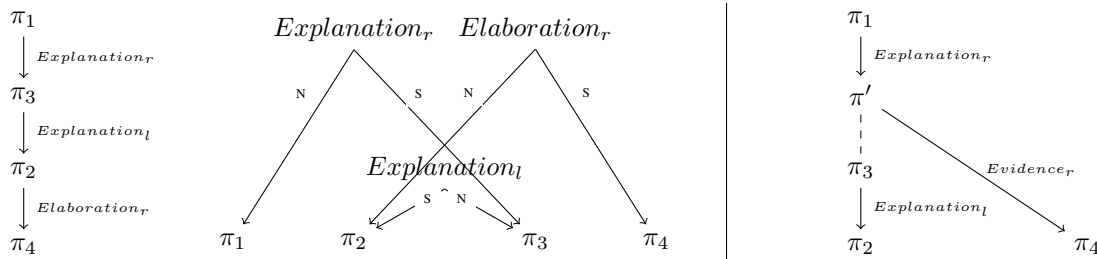


Figure 2: SDRT graphs and dependency DAG for (20)

### 3.2.3 Conclusion

Both the acceptabilities and the analyses of the discourses presented in the preceding sections are debatable<sup>17</sup>. Further investigation on real corpus examples is clearly needed. However, it seems that we can draw the following conclusion. A left Satellite which occurs at the beginning or in the middle of a discourse can be an available attachment node but only in a restricted way, i.e. only as the first argument of a right subordinating relation (excluding any left subordinating and coordinating relation), and not for any right subordinating relation. When it occurs in the middle of a discourse, the attachment of new information to it could lead to a crossing dependency, although it is often claimed that discourse structures don't exhibit crossing dependencies [Mann and Thompson1988, Asher and Lascarides2003, Egg and Redeker2008, Webber et al.2003].

With this conclusion, we are able to define the right frontier for discourses in canonical and non-canonical order. The right frontier is identical to the right frontier of SDRT graphs. Namely it includes the last atomic node, any node that dominates it **and** any node that it dominates<sup>18</sup>. RFC as an attachment constraint is stated as follows: the available nodes for attachment of new information are those on the right frontier, although attachment to a left Satellite – a node that is dominated by the last atomic node – is constrained in a way which needs further investigation.

### 3.3 Pronominal anaphora resolution

First, we can note that (22abc) and (22abd) are unacceptable. (22) differs from (15) in the subject of the second sentence (b) which is *Mary* instead of *he* referring to *John*. The fact that the acceptability of (22abc) or (22abd) is worse than the one of (15abc) or (15abd) can be explained by the fact that the pronoun *he* in the third sentence (c) or (d) has its antecedent in the left Satellite  $\pi_1$ , which is intuitively eclipsed by the Nucleus  $\pi_2$ .

- (22) a. As John was sick,  
 b. Mary didn't go to the meeting on Tuesday.  
 c. # He had gone jogging in the rain the day before.  
 d. # He had a bad flu.

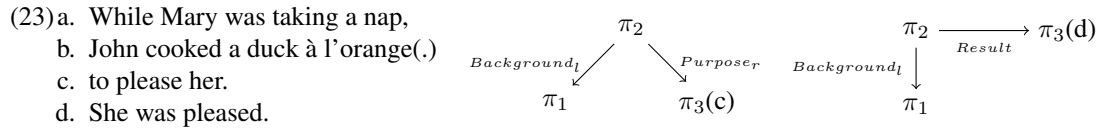
<sup>16</sup>In the discourse structure with a crossing dependency, we have not considered that  $\pi_3$  and  $\pi_2$  linked via  $Explanation_l$  form a complex constituent  $\pi'$ . This is not a crucial point because there is a crossing dependency with or without  $\pi'$ , as the reader can check.

<sup>17</sup>The reader can also examine the discourse structure of (21), in which (b) is introduced by no connective but is linked to (c) by *and*.

- (21) a. John broke his leg yesterday.  
 b. He was drunk  
 c. and he fell down on the stairs.  
 d. He had 2 grammes of alcohol in his blood.

<sup>18</sup>In most cases, the last atomic node dominates one node at most, although it can dominate two nodes, see (11).

However, (23abc) and (23abd) are totally acceptable. In these discourses,  $\pi_3$  is attached to  $\pi_2$  via the right subordinating relation *Purpose<sub>r</sub>* in (23abc) and via the coordinating relation *Result* in (23abd). The third sentence (c) or (d) contains a pronoun whose antecedent is *Mary* introduced in the first sentence (a). Therefore, these discourses show that the left Satellite  $\pi_1$  must be considered as a member of the right frontier. It has to be stressed that the acceptability of (23abc) is an argument against Veins Theory [Cristea et al.1998], see the quote “*The vein definition formalizes the intuition that in a sequence of units A B C, where A and C are satellites of B, B can refer to entities in A (its left satellite), but the subsequent right satellite, C, cannot refer to A due to the interposition of nuclear unit B.*”



In conclusion, we can keep the definition of the right frontier for discourses in canonical and non-canonical order which was given in Section 3.2.3. RFC as a pronominal anaphora resolution constraint is valid both for discourses in canonical and non-canonical order.

### 3.4 Cataphora resolution

It is well known that a proposed subordinate clause can include a cataphoric pronoun, see (24) in which *he* in (a) refers to *John* in (b).

- (24)a. As he was sick,  
 b. John didn’t go to the meeting today.

Cataphora are not dealt with in SDRT since only discourses in canonical order are dealt with. They obviously violate RFC as a pronominal anaphora constraint. Therefore, this constraint must be redefined with the addition of the case where a pronoun appears in the first clause of a discourse (generally within a proposed subordinate clause): this pronoun finds its referent in the subsequent clause.

## 4 Conclusion and future research

When taking into account discourses in non-canonical order with Satellites preceding Nuclei, the right frontier can be redefined so that it contains the last atomic node, any node that dominates it **and** any node it dominates. With this new definition, RFC holds as an attachment and pronominal anaphora resolution constraint both for discourses in canonical and non-canonical order. However, attachment of new information to a left Satellite on the right frontier is constrained: a new constituent can be attached only via a right subordinating relation (to the exclusion of left subordinating and coordinating relations), and this kind of attachment is not always allowed.

This study has been based on artificial constructed examples. Obviously, it needs to be validated by real corpora examples. For that, we intend to find left Satellites in the Penn Discourse Treebank [The PDTB-Group2008] for English and in the annotated corpora Annodis for French.

Further research will concern other types of discourse in non-canonical order, for example discourses including Satellites embedded in Nuclei.

### Acknowledgments

We want to thank Nicholas Asher, Benoît Sagot and Alexandros Tantos for their very fruitful comments, and also André Bittar for editing our English.

### References

[Asher and Lascarides2003] Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

[Asher2008] Nicholas Asher. 2008. Troubles on the right frontier. In A. Benz and P. Kühnlein, editors, *Constraints in Discourse*. Benjamins.

- [Cristea et al.1998] Dan Cristea, Nancy Ide, and Laurent Romary. 1998. Veins theory: A model of global discourse cohesion and coherence. In Christian Boitet and Pete Whitelock, editors, *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*, pages 281–285, San Francisco, California. Morgan Kaufmann Publishers.
- [Danlos2001] Laurence Danlos. 2001. Event coreference between two sentences. In H. Bunt, editor, *Computing Meaning (volume 2)*, pages 271–288. Kluwer Academic Publishers, Amsterdam.
- [Danlos2004] Laurence Danlos. 2004. Discourse dependency structures as constrained DAGs. In *Proceedings of SIGDIAL'04*, pages 127–135, Boston.
- [Danlos2008] Laurence Danlos. 2008. Strong generative capacity of RST, SDRT and discourse dependency DAGs. In A. Benz and P. Kühnlein, editors, *Constraints in Discourse*. Benjamins.
- [Egg and Redeker2008] Markus Egg and Gisela Redeker. 2008. Underspecified discourse representation. In A. Benz and P. Kühnlein, editors, *Constraints in Discourse*. Benjamins.
- [Grosz and Sidner1986] Barbara Grosz and Candace Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- [Knott1996] Alistair Knott. 1996. *A Data-driven Methodology for Motivating a Set of Coherence Relations*. Artificial intelligence thesis, University of Edinburgh.
- [Mann and Thompson1988] William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- [Polanyi1985] Livia Polanyi. 1985. A theory of discourse structure and discourse coherence. In P.D. Kroeber, W.H. Eilfort, and K.L. Peterson, editors, *Papers from the General Session at the 21st Regional Meeting of the Chicago Linguistics Society*.
- [The PDTB-Group2008] The PDTB-Group. 2008. The Penn Discourse Treebank 2.0. Annotation Manual. Technical Report IRCS-08-01, Institute for Research in Cognitive Science, University of Pennsylvania.
- [Webber et al.2003] B. Webber, A. Joshi, M. Stone, and A. Knott. 2003. Anaphora and discourse structure. *Computational Linguistics*, 29(4):545–587.
- [Wolf and Gibson2005] Florian Wolf and Edward Gibson. 2005. Representing discourse coherence: a corpus-based study. *Computational Linguistics*, 31(3):249–287.