Représentation sémantique sous-spécifiéepour les conjonctions de subordination

Laurence Danlos

MTT 2003, Paris, 16–18 juin 2003

Discourse dependency structures as DAGs

Laurence Danlos
TALANA/ LATTICE
Université Paris 7
Laurence.Danlos@linguist.jussieu.fr

Mots-clefs – Keywords

Discourse
MTT (Meaning-Text Theory ou Théorie Sens-Texte)
RST (Rhetorical Structure Theory)
SDRT (Segmented Discourse Representation Theory)

Abstract – Résumé

We show that the semantic level for discourses, understood as a dependency representation, can be mathematically characterized as DAGs. Our argumentation is based on discourses with three clauses and two discourse connectives. This simple case allows us to state that DAGs for discourses are not arbitrary: they present structural constraints.

Nous montrons que les représentations sémantiques pour les discours, conçues comme des représentations de dépendances sémantiques, ont la structure mathématique d'un DAG. Notre argumentation repose sur des discours très simples, à savoir sur les discours comportant trois propositions et deux connecteurs de discours. Ces discours, bien que simples, permettent d'aboutir à une conclusion importante : les DAG représentant les discours ne sont pas arbitraires, ils présentent des contraintes structurelles.

1 Introduction

Within a multi-level approach for discourse processing, this paper focuses on the semantic level. This level reflects the discourse structure (*how things are said*, *how the discourse is rhetorically organized*). This structure plays an important role, e.g., it constraints both anaphora resolution and the attachment of incoming propositions in understanding. The informational content level (*what is said*) will not be discussed here, neither the syntactic level.

The semantic representation is understood as a dependency representation, as in MTT[1] although MTT deals only with (complex) sentences, discarding discourses. An extension of MTT from sentences to discourses requires (at least) to take into account adverbial connectives (*therefore*, *next*) on the top of other connectives such as subordination and coordination conjunctions. As advocated in most discourse theories, e.g. RST and SDRT but not DRT[2], we assume that a connective denotes a "discourse relation", which is viewed as a predicate with two arguments. The discourse structure in RST corresponds roughly to a semantic dependency representation. This is not the case in SDRT, although one can infer a dependency representation from a SDRS boxed representation (as we will do it in this paper).

One concern in this paper is to determine to which mathematical object dependency structures for discourses correspond. In RST, it is a principle that this object is a tree. In SDRT, the issue is not discussed. We will show that this object is a directed acyclic graph (DAG), more precisely, a labeled ordered DAG (abbreviated as LODAG). This result fits with MTT to the extent that the semantic representations for sentences in MTT are DAGs with pointers to "dominant" nodes[3].

Our argumentation is based on one of the simplest cases of discourses, namely discourses of type *S1* $Conn_a$ *S2* $Conn_b$ *S3* with two discourse connectives ($Conn_{a/b}$) and three clauses (*Si*). We will show (Section 2) that they are topologically only four types of LODAGs for these discourses. This allows us to state that LODAGs for discourses are not arbitrary: they present structural constraints (Section 3). This is an important result since many authors in the discourse community hang on trees as discourse structures, even if it means to use artificial trees as shown *infra*. They reject DAGs because they view them as completely unconstrained (except the acyclicity constraint) and so as unusable in discourse processing.

Only discourses in which sentences are linked by discourse connectives are considered here. I stipulate that the results obtained can be extrapolated to discourses in which sentences are simply juxtaposed without discourse connective. Note that this paper concerns only the topological structure of semantic dependency representations for discourses: not a word is said on how to compute them (neither in understanding nor in generation).

Before getting to the heart of the matter, let us give some preliminaries.

Semantic dependency representation for *S1* $Conn_a$ *S2*. I adopt a conventional semantic dependency representation for discourses of the type *S1* $Conn_a$ *S2*. Namely, $Conn_a$ denotes a discourse relation R_a. R_a is a predicate with two arguments $\pi 1$ and $\pi 2$, which correspond to the semantic representations of *S1* and *S2* respectively. These arguments are both ordered ($\pi 1$

[1] MTT stands for Meaning-Text Theory (Mel'cuk, 1997).

[2] RST stands for Rhetorical Structure Theory (Mann & Thompson, 1987). SDRT stands for Segmented Discourse Representation Theory (Asher, 1993) (Asher & Lascarides, 2003). DRT stands for Discourse Representation Theory (Kamp & Reyle, 1993). DRT does not involve discourse relations.

[3] A DAG which includes pointers to "dominant" nodes is roughly equivalent to a LODAG (Candito & Kahane, 1998).

precedes $\pi 2$)[4] and labeled, with the labels nucleus (noted as N) and satellite (noted as S) which are taken from RST. $\pi 1$ is said to be the nucleus of R_a, $\pi 2$ its satellite. In the tradition of dependency grammar, this semantic representation is represented as the LODAG in the left-hand side of Figure 1. This LODAG is written as $\mathrm{R}_a(\pi 1, \pi 2)$ in an abbreviated form. In SDRT, it corresponds to the boxed representation shown in the right-hand side of Figure 1 (e_i denotes the main eventuality of *Si*)[5].

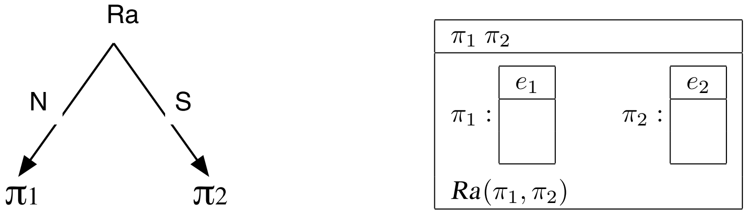

Figure 1: LODAG and SDRS for *S1 Conn*$_a$ *S2*

Compositionality principle. I assume the following principle: if a discourse D_n with n sentences is such that a sub-discourse D_p with p sentences and $1 \prec p \prec n$ can be inferred, then the DAG for D_n must include the DAG for D_p. This intuitive principle needs to be formally precised.

Interpretation of dependency relations in trees. Two different ways can be used to interpret dependency relations in trees: the standard one used in mathematics and computer science, and the "nuclearity principle" put forward in RST (Marcu, 1996). Let us illustrate them with the tree in Figure 2. With the standard interpretation, the nucleus of R_c is its left daughter (the *tree* rooted at R_a), while with the nuclearity principle, it is π_1 (the *leaf* which is the nucleus of R_a). Similarly, with the standard interpretation, the satellite of R_a is its right daughter (the *tree* rooted at R_b), while with the nuclearity principle, it is π_2 (the *leaf* which is the nucleus of R_b). To put it in a nutshell, the arguments of a discourse relation can be intermediary nodes or leaves with the standard interpretation, while they can only be leaves with the nuclearity principle.

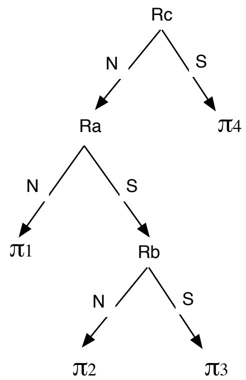

Figure 2: Binary tree

[4]Subordinate conjunctions are the only connectives which offer linear order variants: a subordinate clause can be preposed to the main clause. In (Danlos, 2003), I discuss preposed subordinate clauses in detail both at the syntactic and semantic level, and I propose an (underspecified) semantic dependency representation for subordinate clauses which does not depend on their position. So it can be assumed in this paper, which deals only with semantics, that subordinate clauses are always postposed. The order of the semantic arguments $\pi 1$ and $\pi 2$ corresponds thus to the order of the sentences (*S1* precedes *S2*) for any connective *Conn*$_a$.

[5]In fact, in dependency formalisms, the arguments of a discourse relation are semantic dependency representations (i.e. LODAGs), while in SDRT they are labels (π_i) for semantic representations. Figure 1 skips this difference: in the LODAG, π_i stands for a LODAG, in the SDRS, it stands for a label. We will come back to this difference at the end of Section 2.

I will show (Section 2) that the standard interpretation should be adopted. The point I want to make now is that one could argue that the nuclearity principle should be adopted instead, but one should not feel free to use both interpretations for the same tree. This is however what is done by some authors. For example, the discourse in (1) is represented as the tree on the left-hand side of Figure 3 by (Webber *et al.*, 2003) (this tree is a copy of theirs: it is the reason why the edges are neither labeled nor directed).

(1) a. Although John is very generous -
b. if you need some money,
c. you only have to ask him for it -
d. he's very hard to find.

Let us show that some predicate-argument relations are given by the nuclearity principle, while other ones are given by the standard interpretation in their tree. From (1), (2) can be inferred. This is evidence that the arguments of the discourse relation "concession" in their tree are a and d. These predicate-argument dependencies are given by *the nuclearity principle*.

(2) a. Although John is very generous,
d. he's very hard to find.

From (1), (3) can also be inferred. This is evidence that the arguments of "elaboration" in their tree are a and the tree rooted at "condition". These dependencies are given by *the standard interpretation*.

(3) a'. John is very generous -
b. if you need some money,
c. you only have to ask him for it.

Nevertheless, one should not feel free to use artificial trees relying on a mixed interpretation (the standard one and the nuclearity principle) just to avoid DAGs. One has to admit that discourse structures are DAGs, for example, the DAG on the right hand side of Figure 3 for (1). This DAG is conform to our compositionality principle: it can be viewed as the fusion of the dependency graphs for (2) and (3), with the discourse in (1) being viewed as the fusion of the discourses in (2) and (3).

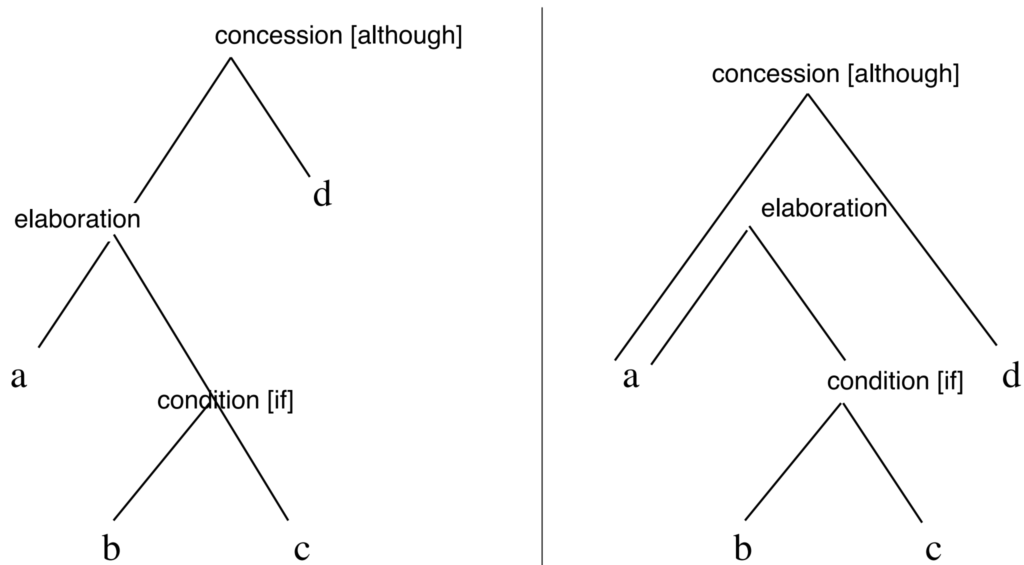

Figure 3: Artificial tree and DAG for (1)

2 LODAGs for *S1 Conn*$_a$ *S2 Conn*$_b$ *S3*

In discourses of the type *S1 Conn*$_a$ *S2 Conn*$_b$ *S3*, it is compulsorily the case that the nucleus of R$_a$ is π_1 and the satellite of R$_b$ is π_3. On the other hand, the satellite of R$_a$ and the nucleus of R$_b$ may vary depending on scope. More precisely, the satellite of R$_a$ may *a priori* be:

- either the representation of the whole right hand side of *Conn*$_a$, i.e. the semantic representation of *S2 Conn*$_b$ *S3*. I call this case "wide scope" of *Conn*$_a$ or R$_a$. It leads to graph (A) in Figure 4. The dependency relations in (A), which is tree shaped, must be interpreted in the standard way: the satellite of R$_a$ is its right daughter, i.e. the tree rooted at R$_b$.

- or the representation of one of the two sentences in the right hand side of *Conn*$_a$. This case leads either to graph (A1) or (A2) in Figure 4.

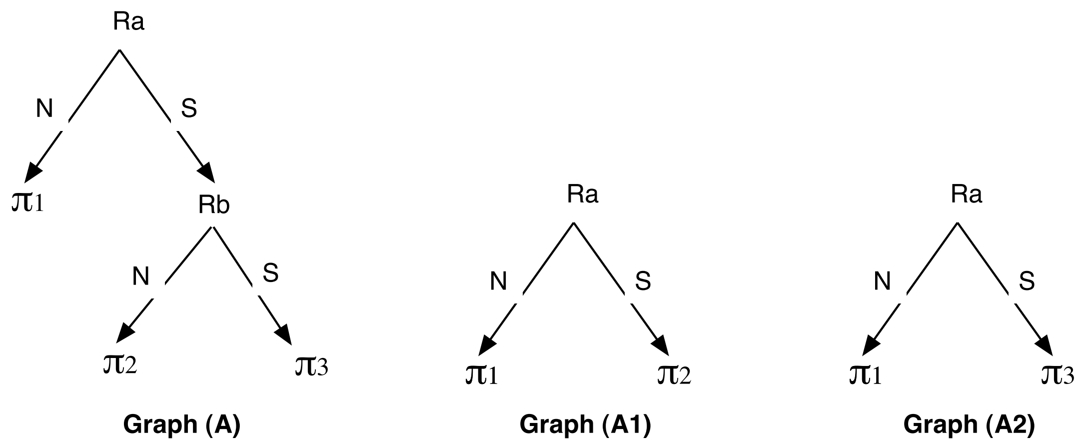

Figure 4: LODAGs for the various scopes of R$_a$

Similarly, depending on the scope of *Conn*$_b$, the nucleus of R$_b$ may *a priori* be R$_a(\pi_1, \pi_2)$, π_2 or π_1 as shown in the LODAGs in Figure 5.

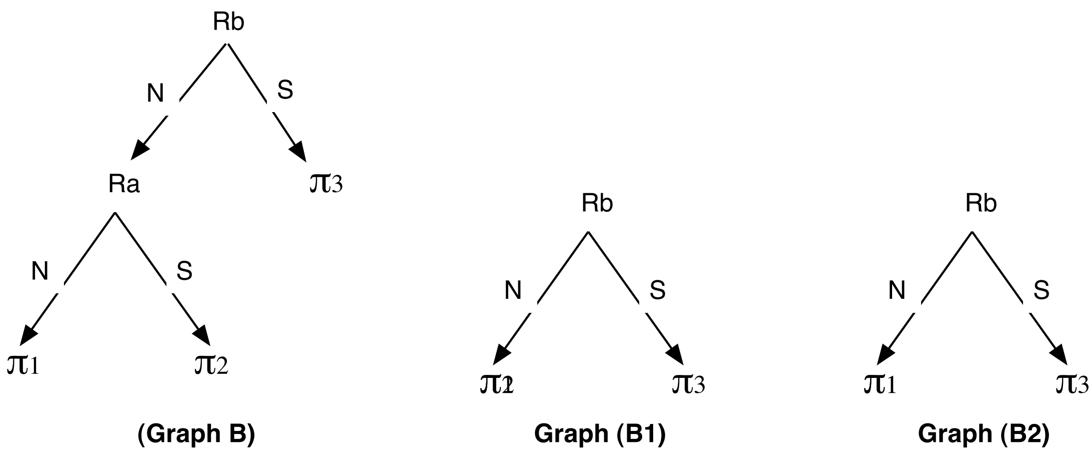

Figure 5: LODAGs for the various scopes of R$_b$

We are now ready to study the combinatory coming from the fusion of LODAGs (Ai) and (Bj). The goal is to distinguish the LODAGs which correspond to discourses *S1 Conn*$_a$ *S2 Conn*$_b$ *S3* from those which do not (i.e. which cannot be linguistically realized).

A) Graph (A). This graph is linguistically realized in (4a)[6]. The wide scope of *Conn*$_a$ = *because* can be seen in the dialogue in (4b-c) in which the answer is *Because S2 Conn*$_b$ *S3*[7]. Moreover, from (4a), *S2 Conn*$_b$ *S3* can be inferred: if (4a) is true, then it is true that *Fred played tuba while Mary was taking a nap*. On the other hand, *S1 Conn*$_a$ *S2* or *S1 Conn*$_b$ *S3* cannot be inferred: if (4a) is true, neither *Mary is in a bad mood because Fred played tuba*[8] nor *Mary is in a bad mood while she was taking a nap* is true. The reader will check that the adverbial *Conn*$_a$ = *therefore* in (4d) has also wide scope.

(4) a. Mary is in a bad mood because Fred played tuba WHILE she was taking a nap.
b. - Why is Mary in a bad mood?
c. - Because Fred played tuba WHILE she was taking a nap.
d. Fred wanted to bother Mary. Therefore, he played tuba WHILE she was taking a nap.

B) Graph (B). This graph is linguistically realized in (5a). The wide scope of *Conn*$_b$ = *in order that/to* can be seen in the dialogue in (5b-c) in which the question is *Why S1 Conn*$_b$ *S2?* From (5a), *S1 Conn*$_a$ *S2* can be inferred, while neither *S2 Conn*$_b$ *S3* nor *S1 Conn*$_b$ *S3* can be so. The adverbial *Conn*$_b$ = *therefore* in (5d) has also wide scope.

(5) a. Fred played tuba WHILE Mary was taking a nap in order to bother her.[9]
b. - Why did Fred play tuba WHILE Mary was taking a nap?
c. - In order to bother her.
d. Fred played tuba WHILE Mary was taking a nap. Therefore, she is in a bad mood.

C) Graphs (A1) and (B1). The fusion of (A1) and (B1) leads to LODAG (C) in Figure 6. This LODAG is not tree shaped: $\pi 2$ has two parents, it is the satellite of R$_a$ and the nucleus of R$_b$. It is linguistically realized in (6a), in which S2 is said to be "factorized" since both *S1 Conn*$_a$ *S2 = Mary is in a bad mood because her son is ill* and *S2 Conn*$_b$ *S3 = Her son is ill. Precisely, he has an attack of bronchitis* can be inferred from (6a). On the other hand, *S1 Conn*$_b$ *S3 = Mary is in a bad mood. Precisely, her son has an attack of bronchitis* cannot be inferred. A similar situation is observed in (6b) and (6c).

(6) a. Mary is in a bad mood because her son is ill. Precisely, he has an attack of bronchitis.
b. Fred played tuba. Next he prepared a pizza to please Mary.
c. Fred was in a foul humor because he hadn't slept well that night because his electric blanket hadn't worked.[10]

D) Graphs (A1) and (B2). The fusion of (A1) and (B2) leads to LODAG (D) in Figure 6. This LODAG is not tree shaped: $\pi 1$ has two parents, it is the nucleus of both R$_a$ and R$_b$. It is linguistically realized in (7), in which S1 is said to be "factorized" since both *S1 Conn*$_a$ *S2 = Fred prepared a pizza to please Mary* and *S1 Conn*$_b$ *S3 = Fred prepared a pizza. Next he took a nap* can be inferred. On the other hand, *S2 Conn*$_b$ *S3* cannot be inferred.

(7) Fred prepared a pizza to please Mary. Next, he took a nap.

Let us underline the following point. Interpreting tree shaped graphs (A) and (B) with the "nuclearity principle" amounts to interpreting (A) as (C), and (B) as (D)[11]. But then, cases with

[6]To indicate that it is stressed when spoken, the word *while* is written in capital letters in (4).

[7]When *while* is not stressed, the question in (4b) may be given as answer only *Because S2*. The interpretation of (4a) corresponds then to LODAG (C) in Figure 6. The ambiguity of discourses such as (4a) or (5a) below, which include two subordinating conjunctions, is discussed in detail in (Danlos, 2003).

[8]Maybe, Mary enjoys tuba, but not when she is taking a nap.

[9]When *while* is not stressed, the interpretation of (5a) may correspond to LODAG (D) in Figure 6.

[10]This discourse is a modified version (including discourse connectives) of an example taken in (Blackburn & Gardent, 1998). These authors acknowledged that the structure of this discourse is a "re-entrant graph".

[11]With the nuclearity principle, the satellite of R$_a$ in (A) is $\pi 2$, and the nucleus of R$_b$ in (B) is $\pi 1$.

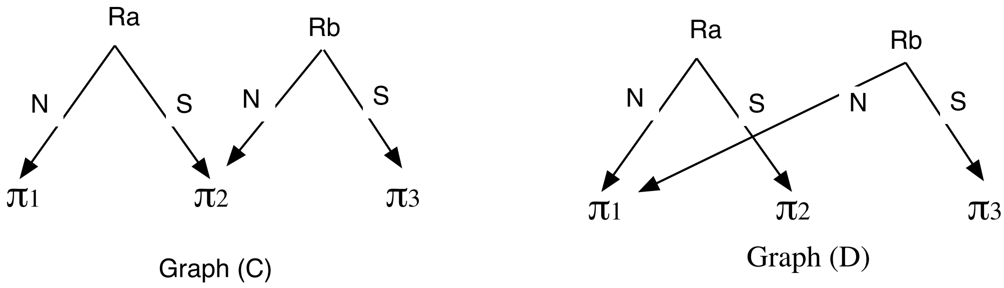

Figure 6: Graphs (C) and(D)

wide scope are not represented at all, they are not taken into account, which is unacceptable. This is the reason why trees should be interpreted in the standard way and not with the nuclearity principle.

E) Graphs (A2) and (B1). The fusion of (A2) and (B1) leads to LODAG (E) in Figure 7, in which $\pi 3$ has two parents. I cannot find any discourse corresponding to (E), i.e. with *S3* factorized.

F) Graphs (A2) and (B2). The fusion of (A2) and (B2) leads to LODAG (F) in Figure 7. This graph cannot represent a discourse *S1 Conn*$_a$ *S2 Conn*$_b$ *S3* since it does not include π_2.

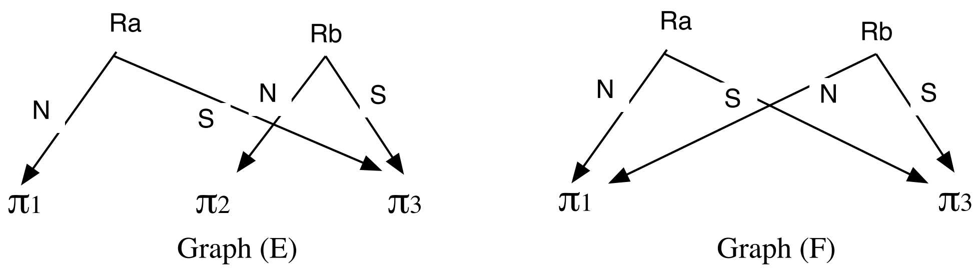

Figure 7: Graphs (E) and (F)

So far, we have examined only cases where a given discourse relation has only one satellite and one nucleus. We are left with multi satellite or nucleus cases.

G) Graphs (A1), (A2) and (B2). The fusion of (A1), (A2) and (B2) leads to LODAG (G) in Figure 8. This LODAG could be said to be linguistically realized in (8a). In (8a), R$_a$ = Elaboratio has two satellites, π_2 and π_3, since both *S1* $Conn_a$ *S2* and *S1* $Conn_a$ *S3* can be inferred. R$_b$ = Narration links π_2 and π_3. The following question arises: is R$_b$ in a satellite dependency relation with R$_a$? It is hard to give an answer for (8a). However the answer seems positive for (8b) with R$_a$ = Explanation, R$_a(\pi 1, \pi 2)$ and R$_a(\pi 1, \pi 3)$ (multi-satellite case), R$_b$ = Joint and R$_b(\pi 2, \pi 3)$. This leads to LODAG (G') in Figure 8. However, consider (8c) which differs from (8b) only by the use of *or* instead of *and*. Graphs (G) or (G') would not do justice to this discourse: neither $R_a(\pi_1, \pi_2)$ nor $R_a(\pi_1, \pi_3)$ can be laid down. (8c) can only be represented as graph (A) (repeated in Figure 8) with R$_a$ = Explanation and R$_b$ = Disjunction.

(8) a. Guy experienced a lovely evening last night. More precisely, he had a fantastic meal. Next he won a dancing competition.[12]
b. Mary is in a bad mood because she had'nt slept well and it is raining.
c. Mary is in a bad mood because she had'nt slept well or it is raining.

[12] This discourse is a modified version (including discourse connectives) of an example taken in (Lascarides & Asher, 1993).

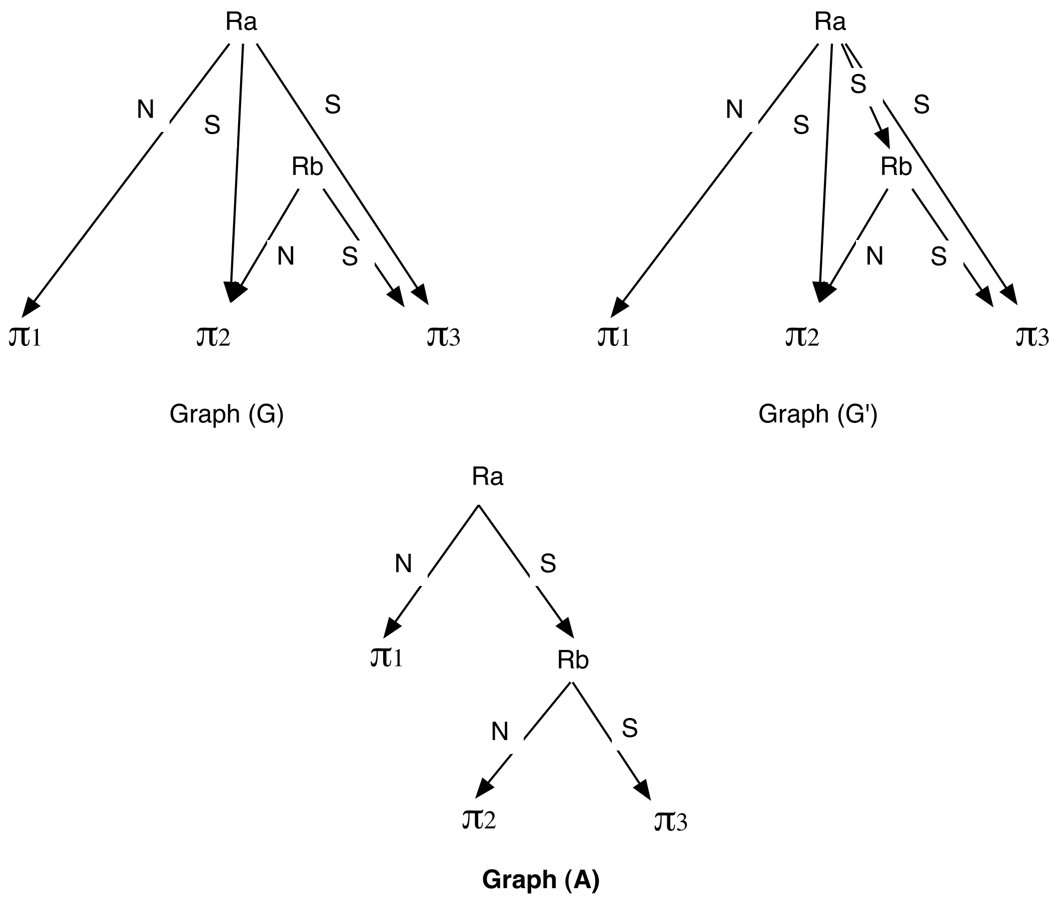

Figure 8: Graphs (G), (G') and (A) (repeated)

We touch here a crucial question in discourse processing (within a multi-level approach): to what extent the semantic (dependency) level (*how things are said*) should echo the informational content level (*what is said*)? I don't pretend to give a general answer to this fundamental question. However, I claim that (8b) and (8c) should be represented at the semantic level as the very same graph. This graph can only be (A), which is the only possibility for (8c). For the sake of homogeneity, (8a) should also be represented as (A)[13]. Recall moreover that (4a) with wide scope of $Conn_a$ is also represented as (A). All in all, (A) happens to be a semantic representation which is shared by discourses whose informational content is quite different. Is it a problem? I would say no, because, from (A), semantic to content rules, based on the values of R_a and R_b, can make the difference: they can compute the following (simplified) logical forms, which show that the discourses in (8) and (4a) do not have the same type of informational content, althoug they share the same (dependency) semantic representation:

- for (8a) with R_a = Elaboration and R_b = Narration: $e_1 \wedge e_2 \wedge e_3 \wedge precede(e_2, e_3) \wedge subevent(e_2, e_1) \wedge subevent(e_3, e_1)$.
- for (8b) with R_a = Explanation and R_b = Joint: $e_1 \wedge e_2 \wedge e_3 \wedge cause(e_1, and(e_2, e_3))$, i.e., $e_1 \wedge e_2 \wedge e_3 \wedge cause(e_1, e_2) \wedge cause(e_1, e_3)$
- for (8c) with R_a = Explanation and R_b = Disjunction: $e_1 \wedge e_2 \wedge e_3 \wedge cause(e_1, or(e_2, e_3))$, i.e., $e_1 \wedge e_2 \wedge e_3 \wedge (cause(e_1, e_2) \vee cause(e_1, e_3))$
- for (4a) with R_a = Explanation and R_b = Circumstances [*while*]: $e_1 \wedge e_2 \wedge e_3 \wedge overlap(e_2, e_3) \wedge cause(e_1, overlap(e_2, e_3))$

[13]This analysis corresponds to the SDRS proposed by (Lascarides & Asher, 1993) for (8a) modulo some adjustment explained at the end of this section.

H) Graphs (A1), (B1) and (B2). The fusion of (A1), (B1) and (B2) leads to LODAG (G) in Figure 9. This LODAG could be said to be linguistically realized in (9). In (9), R_b = Circumstances has two nuclei, π_1 and π_2, which are linked by R_a = Joint. However, as in the previous case, one may well wonder if the graph for (9a) should not be Graph (H') with a nucleus relation between R_b and R_a or graph (B) with wide scope of R_b. By the same argumentation as previously, graph (B) should be adopted.

(9) Fred washed the dishes and Guy cleaned up the bathroom, while Mary was taking a nap.

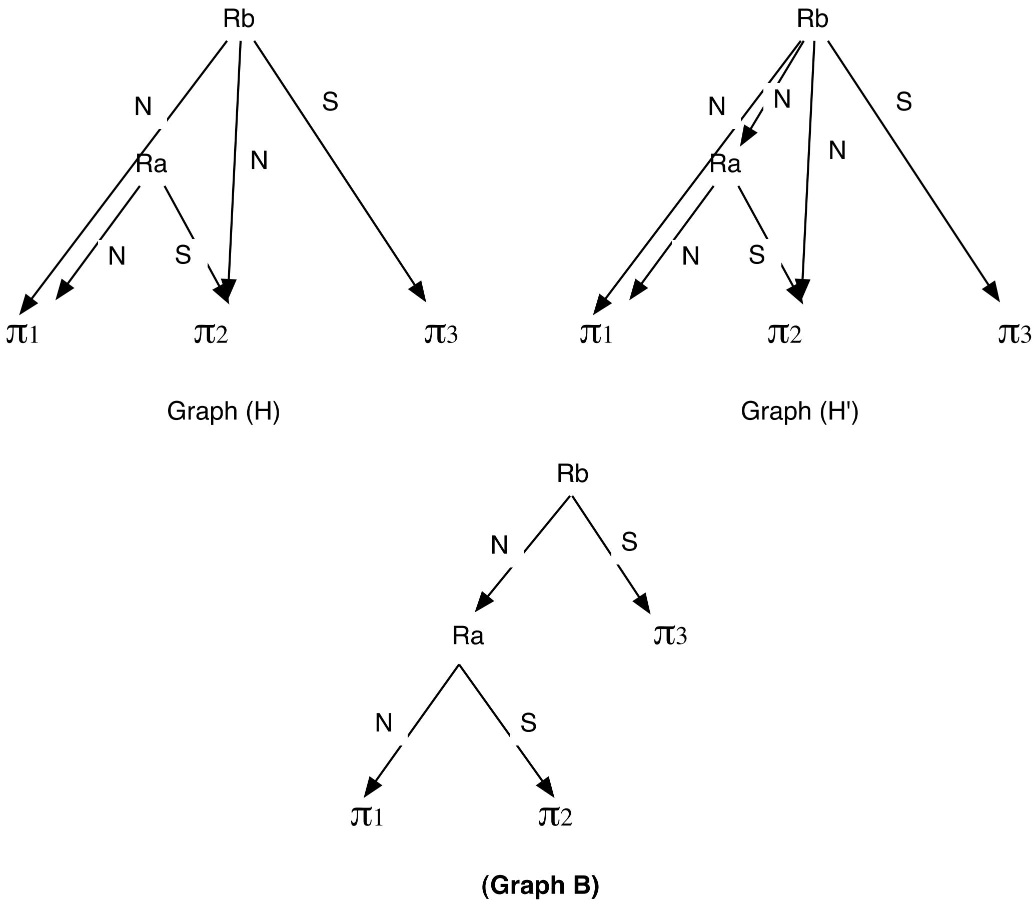

Figure 9: Graphs (H), (H') and (B) (repeated)

I) Graphs (A1), (A2) and (B2). The fusion of these graphs lead to LODAG (I) in Figure 10. I cannot find any example corresponding to this LODAG, which would be a case of multi satellites for the same relation (π_2 and π_3 are both satellites of R_a) with no link between the satellites but with a link between the factorized nucleus (π_1) and the non adjacent satellite (π_3).

J) Graphs (A2), (B1) and (B2). Along the same lines, the fusion of these graphs lead to LODAG (J) in Figure 10 and I cannot find any example corresponding to this LODAG.

No other fusion of graphs (A_i) and (B_j) leads to a LODAG which corresponds to a discourse. So we have arrived at the following result.

> The dependency structure of a discourse *S1 Conn$_a$ S2 Conn$_b$ S3* is one of the four LODAGs (A), (B), (C) and (D). (A) and (B), which are tree shaped, cover wide scope cases (including multi satellite or nucleus cases), (C) and (D), which are not tree shaped, cover multi parent cases (factorization of a sentence).

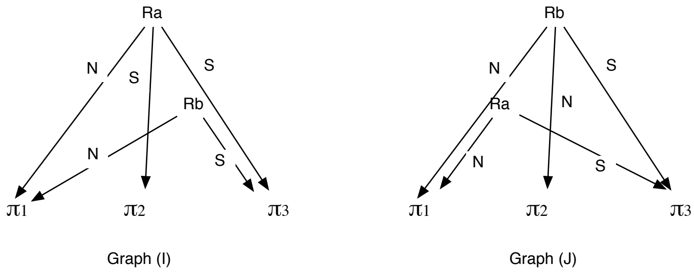

Figure 10: Graphs (I) and (J)

Before commenting on this result, let us examine the semantics-syntax interface in MTT when both R$_a$ and R$_b$ are lexicalized by subordination conjunctions. While the non tree shaped graphs (C) and (D) do not raise any problem[14], the trees (A) and (B) do so. In these trees, the root, a discourse relation node, is the mother of another discourse relation node. In the standard case, which leads to a staightforward semantics-syntax rule, a discourse relation node is the mother of sentential nodes. Therefore a specific rule has to be called upon in the semantics-syntax interface so that the syntactic tree obtained from (A) (respectively (B)) is equivalent to the tree obtained from (C) (respectively (D)) (Danlos, 2003).

Consider now SDRT. The interpretation given in the tree shaped graph (A) corresponds to the SDRS shown in the leftmost box of Figure 11. This SDRS can be schematized as in graph (A') in which the tree rooted at R$_b$ is circled and labelled by π_4 (see note 4). The interpretation given in the non tree shaped graph (C) corresponds to the SDRS in the bottom part of Figure 11. LODAGS (B) and (D) also correpond to well-formed SDRSs. In conclusion, LODAGS (A)-(D) translate straightforwardly in SDRT, modulo a circling and labeling operation for discourse relations with wide scope (e.g., transforming (A) as (A')). The reader can check that all the LODAGS proposed for the discourses we have analyzed are compatible with a SDRT analysis.

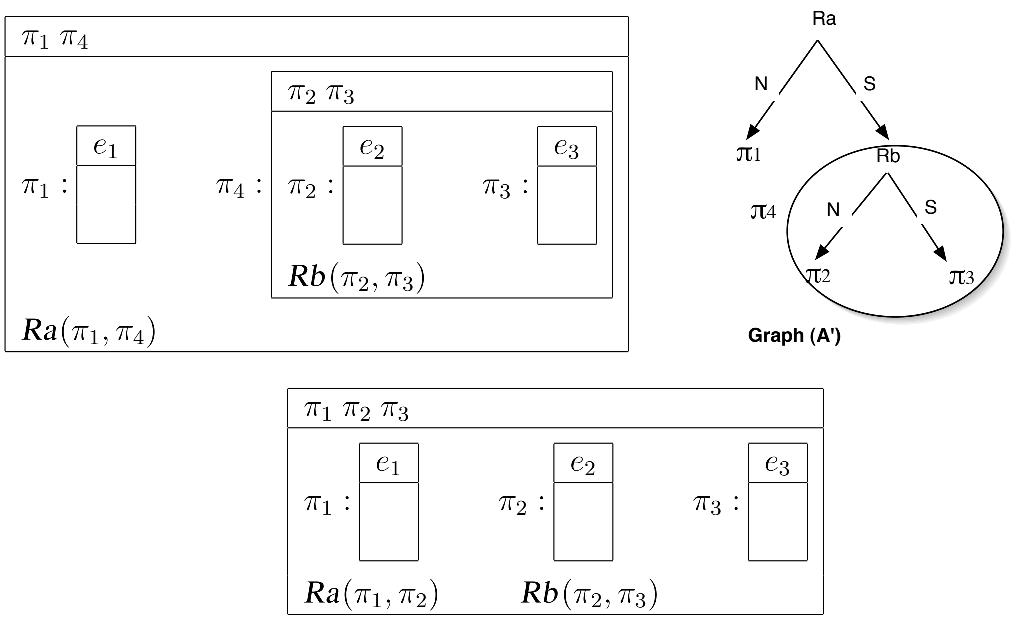

Figure 11: SDRS corresponding to (A), graph (A'), SDRS corresponding to (C)

[14]In MTT, there is no problem in transforming a non tree shaped semantic LODAG into a syntactic tree (Candito & Kahane, 1998).

3 Analysis of the results and conclusion

The summary of the results we arrived at does not take into account the discourse connectives / relations at stake. However, for a given pair of connectives, it may happen that only some of these LODAGs are observed. For example, if *Conn*$_a$ is an adverbial and *Conn*$_b$ a subordinate conjunction, then (B) with wide scope of R_b should be excluded. On the top of part of speech considerations, the lexical value of each connective may exclude some of these LODAGs. Finally, the distinction between "coordinating and subordinating" discourse relations (Lascarides & Asher, 1993) must certainly be taken into account[15]. Technically, it could be done by considering that there are two kinds of branching nodes in LODAGs. Then, it ought to be established which kind(s) of branching nodes is (are) allowed in each LODAG (A)-(D). To conclude, there is a maximum of four LODAGs representing the semantic analyses of discourses *S1 Conn*$_a$ *S2 Conn*$_b$ *S3*. A linguistic (corpus) study should refine this result according to the specific values of *Conn*$_a$ and *Conn*$_b$. I stipulate that theses results can be extrapolated to cases where sentences are simply juxtaposed without discourse connective.

It can be considered that there is only a few LODAGs corresponding to discourses with three clauses[16]. A number of *a priori* possible LODAGs do not correspond to coherent discourses, e.g. LODAG (E). As a consequence, some structural constraints can be brought out, as the following one: *R*$_a$ *must "left-dominate"* π_2. The notion of left-dominance is introduced in (Danlos, 2003)[17]. Let us here examine the consequences of this left-dominance constraint in non formal terms. R_a is the mother of π_1 (Section 2) and must left-dominate π_2. This means that R_a establishes some semantic link between *S1* and *S2*[18]. This result may sound trivial on psycho-linguisitics grounds: what would be a discourse in which the second sentence is not linked at all to the first one?[19] It has an important consequence: the semantic representation of a discourse with n clauses, $n \succ 3$, cannot be or include LODAG (K) in Figure 12, since in (K) R_a does not left-dominate π_2, or informally, there is no link between *S1* and *S2*. A graph such as (K) is a "cross dependency" case in the terms of (Webber *et al.*, 2003) who stipulate that "*discourse structure itself does not admit crossing structural dependencies"*. I agree with them, but not on a stipulative ground. To conclude, semantic dependency structures for discourses are LODAGs but these LODAGs present some structural constraints, which can help us to cut down the number of possibilities when processing discourses, be it in analysis or generation, or be it in SRDT or in an extension of MTT to discourses.

[15]Roughly, Explanation and Elaboration are subordinating relations while the others are coordinating ones.

[16]In RST, there are only 2 trees (2 is the number of binary trees with 3 leaves), namely trees (A) and (B), which are supposed to be interpreted with the nuclearity principle (being so interpreted as (B) and (D) respectively). We have seen that this is too restrictive: wide scope cases are not taken into account (neither multi satellite or nucleus cases).

[17]The definition of left-dominance is the following: a node X in a tree left-dominates a node Y iff Y is a daughter of X (immediate dominance) or there exists a daughter Z of X such that Y belongs to the left-frontier of the tree rooted at Z. For example, R_a left-dominates $\pi 1$ and $\pi 2$ in (A), while R_b left-dominates $\pi 1$ and $\pi 3$ in (B).

[18]This result can be seen as a weaken version of the adjacency constraint in RST, i.e. two adjacent clauses are linked by a discourse relation. This constraint does not hold because it is too restrictive: for example, in discourses analyzed as (D), *S3* is linked to *S1* (which is not adjacent) and not to *S2* (which is adjacent).

[19]The link between the first two sentences can be given by a third sentence, as in (10) in which *S3* establishes a joint link between *S1* and *S2* through its plural subject. The LODAG for (10) is topologically (B).

(10) It is raining. Ted arrived late. These two facts, which are not related, irritated Mary.

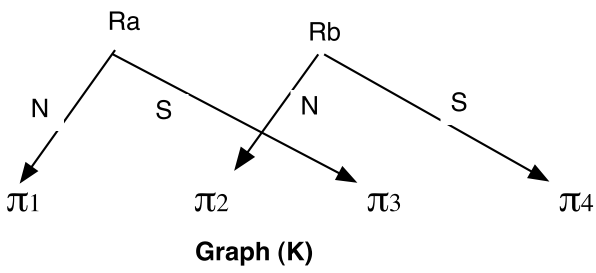

Figure 12: Graph (K)

References

ASHER N. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.

ASHER N. & LASCARIDES A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press. (forthcoming).

BLACKBURN P. & GARDENT C. (1998). A specification language for discourse semantics. In *Proceedings of LACL'98*, p. 61–67, Grenoble.

CANDITO M.-H. & KAHANE S. (1998). Can the TAG Derivation Tree represent a Semantic Graph? In *Proceedings of TAG+4*, p. 25–28, Philadelphia.

DANLOS L. (2003). Représentation sémantique sous-spécifiée pour les conjonctions de subordination. In *Actes de TALN 2003*, Batz-sur-Mer.

KAMP H. & REYLE U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.

LASCARIDES A. & ASHER N. (1993). Temporal interpretation, discourse relations, and common sense entailment. *Linguistics & Philosophy*, **16**(5), 437–493.

MANN W. C. & THOMPSON S. A. (1987). Rhetorical structure theory: Description and construction of texts structures. In G. KEMPEN, Ed., *Natural Language Generation*, p. 85–95. Dordrecht: Martinus Nijhoff Publisher.

MARCU D. (1996). Building up rhetorical structure trees. In *The Proceedings of the 13th National Conference on Artificial Intelligence*, p. 1069–1074, Portland.

MEL'CUK I. (1997). *Vers une linguistique Sens-Texte*. Paris: Leçon inaugurale au Collège de France, Collège de France,.

WEBBER B., JOSHI A., STONE M. & KNOTT A. (2003). Anaphora and discourse structure. *Computational Linguistics*, **44**, 1–45.