

# The Use of Referential Constraints in Structuring Discourse

Violeta Seretan<sup>♦</sup>, Dan Cristea<sup>♦</sup>

<sup>♦</sup>University of Geneva, Language Technology Laboratory  
2, rue de Candolle, CH-1211 Genève 4, Switzerland  
Violeta.Seretan@lettres.unige.ch

<sup>♦\*</sup>"Al.I.Cuza" University of Iași, Dept. of Computer Science  
16, Berthelot St., 6600 – Iași, Romania  
dcristea@infoiasi.ro

## Abstract

The quality of discourse structure annotations is negatively influenced by the numerous difficulties that occur in the analysis process. In contrast, referential annotation resources are considerably more reliable, given the high precision of the existent anaphora resolution systems. We present an approach based on the Veins Theory (Cristea, Ide, Romary, 1998), in which successful reference annotations of texts are exploited in order to improve arbitrary structural analyses; in this way, the large amount of corpora annotated at reference level can be used for the acquisition of discourse structure annotation resources.

## 1. Introduction

Discourse analysis is an important but difficult task, requiring a lot of effort in order to capture the relations between text constituents and thus to decode the author's communicative intentions.

The popular theory of discourse structure, Rhetorical Structure Theory (RST; Mann and Thompson, 1988) is proved to account for the manner in which the text is organized in correspondence with the speaker's intention. Recent research on the relationship between the structure of the text and intentions showed (Moser and Moore, 1996; Marcu, 1999) the similarity between the intention-based discourse structure (Grosz and Sidner, 1986) and the RST tree-like structure built for a text.

Despite its popularity and usability, RST lacks important prescriptions on criteria for the hierarchical aggregation of the pieces of text, this making structural analysis an ambiguous task and leading to inappropriate text interpretations.

Many elements provide hints on the discourse structure, e.g. delimiters, cue-phrases, time etc., that are extensively used for the automatic computation of the coherence relations (Marcu 1997, Kurohashi and Nagao, 1997).

We will focus on those elements that better indicate the structure of a text: the co-references in it. A reference from an anaphor to its antecedent indicates a structural relation between the textual units involved. The referential chains in discourse (the repeated references to the same discourse entity) contain important information about the text organization; therefore, they should also be considered when structuring discourse.

While largely agreed that there is a straight relation between the references in a text and its structure (Fox, 1987; Vonk et al., 1992), up to now most of the attention was concentrated in only one direction, i.e. on the way in which the process of anaphora resolution is influenced by the hierarchical organization of the text. For example, in (Fox, 1987) it is indicated that the treatment of anaphora should consider the hierarchal structure of texts, and in (Cristea et al., 2000) it is shown that the anaphora resolution can benefit from following a discourse organization.

We concentrate, instead, on the way the use of referring expressions restricts the discourse interpretation, and we intend to use them as disambiguation clues during structure derivation, or as structural constraints for correcting arbitrary (possibly automatically generated) analyses.

A reliable reference annotation can be used to impose constraints on the partial or complete discourse structure built for a text, by means of the prescriptions on the relationship between the discourse structure and references stated in Veins Theory (VT; Cristea, Ide, Romary, 1998). According to these prescriptions, the reference chains from text are associated to sets of structurally related units, the "veins" of discourse. The references from a given unit are mostly to preceding units that are contained in the unit's vein.

In a well-formed structure, anaphora are expected to be resolved along the veins. If this is not the case, it is likely that errors have occurred in the process of interpretation, which disrupted the structural connection, i.e. the path given by the vein, between the pairs of concerned units.

The goal of our approach is to systematically detect and correct the structural errors which are likely to occur during the structural annotation, and which can be signalled by the resolution of anaphors outside their domains of accessibility as given by the structure.

The next section briefly revises the Veins Theory main concepts and ideas on which our approach is based. It is followed by a section that describes in which way the VT prescriptions are used in better structuring discourse, and presents a method for the local and global correction of a structure. Afterwards, we present the results of an empirical study on a corpus of texts, the conclusions it allows us to draw and, finally, the comparison with the related work.

## 2. Global Discourse Cohesion in VT

Veins Theory extends and formalizes the relation between discourse structure and reference proposed by Fox (1987). Its central notion is the "vein", defined over discourse structure trees built according to the RST requirements.

## 2.1. The Vein Concept

VT's fundamental assumption is that references in a text occur mostly between units that are in structural relation with one another, even if they are distant in the text.

The sets of structurally related units form the main threads of discourse, called *veins*, and are defined on the basis of the nuclearity of the constituents in the discourse structure built for the text. They express the idea that, in order to understand a unit in the context of the whole text, only part of the discourse units, including the one under examination, are required. Considering that these units form a chain, all references from the examined unit should be resolved along the sub-chain of preceding units. In a left-polarized tree, the references are mainly to nuclei rather than to satellites. In addition, the vein expression

accounts for the cases when the discourse structure is not left-polarized, that is, when a satellite can precede a nucleus. In such a case, the antecedents from the satellite should also be accessible for further referring, as long as the anaphor's unit is not descendent of a right satellite node. The interposition of a nuclear node blocks the accessibility between the left satellite and the right subsequent satellites.

We illustrate the computation of the vein on an example taken from the corpus we examined (see Figure 1). Unit 7, for instance, is contained in the vein formed by units 1, 2 (that are nuclei in the most important preceding nodes), 6 (the sister that is a left satellite), and 7 (the unit itself). The vein of unit 8 comprises units 1, 2, 7 (as before), but not unit 6, as discussed before (nuclear unit 7 interpose between units 6 and 8).

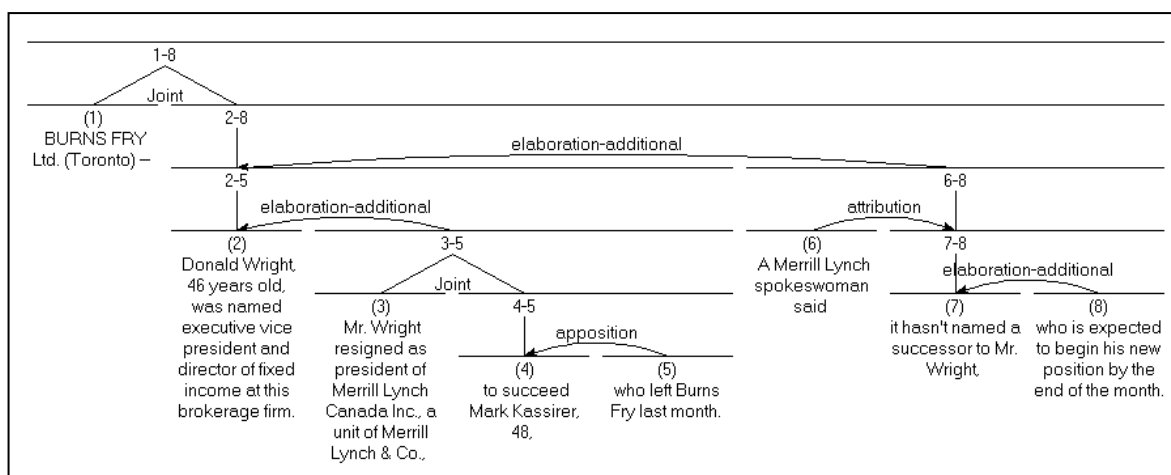


Figure 1: The representation of the RST analysis for a text<sup>1</sup>.

## 2.2. Direct and Indirect Vein References

The reference annotation of texts widely adopts the convention of textual proximity: if an entity is referred more than once in a text, the co-referential links are marked from each anaphor to the closest antecedent in the text that points to the same entity. In the text in Figure 1, for example, the entity "Mr. Wright" is referred to 3 times, in units 2, 3 and 7, and the co-referential links are marked from unit 3 in unit 2 and from unit 7 in unit 3.

This process doesn't follow a discourse structure. Obviously, the pairs of units involved in an anaphoric relation are structurally connected by some path along the coherence relations; the closest the antecedent along this path, the more entitled to be considered as the right target of the co-reference link. One could find more intuitive to mark, in our example, the link from unit 7 directly to unit 2.

VT considers that the relation of co-reference induces equivalence classes over the set of referential expression, and distinguishes between *direct* and *indirect* vein references.

A *direct* reference situation is that in which the unit of the closest antecedent belongs to the vein of the anaphor's

unit. An example is the reference to "Mr. Wright" from unit 3 to unit 2: the vein of unit 3 is "1 2 3 4".

In the case of *indirect* reference, the unit of the closest antecedent of an anaphor is not on the vein of anaphor's unit. However, a more distant antecedent exists in a unit belonging to the anaphor's unit vein. It is the case of the reference to "Mr. Wright" from unit 7: the closest antecedent is in unit 3, which is not contained in the vein expression of unit 7 ("1 2 6 7"), but a previous antecedent in the same reference chain is in unit 2, present in the vein.

In the first case, the target of the co-reference that was indicated by the annotation corresponds indeed to the structurally signalled target. In the latest case, the target proposed, i.e. the linearly closest antecedent in the list, was not the closest one along the structural path.

For both direct and indirect reference types, the anaphor is resolved on its corresponding vein, henceforth this situation is referred to as *vein resolution*.

## 2.3. Vein Resolution Exceptions

VT indicates that most of the references from a coherent text will be resolved along the veins, directly or indirectly, thus being easier and quicker to interpret.

<sup>1</sup> The figure is depicted using the RST-Tool (O'Donnell, 1997).

The references for which no antecedent from the same chain occurs on the vein are supposed to be of *pragmatical* nature, in the sense that the anaphor can be understood without any antecedent, using the general world knowledge, like if it were introduced for the first time in discourse.

Anyhow, few cases are reported, in (Cristea et al., 2000) for instance, where out-of-vein resolution exceptions are not of a pragmatical nature. It is unclear whether they are due to error in the structure annotation.

We tend to believe that such an exception occurs when the relation between the two parts of text given by the reference is not conveyed by the structure, and that, probably, a non-valid interpretation was produced in the process of discourse analysis.

This is, for example, the case of the structure in Figure 1. The reference to the entity "Merrill Lynch Canada Inc." from unit 6 to unit 3 is an exception, since unit 3 is not on the vein of unit 6, which is "1 2 6 7". Later we will see that this exception is due indeed to structure misconfiguration.

### 3. From Cohesion To Coherence Using VT

Our goal is to perform slight modifications on the current structure configuration (which can be either partial or complete), in the areas indicated by the exceptions in the vein resolution, in order to enable the referential accessibility between the two parts of text involved and to reconstruct the structural relation that exists between them.

The content of this section is the following: we will first look at the causes that could lead, during the discourse analysis, to structures where a hierarchical relation between two units is not allowed, in spite of the existence of a co-reference from one to another.

Corresponding to the nature of these causes, we will present several correcting operations which will be applied on the affected structure, that aim at its reconfiguration so that it satisfies the constraints imposed by the references. Furthermore, we will show how these operations on the structure influence the vein-resolution for the involved references.

Finally, we show how the local and global corrections are applied, looking at the references in the text, in order to reconfigure an arbitrary structural analysis so that the hierarchical relations obey to the constraints given by the use of references.

#### 3.1. Misleading Factors in Discourse Structure Analysis

With respect to text structure, different theories exhibit many commonalities. For instance the "fix-point" of any RST-like analysis is given, at least, by: the elementary units of text structure are non-overlapping spans of text, some textual units play a more important role than others, the abstract structure of a text is a tree (Marcu, 2000). But it is also well known that, more often than not, more than just one analysis could be drawn, as one should face at least the inherent ambiguities of rhetorical structure and the scarcity of elaborated theoretical knowledge on the way hierarchical aggregation of structure constituents in an RST tree is to be pursued. Indeed, RST ambiguity arises in what concerns:

- the way the elementary textual spans are defined;

- the way the constituents of a rhetorical relation are identified;
- whether the speaker intends to assign, for some relations, a more important role to one component or to the other (which constituent is "nucleus" and which one is "satellite");
- under what conditions two spans of text can be combined into a higher structure, using a rhetorical relation.

Like in VT, we consider the structure of discourse being that of a binary tree, therefore we see a rhetorical relation holding between two sibling descendants. Also like in VT, we ignore the name of relations while keeping at value their polarity given by the nucleus-satellite dichotomy. As such, we abstract away from the relation name ambiguity as well as from the dispute on relations taxonomy, focusing instead on issues of text interpretation that are common to all structural theories. We claim that doing that way the approach gains in generality and provides a wider range of applicability. If the intention is to acquire a higher specificity, a mapping to a given set of relations could, ultimately, be added.

Between the factors previously mentioned, the one that we consider that mostly renders difficult the process of analysis is the size of relation's constituents. Human annotators generally agree on the segmentation of text in elementary units and on the relative importance of the constituents (Marcu, Amorrortu, and Romera, 1999), but, in what concerns the spans of text the relation entails, the risk of mistakes is much higher.

Leaving apart the segmentation of discourse, we consider that, in one step of discourse analysis, two possible types of choices affect the well-formedness of structure configuration, in the absence of criteria for compositionality:

- the manner of assigning the nuclear roles for constituents,
- the manner of associating sub-structures in a bigger structure, at a higher level in the hierarchy.

#### 3.2. Discourse Structure Errors indicated by Exceptions: two Examples.

We provide two representative examples in which the ambiguities of the above mentioned nature led to construction errors that were then signalled by the vein resolution exceptions.

The first example, related to the hierarchical ambiguity, is represented by the structure in Figure 1 in section 2.1. The annotator improperly associates two constituents in the hierarchy, adjoining the structure 6-8 in the root of the already created structure 2-5, and not in the node 3-5 situated below on the right frontier. The vein of unit 6 is then affected so that it doesn't contain the unit 3. Consequently, the reference to "Merrill Lynch" from unit 6 to unit 3 appears like an exception.

The correct target for the adjunction of substructure 6-8 to the substructure 2-5 is the node 3-5, as in Figure 2. Actually, the whole span (2-5) is not in the relation of *elaboration-additional* with the substructure 6-8, but only its subspan, 3-5, because the new topic ("no successor was named to Mr. Wright") elaborates only the topic in subspan 3-5 ("Mr. Wright resigned as president of Merrill Lynch"), and not the topic of the whole span ("Donald Wright was named executive prime vice president at

Burns Fry"). This is consistent with the satisfaction of the Compositionality Criterion (Marcu, 1997), which says that a relation that holds between two spans holds also between the most salient units of the constituent spans: the relation also holds between units 3 and 7, while, in the

initial structure, this criterion is not obeyed: the relation doesn't hold between units 2 and 7.

In the correct structure, the reference from unit 6 to unit 3 becomes direct reference on the vein, since the correct vein expression for unit 6 is "1 2 3 4 6 7".

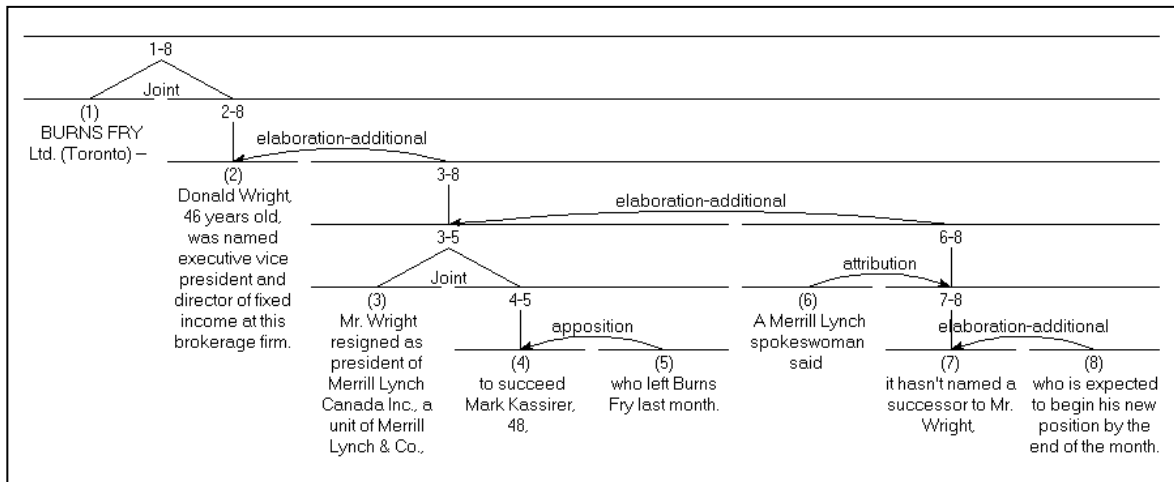


Figure 2: The correct structure proposed for the original structure from Figure 1.

The second example is related to the nuclearity ambiguity. Let's suppose that one assigns a role of S (*satellite*) to a left constituent of a relation instead of N (*nucleus*), by choosing either a relation that is inadequate (from a set of several possible relations), or one whose RST nuclear role assignment is uncertain<sup>2</sup>.

It could happen that the vein accessibility from an anaphor to an antecedent be blocked from further units, like in the text below extracted from our corpus:

- (1) At Chrysler, Mr. York had been a dark-horse candidate to succeed departing **chairman** Lee A. Iacocca.
- (2) Chrysler ultimately chose former General Motors Corp. executive Robert Eaton.
- (3) The succession struggle left relations somewhat strained between Mr. York and Chrysler president Robert Lutz,
- (4) who had been another contender for **the top job**.

The original structure associated by the annotators to the text was that in Figure 3, where the nuclear constituents were underlined. The relations *r1*, *r2* and *r3* were originally identified as, respectively, *background*, *elaboration-object-attribute*, and *consequence*:

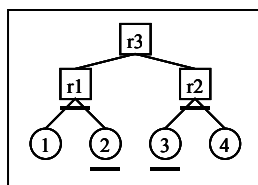


Figure 3: The RST structure of the text in Example 1, represented as a binary tree.

<sup>2</sup> In several cases, e.g. for the *consequence* relation, it is ambiguous as to which unit is nucleus and which one is satellite (Marcu, 1997).

The referential expression "the top job" in unit 4 refers back to the "chairman" position mentioned in unit 1. Still, a reference from unit 4 in unit 1 is not allowed, since unit 1 is not on the vein of unit 4, whose expression is "2 3 4". If unit 1 were nuclear, the reference would have been possible. When verifying the structure, we found that relation *r1* is better interpreted as *sequence*, therefore a binuclear relation, since the intentions realized by its constituents are in a relation of satisfaction-precedence and not dominance, in terms of the intentional structure (Grosz and Sidner, 1986). The correction modifies the vein expression of unit 4 to be "1 2 3 4", therefore allowing a reference from 4 to 1.

There are multiple situations in which the particular choices at each step of structure derivation affect the vein accessibility for certain units of text. We studied the effect of all types of choices over the vein expression of text's units and over the vein-resolution of anaphora (Serețan, 2000).

### 3.3. Veins-Guided Structure Recovery

In the process of correction of an arbitrary discourse structure, we considered two main types of modifications: an operation related to the hierarchy of constituents in the tree-like text structure, and an operation related to the nuclearity of constituents. They correspond to the two choices possible in one step of analysis.

The manner in which these operations will be applied on the discourse structure will be discussed later, in the subsection 3.4.

#### 3.3.1. Basic Recovery Operations

*Structural operations.* The first type of structural modification is relatively complex and is inspired by the operations from Tree Adjoining Grammars (Joshi, 1987). All operations obey the *sequentiality principle* (Cristea and Webber, 1997), according to which at any time during analysis the sequence of nodes on the terminal frontier of the tree corresponds to the sequence of discourse units in

the original text. It consists of two elementary operations on the substructure involved:

- i. a *cut* operation, of extraction of a sub-tree from one substructure's daughter;
- ii. a *paste* operation, of adjunction of the tree obtained at the previous step, as left or right auxiliary tree, on the opposite (right, respectively left) internal frontier<sup>3</sup> of the other substructure's daughter.

Figure 4 below schematizes one type of structural operation, which consists of cutting a sub-tree from the right daughter and adjoining it as left auxiliary tree on the right frontier of the left daughter.

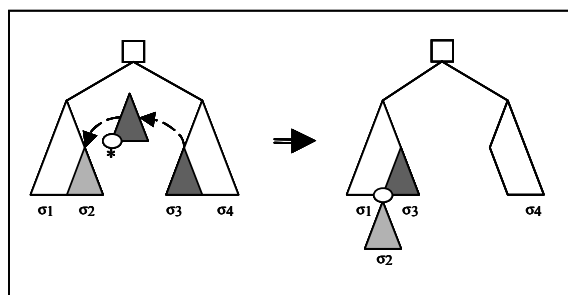


Figure 4: Type of structural operation.

This operation aims at repositioning the constituents of the text in the appropriate place in the hierarchy. It is supposed that the ambiguity on the size of relation's constituents has led to a situation in which, after several steps of analysis, the right constituent of a relation appears as constituent of another, arbitrarily big substructure.

In the first example from subsection 3.2, this kind of operation is applied in order to obtain the correct structure: the node extracted is the whole right sub-tree (6-8) of the structure concerned (2-8).

*Nuclearity operations.* The second type of operation concerns simple modifications of the nuclear roles assignment. The assumption underlying this transformation is that, among several possible relations, the choice did not fall on the most appropriate of them, with consequences on the node's nuclearity.

### 3.3.2. The Effect on the Reference Type

We specified the changes these operations of nuclearity and structural modification have on the vein expressions of the units affected (that are determined by means of the *nuclear path* and *right-satellite condition* notions we used<sup>4</sup>), as well as the effect caused on the vein reference. These were extensively described in (Serețan, 2000).

When the changes in the vein expression are such that new units are added to it, the following transformations are possible for the types of vein reference:

- an exception becomes direct reference;
- an exception becomes indirect reference;

- an indirect reference becomes direct reference<sup>5</sup>.

The opposite transformations take place when removing units from the vein.

We already encountered an example of transformation in subsection 3.2, where an exception became direct vein reference after reconsidering a substructure's place in the hierarchy.

### 3.4. Applying the Operations

In case the proposed discourse structure is such that any given anaphor is resolved on its corresponding vein, it means that the relation between the pairs of the units involved (stated by the reference itself) is structurally well-formed, since the vein is a set of related units in the hierarchy of discourse. Otherwise, the references are exceptions from the point of view of the vein resolution and could indicate misconfigurations of the structure. We then try to apply the operations described on the cases indicated by these exceptions.

A local correction is performed in the substructure determined by an exception, which is represented by the common parent of the two involved units in the hierarchy. The algorithm makes successive tries to put the unit of one of the antecedents on the vein of the unit of the anaphor. An order is considered on the antecedents list, based on the number of other existing references: the more a unit is referred to, the more probable it is to be a nucleus or the target of a structural *paste* operation.

We verify the nuclear roles assignment for the nodes along the path antecedent-root (to see if it is a nuclear path) and root-anaphor (to see if it satisfies the right-satellite condition) and the constituents' association for the nodes on the right and left frontiers of the daughters. We perform either nuclear operation or structural operations, where we found inappropriate relation link.

The global correction method consists of successive error corrections in a linear, bottom-up manner.

## 4. Corpus Study

In our experiment we used a collection of 25 newspaper texts from the MUC corpus (Hirschmann and Chinchor, 1997). They contained a two-level annotation: the original MUC co-reference annotation, and an RST annotation (Marcu, Amorrortu, and Romera, 1999) enriched with information related to the computed veins expressions (Ide and Cristea, 2000).

In a pre-processing phase, we manually verified the referential annotation. We also corrected several segmentation errors in the RST annotation<sup>6</sup>.

We applied the correcting operations like in subsection 3.4 in order to recover the structure, using the constraints of vein resolution of anaphora.

The results of the corpus study showed that most of the resolution exceptions corresponded indeed to mistakes in the structure construction (Marcu's Compositionality Criterion was not obeyed).

<sup>3</sup> The *left (right) internal frontier* of a tree consists of the set of the leftmost (rightmost) non-terminal nodes, at all depths in the tree.

<sup>4</sup> We call *nuclear path* (in a tree-like structure) a path that connects the source and the target along nuclear nodes only. A node in a structure satisfies the *right-satellite condition* if the path that connects it to the root passes through a node that is the right satellite of a relation.

<sup>5</sup> An indirect reference can also be changed so that, although remaining indirect, it is resolved in a unit that is closer on the vein.

<sup>6</sup> We detected a number of 32 referential links in the annotation that were either missing or not plausibly marked. Also, we detected too fine unit segmentations in 3 cases.

We detected 35 cases in which this happened, from the 77 exceptions in the vein-resolution. That is, 45.5% of the exceptions correctly indicated errors of configuration. They were repaired using relatively few correcting operation and usually a single modification in the structure led to the transformation of multiple exceptions in direct or indirect references: 27 basic operations (10 structural and 17 nuclear) were applied in the 35 cases, that actually corrected 20 cases (an average of 1.35 operation/case) and sufficed to also correct the rest of 15, or 42.86% from the total.

The remaining 42 exceptions did not correspond to structure errors, but were either pragmatical references, to entities known from outside the discourse, like "The White House", "The Senate" etc. (11 or 26.1%), or long-distance name references (9 or 21.4%). Also, 17 cases involve an *attribution* relation. This big number suggests, as (Ide and Cristea, 2000) remarked, that the *attribution* relation's nuclearity should be reconsidered, perhaps getting rid of it entirely and allowing for the inclusion of both constituents that it connects into a unique discourse unit. The other 5 unresolved cases concern the use of the *purpose* relation (3 of them) or, interestingly, anaphors realized as definite nouns renaming the antecedents (e.g. "the steelmaker" referring back to "Bethlehem Steel Corp.").

The results show that the exceptions are good indicators of wrongly build areas in discourse structure. They can also provide indications and suggestions on several analysis matters.

## 5. Conclusions

We have shown how VT-derived referential constraints apply to the discourse structure configuration and can be successfully used in better structuring discourse. Although the correction of an initial structure may not be complete, in the sense that it is still uncertain whether the result corresponds or not to the interpretation the author intends for the text, it is clear that the method proposed allows us to provide a better structure which can also accommodate the structural restrictions imposed by the references.

We proposed the basic local correcting operations on the tree-like structure of text, and a global correcting method that uses the successful reference annotations of texts in order to improve given structural analyses.

The method can be applied not only for the correction of human annotation to structure, but also to the analyses automatically derived, for instance on the basis of cue-words, in order to refine them, or during an incremental discourse parsing process, in order to guide or assist it.

We believe that the processes of anaphora resolution and discourse structure building are interdependent to such a degree that discourse analysis should definitely make use of both of them indivisibly, and combine their partial results to acquire the best analysis. In the same way that anaphora resolution can benefit from the discourse structure, already solved anaphora can be used in determining the new structure, which in turn contributes to the resolution of further anaphora.

The work of (Schauer and Hahn, 2001) uses the same idea of considering text cohesion to address the coherence problem of discourse. They have proposed an algorithm for the combined computation of co-references and

discourse structure, using the right frontier of the partially built tree to find both the target to connect a new unit to, and the antecedents of the anaphora in the new unit. Our approach is more specific with regard to the constraints imposed to the structure, given that the vein's definition elaborates the right-frontier principle (Webber, 1991). Differences appear both with respect to the target node in the partial tree the new unit is connected to, and consequently with respect to the resolution of new anaphora. We expect our approach to better account for the cases in which their algorithm implausibly predicts the target unit.

Further research involves the investigations on the way in which additional reference-related factors restrict the text structure: the reference type, the kind of anaphor, the evoking power of the anaphor, the distance in text between anaphor and antecedent.

## 6. References

- Cristea, D., 2000. An Incremental Discourse Parser Architecture. In D. Christodoulakis (Ed.) *Proceedings of the Second International Conference - Natural Language Processing - NLP 2000*, Patras, Greece Lecture Notes in Artificial Intelligence 1835, Springer.
- Cristea, D., and Webber, B.L., 1997. Expectations in Incremental Discourse Processing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid.
- Cristea, D., Ide, N., and Romary, L., 1998. Veins Theory: A Model of Global Discourse Cohesion and Coherence. In *Proceedings of COLING-ACL'98* (pp. 281–85).
- Cristea, D., Ide, N., Marcu, D., and Tablan, M.V., 2000. Discourse Structure and Co-Reference: An Empirical Study. In *Proceedings of the 18th International Conference on Computational Linguistics COLING'2000*, Saarbrueken.
- Fox, B., 1987. Discourse Structure and Anaphora. Written and Conversational English. *Cambridge Studies in Linguistics*, Cambridge University Press.
- Grosz, B.J., and Sidner, C., 1986. Attention, intentions, and the structure of discourse. In *Computational Linguistics*, 12(3):175–204.
- Hirschman, L., and Chinchor, N., 1997. MUC-7 Co-reference Task Definition.
- Ide, N., and Cristea, D., 2000. A Hierarchical Account of Referential Accessibility. *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, ACL'2000*, Hong Kong.
- Joshi, A., 1987. An introduction to Tree Adjoining Grammar. In A. Manaster-Rammer (ed.) *Mathematics of Language*.
- Kurohashi, S., and Nagao, M., 1997. Automatic Detection of Discourse Structure by Checking Surface Information in Sentences. In *Proceedings of Coling'97* (pp. 1123–1127).
- Mann, W.C., and Thompson, S.A., 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3):243–281.
- Marcu, D., 1997. The rhetorical parsing, summarization and generation of natural language texts. Ph. D. Thesis. Dept. of Computer Science, University of Toronto.
- Marcu, D., 1999. A formal and computational synthesis of Grosz and Sidner's and Mann and Thompson's theories.

- In *Workshop on Levels of Representation in Discourse*.  
Edinburgh.
- Marcu, D., 2000. The theory and practice of discourse parsing and summarization, The MIT Press, Cambridge, Massachusetts.
- Marcu, D., Amorrortu, E., and Romera, M., 1999. Experiments in Constructing a Corpus of Discourse Trees. In *ACL'99 Workshop on Standards and Tools for Discourse Tagging*, Maryland.
- Moser, M., and Moore, J.D., 1996. Toward a synthesis of two accounts of discourse structure. In *Computational Linguistics*, 22(3):409–419.
- O'Donnell, M., 1997. RST-Tool: An RST analysis tool. In *Proceedings of the 6th European Workshop on Natural Language Generation*, Duisburg, Germany.
- Schauer, H., and Hahn, U., 2001. Anaphoric Cues for Coherence Relations. In *Proceedings of RANLP'2001* (pp. 228–235).
- Serețan, V., 2000. Considerations regarding the reconciliation of discourse structure with referential chains (in Romanian). Master Thesis, University of Iași.
- Vonk, W., Hustinx, L., and Simons, H.G., 1992. The use of referential expressions in structuring discourse. In *Language and Cognitive Processes*, 7(3,4):301–333.
- Webber, B., 1991. Structure and ostension in the interpretation of discourse deixis. In *Natural Language and Cognitive Processes* 6(2):107–135.