The relationship between discourse structure and referentiality in Veins Theory

Dan Cristea
University "Alexandru Ioan Cuza" of Iaşi, Faculty of Computer Science
16, Berthelot St., 6600 – Iaşi, Romania
dcristea@infoiasi.ro

Abstract
The aim of this paper is to give an intuitive look onto that part of the Veins Theory – a theory of discourse structure and cohesion – that deals with the relationship between discourse structure and referentiality. It formalizes the central notion of vein used to identify domains of referential accessibility for anaphors in the discourse. One application of the theory is to make corrections to discourse structure based on firm referential links.

1 Introduction

For the past 25 years a lot of work has been done to understand what makes a text be a discourse (considered as a sequence of sentences), therefore why a discourse is coherent and what makes it cohesive. People have felt that the sentence and the discourse should have something in common, although a discourse is more than a sentence seen at a different granularity. The main concern of the studies dedicated to discourse has been the discourse structure and the relationship between discourse structure and referentiality.

Grosz and Sidner [10] in the Atentional State Theory (AST) propose a recursive segmental structure of the discourse, which resides in a tree-like representation (the result of considering two relations between discourse segments: dominance and satisfaction-precedence). The AST's account on the dynamic interpretation claims that a stack could accurately model the processes involved in discourse interpretation, including accessibility: the domain of reference for the referential expressions that belong to the state (segment) positioned on top of the stack is given by entities belonging to states below in the stack. However, the stack-based model can run into difficulties when dealing with sub-segments of a dominating segment which appear at the same level.

In Mann and Thompson's Rhetorical Structure Theory (RST) [13], the accent moves towards rhetorical performance: by what means a writer (speaker) could persuade a reader (listener) to accept his communication intentions. Perhaps as a by-product of the main objectives of RST, a lot of post-RST work has been dedicated to the study of discourse structure along the lines traced by the theory. Indeed it is very convenient to see discourse represented as a tree where terminal nodes are elementary discourse units, intermediate level nodes represent (rhetorical) relations between sub-spans of text, and coordination and subordination of compounds is similar to that of syntactic structures. Still, RST gives no clues on the relationship between structure and referentiality. Fortunately, significant similarities between AST and RST (the hierarchical nesting of discourse segments and the use of the stack as a processing mechanism in AST strongly resembles the dynamic processing model of a tree-like structure as the one in RST) were pointed out at least by Moser and Moore [16] and Marcu [14]. These similarities could make the glue of a combined account in which RST contributes with a more refined set of relations, while AST with a discipline on where to look for the antecedents of referring expressions. The union of these two theories gives a reasonable justification to
observations as those made by Fox [8] about pronominal pop-overs (long distance pronominal references to matching antecedents which are not the most linearly recent).

Yet, another theory should be added to the pot: Centering (CT) [9], [1], which gives a convincing explanation of what makes a discourse coherent. Nevertheless, with its account for references between adjacent units, CT is known to be applicable solely inside a segment as defined by AST. But a segment, also as defined by AST, has a recursive structure. If one does not penalize CT in its locality claim based on this observation, then it seems natural to think that CT could be applied to the global discourse in a kind of recursive way, on the recursive segmental structure.

By using the RST notion of nuclearity, Veins Theory (VT) [3] reveals a "hidden" structure in the discourse tree, called vein, that enables to determine the domain of referential accessibility (DRA) for each discourse unit. VT gives an integrated explanation of the common points of these three theories, while it also corrects some AST predictions relative to accessibility domains (the nucleus to nucleus references or references from nuclei to left satellites). It generalizes CT from local to global discourse, adds a view on summarization (consistent with Marcu [15]) and naturally reveals how focused summaries can be produced. VT’s account on the relationship between discourse structure and referentiality can be exploited in three ways:

- to constrain a simultaneous parsing and anaphora resolution process towards that interpretation that requires minimum inferential load in building the structure and in identifying the antecedents of referential expressions [5];
- to help in the interpretation of anaphora when discourse structure is known, by preferably chasing only antecedents contained in the DRA of the anaphor’s unit [7];
- to correct discourse structure when referential links are known (this will be discussed in section 5) [18].

In the following sections we synthesize the cohesion account of VT, in a rather intuitive way. Basic assumptions of VT are presented in section 2 and its main lines are revised in section 3. VT’s claim on discourse cohesion is explained in section 4. Section 5 exemplifies an application of VT for the correction of discourse structure while the last section makes a synthesis and compares VT with related work.

2 Basics

The fundamental intuition underlying the unified account on discourse structure and accessibility in VT is that the RST-specific distinction between nuclei and satellites constrains the range of referents to which anaphors can be resolved; in other words, the nucleus-satellite distinction, superimposed over a tree-like structure of discourse, induces for each anaphor a DRA. More precisely, for each anaphor $x$ in a discourse unit $u$, VT hypothesizes that $x$ can be resolved by examining discourse entities from a subset of the discourse units that precede $u$.

The discourse structure assumptions in VT are, to a great extent, the same as in RST:

- the basic units of a discourse are non-overlapping spans of text: usually a dot-to-dot sentence, but they may be sometimes reduced to a clause;
- discourse structures are represented as trees. Unlike RST, in VT, without any loss of generality, the trees are considered binary (each relation node has exactly two descendents);
- terminal nodes of the tree represent discourse units and non-terminal nodes represent discourse relations. Unlike RST, VT is not concerned with the type of relations among textual spans, but considers only the topological structure of the discourse;
- a polarity, established among the daughters of a relation, identifies at least one node, the nucleus, considered essential for the writer’s purpose; non-nuclear nodes, which include spans
of text that increase understanding but are not essential to the writer’s purpose, are called satellites.

Although not central in VT, the same principles as in RST (not explicitly formulated in the theory) apply:

- the principle of sequentiality [2]: the sequence of nodes on the terminal frontier of the tree corresponds to the sequence of discourse units in the original text;
- the principle of compositionality [15]: a relation that holds between two spans also holds between the most salient units of the spans involved in that relation.

VT adopts an incremental processing approach to referentiality, according to which a referential expression refers an entity already introduced in the discourse, therefore belonging to the span of text already read (or heard) at the moment that particular referential expression is encountered (to the left, in left-to-right writing languages, to the right in right-to-left languages). Consistent with this view, syntactic cathaphors of the type X does something. When he realises something, X (= he) does something else, where he does not represent the first mention of the discourse entity, actually have an antecedent previously present in the text and they refer it. Cathaphors which do not have such an antecedent in the preceding text, as in:

“From the corner of the divan of Persian saddle/bags on which he was lying, smoking, as was his custom, innumerable cigarettes, Lord Henry Wotton could just catch the gleam of the honey-sweet and honey-coloured blossoms of a laburnum…” (O. Wilde – The Picture of Dorian Gray, chapter 1, par. 1)

introduce new entities, as poor in morpho-semantic features as the referential expression itself (in the example above, after the first two pronouns have been processed: \([\text{sem=person, gender=masculine, number=singular}]\)). In this case, the following definite noun phrase or proper noun refers the new entity as an antecedent, while also refining the description or complementing it with new information (in the example, after Lord Henry Wotton has been processed: \([\text{sem=lord, name=”Henri Wotton”, gender=masculine, number=singular}]\)).

The notion of vein was born by synthesizing observations on how references align within the representation of a discourse as a tree. Considering the hierarchical organisation given by the tree structure and the principle of compositionality, which allow long-distance sibling relations between discourse units, these observations could be stated as follows (to simplify the wording, we will say that “a unit A refers a unit B” if we mean “a referential expression belonging to the unit A refers a discourse entity introduced or referred by from unit B”):

- in most cases, right satellites or nuclei refer their left nuclear siblings;
- in less cases, a right nucleus refers a left satellite;
- a nucleus blocks the reference from a right satellite to a left satellite.

3 Definitions

Vein expressions defined over a discourse tree are sub-sequences of the sequence of units making up the discourse. To define vein expressions, the following notations are used:

- each terminal node (leaf node, discourse unit) has an attached label;
- \(\text{mark}(x)\) is a function that takes a string of symbols \(x\) and returns each symbol in \(x\) marked in some way (e.g., with brackets);
- \(\text{simpl}(x)\) is a function that eliminates all marked symbols from its argument, if they

---

1 In this paper, the expressions “to the left” or “left-to-right” must be considered with respect to the first type of languages and should be reversed if applied to the other type of languages.
exist, (e.g. $\text{simp}(a(bc)d(e)) = ad$);

- $\text{seq}(x, y)$ is a sequencing function that takes as input two non-intersecting strings of terminal node labels, $x$ and $y$, and returns that permutation of $x$ concatenated with $y$ that is given by the left-to-right reading of the sequence of labels in $x$ and $y$ on the terminal frontier of the tree. The function maintains the marks, if they exist.

**Figure 1**: Representation of the RST analysis for a text.

**Figure 2**: The VT-like representation of the RST tree in Figure 1 (nuclear nodes are underlined)

---

2 The figure was drawn using the RST-Tool [17].
3.1 Heads and veins

Head expressions
1. The head of a terminal node is its label.
2. The head of a non-terminal node is the concatenation of the heads of its nuclear daughters.

The notion of head in VT is equivalent to that of Marcu’s promotion set [15]. Note that the recursive definition of head induces a bottom-up computation over the tree structure. For example, for the text in Figure 1, with the displayed rhetorical structure\(^3\) sketched again in Figure 2 with the graphical notation common in VT, the head expression of the span [4,5] is unit 4 because the head of its nuclear daughter – the elementary unit 4 – is 4. However, the head of the span [3,5] is the list <3,4> because both immediate daughters are nuclei of a join multinuclear relation (see Figure 4).

Vein expressions
1. The vein expression of the root is its head.
2. For each nuclear node whose parent node has the vein $v$, the vein expression is:
   - if the node has a left non-nuclear sibling with head $h$, then $\text{seq}(\text{mark}(h), v)$ (see Figure 3a);
   - otherwise, $v$ (see Figure 3b).
3. For each non-nuclear node of head $h$ whose parent node has vein $v$, the vein expression is:
   - if the node is the left daughter of its parent, then $\text{seq}(h, v)$ (see Figure 3c);
   - otherwise, $\text{seq}(h, \text{simpl}(v))$ (see Figure 3d).

The vein expression of a node intends to reflect the sequence of units that are significant to understand the span of text covered by the node, in the context of the whole discourse. Each vein expression contains the significant units within the span together with other surrounding units (the “visible” preceding units and the most important following units).

Figure 3: Computing veins expressions. The current node is dark, nuclei are underlined.

\(^3\) The rhetorical structure is represented using the conventions proposed by Mann and Thompson [13].
Once each node of the tree is marked for the head expression, vein expressions are computed top-down for each node in the tree.

3.2 Anaphoric references

In agreement with many scholars that have studied anaphora resolution (see for instance [11]), VT sees anaphoric relations as having a semantic nature. When the dynamic interpretation of anaphora is stressed, as shown in [6], [7], arguments can be found to view the anaphora resolution process as involving three layers: the text layer – populated with referential expressions (REs) –, the intermediate layer – where feature structures are projected from the text –, and the deep semantic layer – where descriptions of discourse entities (DEs, also called centers in this paper) are placed (see Figure 5 for a co-referring example).

If we don’t consider phenomena related to the interpretation of anaphors in its dynamics and restrict only to the final representation of anaphoric links, then the intermediate layer can be
hidden. At this level of abstraction, co-referential relations induce equivalence classes over the set of REs in a text. These classes are given by the discourse entities on the semantic layer (co-referential expressions are the ones that point to the same DE).

When hierarchical adjacency is considered, an anaphor may be resolved to a referent that is not the closest in a linear interpretation of a text. However, because co-referential expressions are organized in equivalence classes, it is sufficient that an anaphor is resolved to some member of the set. This is consistent with the distinction between "direct" and "indirect" references discussed in [3] and [5] and reproduced again below:

**Direct references.** If $A$ and $B$ are units in this textual order, $A$ belongs to the DRA of $B$ and $a \in A$ is linearly the most recent (to $B$) RE that evokes the same center as $b \in B$, we say that $b$ **directly co-refers** the center evoked by $a$ (see Figure 6).

If $A$ and $B$ are units in this order, $A$ belongs in the DRA of $B$ and $a \in A$ is linearly the most recent RE (to $B$) that evokes a role of $b \in B$, we say that $b$ **functionally directly refers** the center evoked by $a$. To put into evidence the DRAs of units in the following graphical representations, the units and the veins they belong to are drawn above REs. Also, to simplify the notation, only the first and the third layers are shown.

**Figure 6: Direct reference:** the linearly most recent RE that evokes the same center as the current RE is shown by a dashed arrow and the evoking relation by a full arrow

**Indirect references.** If $A$, $B$ and $C$ are units in this order, $b \in B$ is linearly the most recent (to $C$) RE that evokes the same center as $c \in C$, $B$ is not on the DRA of $C$, $A$ is linearly the most recent (to $B$) unit that is both on the DRA of $B$ and of $C$, and it contains a RE $a \in A$ such that $b \in B$ evokes the same center as $a$, we say that $c$ **indirectly co-refers** the center evoked by $a$ (see Figure 7). A similar definition applies for indirect functional references.

**Figure 7: Indirect reference**

In other words, an indirect reference occurs when: the DRA expression of the anaphor’s unit intersects that of the most recent antecedent, there is a unit on this common segment that evokes the same center as the one referred by the anaphor, and this unit is not linearly the most recent to the anaphor. That is, an indirect reference occurs when the chain of units of backward looking references intersects the DRA of anaphor’s unit in a unit that is not the most recent to the anaphor’s unit.
All references that do not obey the direct or indirect references are **inferential references**. This happens when there is no intersection between the backward looking chain of the units that contain the referential expressions and the DRA of the anaphor's unit.

A particular category of inference references could be called **pragmatic references** (a kind of pseudo references). These are expressions that can be understood without pointing to an entity already introduced in the discourse, although an identically evoked entity has already been introduced. If one says that he has woken up this morning at the same time with the sun and, after a few sentences, he says that no cloud covers the sun, it is questionable whether in the latter sentence he refers back to his first mention of the celestial body. One can understand the last sentence without making the connection with the first one. We all know about sun and anytime we are told about it we know what the concept is because it is outside the text, it is the same to all of us, it is pragmatic knowledge. In the same category: the Senate, the White House, well-known names of cities, countries, personalities, etc. But if we tell a story about the sun which is passing through different events, from its raising to its dawning, then we have references. Although everybody knows what the sun is, one can understand the story only linking the various sequential mentions of the sun to one single situational entity. Because the different hypostases of the sun in the story participate in connected events, it is important for someone to have the same mental representation about all the mentioned hypostases, and not only to know them as having identical functionality.

The difference between circumstances as the ones mentioned could be evidenced by considering the references in connection with the vein structure. It is most plausible that references in the first example could appear as inferential references, while references in the second example could be recovered as direct or indirect references.

### 4 VT’s claim on cohesion

**Cohesion conjecture**: References from a given unit are possible only in its domain of accessibility.

The domain of referential accessibility (DRA) of a unit $u$ is given by the units in the vein of $u$ that precede $u$. For example, for the RST tree and text in Figures 1 and 2, the vein expression of unit 7, which is given by the sequence of units <1 2 (6) 7>, suggests that anaphors of unit 7 should be resolved only to referential expressions in units 1, 2, 6 and 7. Indeed, the reference Mr. Wright in 7 is satisfied by the antecedent [Mr. Wright] introduced in unit 2. Since unit 6 is a satellite to unit 7, it is considered to be “blocked” to referential links from unit 8, another satellite of 7, but to its right. Indeed no referential expression exists in 8 to be resolved by a discourse entity of 6. Both who and his have as antecedents the same [Mr. Wright], who can be found in 7. Figure 4 shows the heads and veins of all nodes.

Empirical evidence on VT claims was reported in [3], [4] and [12]. In particular, these studies reveal the following:
- in most cases the references are direct;
- in less cases the references are indirect;
- in even less cases the references are pragmatic;
- inferential references which are not pragmatic signal a hard-to-make inference or a failed discourse.
5 Correcting discourse structure

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 be also considered when structuring discourse.

When there is strong evidence for reference links (as, for instance, given by proper nouns to homonymous proper nouns), that do not obey the cohesion conjecture, questions arise on the correctness of the discourse structure itself. In [18] a method is proposed for systematic detection and correction of structural errors which are likely to occur during the structural annotation. Slight modifications are made to the current structure, in the areas indicated by the exceptions in the vein resolution, in order to allow a possible antecedent of the anaphor’s unit to appear on the DRA of that unit. To exemplify, let’s notice the case of such an exception on the discourse structure of Figures 1, 2 and 3: the only antecedent for *Merril Lynch* in unit 6 is the DE *[Merril Lynch]* introduced in unit 3, which is not on the vein of unit 6 (1 2 6 7, as shown in Figure 4). The correction results in the structure drawn in Figure 8. A thorough analysis reveals, indeed, that it is not the case that the whole span (2-5) is 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 subsupan 3-5 ("Mr. Wright resigned as president of Marrill 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 [15]: the relation also holds between units 3 and 7, while, in the initial structure, this criterion is not obeyed as the relation doesn't hold between units 2 and 7.

![Figure 8: The correct structure proposed for the original structure of Figure 1](image)

The investigation of this kind of errors on a corpus annotated for discourse structure showed (see [18]) that most of the resolution exceptions corresponded indeed to mistakes in the structure construction (Marcu's Compositionality Criterion was not obeyed). The results proved that the coherence conjecture exceptions are good indicators of wrongly build areas in the discourse structure.
6 Discussion

The fundamental assumption underlying VT is that an inter-unit reference is possible only if the two units are in a structural relation with one another, even if they are distant from one another in the text stream. Furthermore, inter-unit references are rather to nuclei than to satellites, reflecting the intuition that nuclei assert the writer’s main ideas and provide the main “threads” of the discourse [13]. This is shown in the computation of veins over (binary) left polarised discourse trees, where any reference from a nuclear unit must be to entities contained in linguistic expressions from the previous nuclei (although perhaps not any nucleus). On the other hand, satellites depend on their nuclei for their meaning and hence may refer to entities introduced within them. The definition of veins formalizes these relationships. Given the mapping of Grosz and Sidner's [10] stack-based model of discourse structure on RST structure trees outlined by Moser and Moore [16] and Marcu [14], the domains of referentiality defined for left-polarized trees using VT are consistent with those defined using the stack-based model.

However, in cases where the discourse structure is not left-polarized, VT provides a more natural account of referential accessibility than the stack-based model. In non left-polarized trees, at least one satellite precedes its nucleus in the discourse and is therefore its left sibling in the binary discourse tree. 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 entities in A (its left satellite), but the subsequent right satellite, C, cannot refer A due to the interposition of the nuclear unit B. In stack-based approaches to referentiality, such configurations raise problems: as B dominates A it must appear below it on the stack, even though it is processed after A. Even if the processing difficulties are overcome, this situation leads to the postulation of “right” references of cataphora included in satellites that precede their nuclei, which is counter-intuitive.

Inferential references seem to minimize the importance of the domain of referential accessibility, as defined in VT, because references can now “escape” from the domain. Does the domain of accessibility have any significance anymore? Is it an artificial invention or is it defended by a natural characteristic of the manner humans process texts? We claim that there are two significantly distinct types of anaphora resolution processes: evocative (or associative) and post-evocative (or inferential).

The evocative resolution processes are due to direct and indirect references. They are based on associations, which are processes of pattern-matching on feature structures decorated with morpho-semantic attributes. They are performed between a feature structure projected by the anaphor RE and a center that already exists in the DRA of the unit the anaphor belongs to. As such, they are performed between layer two and layer three feature structures. These are fast processes, direct ones being faster and more frequent than indirect ones.

On the other hand, the post-evocative processes are inferential processes that are developed in memory, based on the knowledge accumulated by the preceding discourse, or based on the cultural knowledge the subject owns. We believe these inferences swing the semantic space in an order that is also dictated by discourse structure. Eventually, the target entity can be found based on a pattern-matching process between the projected structure of the anaphor and the center of the antecedent. They are slow (compel to more inference load), require more powerful referencing means (like proper nouns), and are less frequent processes.

A side-effect of corpus research motivated by evaluation of VT claims was the notice that there is a strong relationship between the different kind of referential expressions and their distribution with respect to the three kinds of references put in evidence by VT. It was revealed an alignment between the evoking power and the percentage of different types of referential expressions that did not corresponded to a vein reference. So, the following four types of referring expressions: pragmatic, proper nouns, common nouns and pronouns revealed to have
descending frequencies. Pronouns are very fragile evoking means, and, as such, a message emitter employs them when he is certain that the structure of the discourse allows for easy recuperation of the antecedent in the message receiver's memory. The alignment of the evoking power of referential expressions with the percentage of exceptions of references outside the DRAs shows that the predictions made by VT relative to the domains of referential accessibility are correct. Practically, except for the cases when the pronoun can be understood without an antecedent, it becomes impossible to use a pronoun as an anaphor to refer an antecedent that is outside the DRA.

On the contrary, the almost equal distribution of anaphor types in the exceptions reported by GST, irrespective of their evoking power, shows the weak predicting capability of GST with respect to determining the DRAs.

VT recognises the special status of the left satellite for discourse structure, it predicts that references not contained in the DRA of the corresponding unit are more difficult to process – computationally and cognitively – and it accords this intuition with the different types of referring expressions.

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 structure, which in turn contributes to the resolution of further anaphora.

**Bibliography**


