Motivations And Implications Of Veins Theory.

Cohesion

Dan Cristea

Faculty of Computer Science of the “Alexandru Ioan Cuza” University of Iasi, Romania
Institute of Theoretical Computer Science, Iasi branch of the Romanian Academy
dcristea@info.uaic.ro

Abstract. The paper deals with the cohesion part of a model of global discourse interpretation, usually known as Veins Theory (VT). By taking the notion of nuclearity (though ignoring relations), from the Rhetorical Structure Theory, VT computes strings of discourse units, called veins, from which domains of accessibility can be determined for each discourse unit. VT’s constructs best fit with an incremental view on discourse processing. Linguistic observations that lead to the elaboration of the theory are presented. Cognitive aspects like short-term memory and on-line summarization are explained in terms of VT’s constructs. Complementary remarks are made on anaphora and its resolution in relation with the interpretation of discourse.

1 Introduction

A discourse is different from a text, because a discourse is a text in the progress of reading or hearing in a human brain. So, a discourse exists only as a process and, as such, it has a dynamic nature. When reading comes to an end, the discourse also finishes and only a representation of it remains in the reader's memory.

Some of the main concerns of the studies dedicated to discourse have been on proposing a representation that best describes its structure and on understanding the relationship between structure and referentiality. In Atentional State Theory (AST) (Grosz and Sidner, 1986) the discourse is seen as having a recursive segmental structure residing in a tree-like representation, while the dynamic interpretation uses a stack model in which the references are allowed to occur from the top state elements downwards. The Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) gives only a static representation while ignoring referentiality. Centering Theory (CT) (Grosz et al, 1995; Brennan et al, 1987) uses the notion of segment from AST to propose a local theory of discourse coherence. More recent studies look for discourse markers, cues and particles that can bring indications on the discourse structure (Marcu, 2000; Popescu-Belis and Zufferey, 2007), build implementations able to detect the discourse structure (Cristea et al., 2005), do utterance (Geertzen, 2007) and topic segmentation (Gruenstein et al., 2005), use ontologies in discourse understanding (Niekrasz, 2005), etc.

The main preoccupations in discourse structure are related to understanding (coherence), referentiality (mainly, cohesion) and summarisation (abstraction). Issues involving understanding discuss: how should the meaning of a text be expressed, how is the meaning of a whole made up by combining the meanings of the components, in what way does pragmatics influence the understanding, what is the contribution of the context in decoding the meaning and why is that certain texts are easier to interpret than others. Referentiality means: where in the text should the antecedent of a referential expression be searched, how could coreferential anaphoric relations, metonymy and bridging anaphora be resolved, why is that certain mentioned entities cannot be referred at certain positions by pronouns, but only by very powerful evoking means. And summarisation deals with finding a shorter reformulation of a text, sometimes of only a part of a larger text, or form a short story concerning what the text says about a certain discourse entity.

In this paper we will review the cohesion part of a model of global discourse interpretation, usually known as Veins Theory (VT), while also noticing some new facts about it. VT is not a semantic theory of discourse, therefore it is not concerned with discussing the way the meaning can be abstracted away from text.

By taking the notion of nuclearity (though ignoring relations) from RST, VT (Cristea et al, 1998) reveals a "hidden" structure in the discourse tree, called vein, which gives the minimal span necessary to understand one particular unit in the context of the whole discourse. The vein enables to determine a domain of evocative accessibility (dea) for each discourse unit, as a sub-span of the whole discourse which is opened for immediate reference. In particular, the dea of a particular unit gives the space...
where the weak-power evoking anaphors belonging to that unit can find an antecedent. As such, VT allows for an integrated explanation of the common points of AST, RST and CT, while also correcting some AST predictions relative to accessibility domains (mainly references from nuclei to left satellites).

In the following section we present linguistic observations that lead to the formulation of VT. The basic definitions are revised in section 3. VT’s claim on discourse cohesion is presented in section 4. The last section gives a synthesis of the theory, provides a cognitive argumentation, briefly reviews applications based on the findings of the theory and shows some possible future developments.

2 The Intuitions Underlying VT

The notion of vein appeared by synthesizing observations on how references align within the representation of a discourse as a tree. It is grounded in the hierarchical organization provided by the tree structure and the principle of compositionality in RST, according to which, a relation that holds between two spans also holds between the most salient units of those spans (Marcu, 2000). In particular, this principle allows for long-distance relations between sibling discourse units.

In the notes accompanying the examples that follow we will say that “a unit A refers to a unit B” when we mean “a referential expression (re) belonging to the unit A refers to a discourse element (de) introduced or referred in/from unit B”.

In the examples of this section, as well as further down in the paper, we will mark with indexed \(u\) – units, and with \(R\) or indexed \(R\) – rhetorical relations. An upper \(n\) or \(s\) at the shoulder of an expression representing a span indicates that the corresponding text span is a nucleus, or a satellite, respectively. In italics, we will reproduce parts of discourses/dialogues or only referential expressions. Discourse entities will be noted by relevant names, in bold, placed between square brackets (example, [John]). The names of relations in our commentaries of the examples are taken from RST.

a). Right satellites or nuclei can refer to their left nuclear siblings: in combinations \(u^n R u^s\), or \(u^n R u^s, u_s\) can refer to \(u^n\);

Ex. 1:

1. John left home without an umbrella
2. although he watched the TV morning forecast announcing rain.

The pronoun he in unit 2, a satellite of unit 1, refers to the entity [John], introduced by the referential expression John in the first unit.

b). A right nucleus can refer to a left satellite: in combinations \(u^n R u^s, u_s\) can refer to \(u^n\), as in:

Ex. 2:

1. Although John watched the TV morning forecast announcing rain,
2. he left home without an umbrella.

where he in 2, a nucleus, refers to [John] introduced in 1, a left satellite of 2.

c). A right satellite of a nucleus \(u\) is not accessible from another, more distant, right sibling of \(u\), nuclear or satellite: in combinations \((u^n R_1 u^n)^R_2 u^n\) or \((u^n R_1 u^n)^R_2 u^n\), \(u^n\) can refer to \(u^n\) but not to \(u^n\).

Ex. 3:

1. John told Mary that he loves her.
2. He has never been married
3. and lived until 40 with his mother.
4. She, on the contrary, has been married twice.

Sequence 2-3-4 ELABORATEs on 1. Sequence 2-3 is in a relation of CONTRAST (a paratactic relation) with respect to 4, while unit 3 ELABORATEs on 2. The structure is therefore: \(u^n R_1 ((u^n R_2 u^n)^R_3 u^n)'\). For most readers, she in unit 4 must be [Mary], and not [John’s mother], although [John’s mother] is the most recent entity from the position of unit 4 in agreement in gender and number with the pronoun she. The reason why the reader prefers Mary instead of the mother is because s/he recognizes unit 4 as being in a CONTRAST relation with unit 2 (evidenced by on the contrary), which makes the two units to be perceived as adjacent, and having the same status with respect to a common nucleus, unit 1. Their proximity however is not linear but hierarchical, on the structure. This
makes unit 3 to be closed for reference from unit 4, and the pronoun she in 4 will find its antecedent in the common upper nucleus – unit 1.

The same happens in a structure like: \((u_1^" R_j u_2^" R_z u_j^\prime\prime)\), where \(u_j^\prime\prime\) cannot refer \(u_j^\prime\).

d). A nucleus blocks the reference from a right to a left satellite: in combinations \((u_j^\prime R_j u_2^\prime R_z u_j^\prime\prime)\), \(u_j^\prime\) can refer to \(u_2\) but not to \(u_j\).

Ex. 4:

1. With one year before finishing his mandate as president of the company,
2. Mr. W. Ross has begun to bring about its bankruptcy.
3. There were rumours that he has obtained it by fraud.

In this example, the pronoun its in unit 2 refers to [the company] (only a company can be bankrupted, not a mandate). Then, in unit 3 most readers would see the antecedent of the pronoun it as being [the company’s bankruptcy] and not [Mr. Ross’ mandate as president]. As such, units 1 and 3 are both satellites of unit 2: 1 is in a CIRCUMSTANCE relation with respect to 2, while 3 is intended to give a BACKGROUND for 2. An interpretation in which it in 3 would refer back to [Mr. Ross’ mandate as president] (an antecedent found in unit 1), although semantically compatible, is difficult to infer. To acquire, however, this last interpretation (in which rumors say that Mr Ross has obtained the mandate as president, and not about his bankruptcy patchworks), the discourse can be modified in two ways:

One way is to eliminate the reference:

Ex. 5:

1. With one year before finishing his mandate as president of the company
2. Mr. W. Ross has begun to bring about its bankruptcy.
3. There were rumours that he has obtained it by fraud.

The result is a comprehensible discourse, although the interpretation of unit 3 as a satellite of 1 is not easy and it should be taken as a flashback.

The other way is to reverse the first two sentences:

Ex. 6:

1. Mr. W. Ross has begun to bring about the bankruptcy of his company
2. with one year before finishing his mandate as president.
3. There were rumours that he has obtained it by fraud.

The discourse in Ex. 6 has a different structure then the one in Ex.4: unit 2 is a satellite of 1, and unit 3 is a satellite of 2. The reference its=[Mr. Ross’ mandate as president] is the only one that can be recuperated without difficulty, making 3 to reproduce a gossip occasioned by an element introduced in unit 2.

It is perhaps not realistic to put the failing of it of Ex. 4 - unit 3 to access an antecedent introduced in unit 1 on structural obstacles, as incurred by an interposing nucleus, more than on its longer linear distance to that antecedent, as compared to the successful variant. It is also difficult to state what comes first: structure identification or the resolution of references. But it is clear that both these two aspects correlate to obtain an optimum mental interpretation and that in the acquired result of this tendency towards the optimum certain rules can be identified. Once a structure is clearly highlighted by discourse markers, the references do observe certain restrictions. Vice-versa, if no explicit discourse markers are used, the disambiguation of anaphoric references can help in the recuperation of the structure. If contradictions occur, then the discourse is harder to process, it can give rise to interpretation traps and garden paths, or even it becomes incomprehensible.

Anaphoric phenomena involve mentions of entities that should be put in correspondence during the interpretation with previous or subsequent similar mentions of the same discourse entities. Although anaphora and cataphora, the two types of anaphoric relations, are usually considered to be characterised by opposing reference directions between the two terms, the inherently linear unfolding of the discourse forces the search space, during anaphoric interpretation, to always include discourse structures already stabilised, covering that part of the text which has already been read/heard and interpreted. Considering spoken discourse this means that an anaphor should look for an antecedent in the already uttered discourse, and considering a text which belongs to a left-to-right writing language it means that an antecedent should be searched for while reading strictly to the left of the anaphor. If resolution of direct anaphora is comfortable with this restriction, cataphora resolution, on the contrary,
apparently contradicts it. Cristea & Dima (2001) argue in favour of a left-only searching direction in all cases of anaphoric resolution, therefore in the case of both anaphora and cataphora. Their argumentation is based on the assumption that discourse entities are firstly introduced (proposed) and only then mentioned (evoked), in discourse. It is clear, however, that entities can be introduced with different degrees of detail. If the entity is introduced by semantic features (being a person, sex, profession, name), then it can be referred to by a pronoun. This is the case of anaphora. But the cases when the first mention is realised by a pronoun are also not rare (see Ex. 7).

Ex. 7:

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)

The pronoun initiates in the mind of the reader/hearer a rather shallow representation, announcing the entity vaguely, by only few semantic features (in the case of Ex. 7 – one masculine person). This initial scarce set of features will be later complemented with others (profession, social rank, a name, etc.). But this late disclosure constitutes a mere addition of data that complements the set of features characterizing an entity which already populates the mental space of the reader.

Ex. 8 is an example from another register, the dialogues.

Ex. 8 (after Cristea, 2005):

1. A: So, you didn’t know that I finished with Michael?!
2. A: It happened last month after we came back from Mexico.
3. B: Oh, I’m sorry. Do you have already someone else?
4. A: Negative! I need a period of loneliness.
5. B: you cannot resist long like this. I know you...
6. B: so, have you seen the pyramids there?

In this dialogue, units 1 and 2 belong to the same theme (thread), as signalled by the anaphoric pronoun it in the second utterance, anchored in the first utterance: it refers [the separation of A from Michael]. Then the turns 3-5 develop on this same theme, refining it further, while 6 refers back to A’s trip to Mexico, making evident the initiation of a separate thread. Now consider that instead of unit 6 the dialogue would proceed as follows:

Ex. 8’

... 6’. B: how often have you been there?

The new dialogue is much more difficult to process or even perceived as failed. The cause of this is that the evoking power of the pronoun there is much less powerful (called weak referential means in (Gundel et al., 1993)) than the mention of the pyramids. As theories of right frontier explain (Webber, 1991; Cristea&Webber, 1997), unit 2 is not visible from unit 6. This is true in both cases. Still, the dialogue of Ex. 8 is acceptable while the one of Ex. 8’ is not. The easiness to access unit 2 in Ex. 8 as compared to the difficulty to access it in Ex. 8’ should be put on the use of strong evoking means that accompany there in the first case and their absence in the second.

3 VT’s Basics

The fundamental intuition underlying the unified account on discourse structure and accessibility in VT is that an inter-unit reference is possible only if the units of the anaphor and antecedent are in a structural relation one with respect to the other. 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 a dea for each anaphor. 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. If the antecedent of x belongs to a unit that resides beyond the dea of u, then the link anaphor-antecedent is
found with difficulty or, in order to realize it, strong referential means should be surfaced (as, for instance, proper names).

The discourse structure assumptions in VT are, to a great extent, the same as in RST: a) the basic units of a discourse are non-overlapping spans of text, usually a clause of a sentence (expressing an event, or a situation); b) discourse structures are represented as trees. Unlike RST, in VT, without any loss of generality, the trees are considered binary; a similar representation is used by Marcu (2000); c) terminal nodes of the tree represent elementary discourse units (edus) and non-terminal nodes represent discourse relations. Unlike RST, VT is not concerned with the type of relations, but considers only the topological structure of the discourse; d) a polarity, established among the daughters of a relation, identifies at least one node as being nuclear, 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. The root of the discourse tree, by convention, is always satellite.

By keeping the polarity and discarding relation names, VT abstracts away from the typology of rhetorical relations. In the literature, there are disputes on the number of rhetorical relations needed to completely represent the discourse, for instance (Mann&Thompson, 1988; Knott, 1996). Meanwhile, it seems that the approaches that try to exploit a complete model of the rhetorical structure are few. Our model shows that important clues can be drawn solely on the basis of the topology of the tree structure and the binary labelling of its nodes. These trees will be called in the following VT-trees.

We will define for each node of a VT-tree two expressions, called head expression and vein expression respectively. Having a discourse tree structure, first, head expressions of all nodes are computed bottom-up, then the vein expressions of all nodes can be computed top-down, and, finally, based on the vein expressions, the domains of evocative accessibility of terminal nodes can be computed.

To define vein expressions and deas, the following notations will be used:
- each terminal (leaf) node (elementary discourse unit, edu) has an attached symbolic label (apart from the satellite-nucleus labelling), which, by itself, makes explicit a relation of total ordering among terminal nodes (for instance, integers from 0 to $N-1$, $N$ being the length of the discourse). As such, the whole discourse can be seen as the ordered maximal sequence of these symbolic labels;
- $. (the dot)$ is the concatenation operator: if $\alpha$ and $\beta$ are two sequences of symbols, then $\alpha \cdot \beta$ is the string containing the sequence in $\alpha$ followed by the sequence in $\beta$;
- $\text{mark}(\alpha)$ is a function that takes a string of edu symbols $\alpha$ and returns each symbol in $\alpha$ marked within brackets;
- $\text{unmark}(\alpha)$ is the reverse function of $\text{mark}(\cdot)$. It removes the marking attached to the symbols in the expression $\alpha$. (e.g. $\text{unmark}(\alpha \cdot \text{mark}(\beta) \cdot \gamma) = \alpha \cdot \beta \cdot \gamma$);
- $\text{simp}(x)$ is a function that eliminates all marked symbols from its argument, if they exist, e.g. $\text{simp}(\text{mark}(\alpha \cdot \beta)) = \emptyset$, the empty string, and $\text{simp}(\alpha \cdot \text{mark}(\beta) \cdot \gamma) = \alpha \cdot \beta \cdot \gamma$;
- $\text{seq}(\alpha, \beta)$ is a sequencing function that takes as input two non-intersecting strings of terminal node labels, $\alpha$ and $\beta$, and returns the ordered sequence of $\alpha$ concatenated with $\beta$. The function keeps unchanged the markings, if they exist, and $\text{seq}(\emptyset, \alpha) = \alpha$, $\text{seq}(\alpha, \text{seq}(\beta)) = \text{seq}(\text{seq}(\alpha), \beta) = \text{seq}(\text{seq}(\alpha), \text{seq}(\beta)) = \text{seq}(\alpha, \beta)$;
- we will note with $H$ and $V$, the head and vein expressions, respectively; the node $n$ they belong to will appear either as an index or in parenthesis;
- $\text{pref}(u, \alpha)$ retains the prefix of the expression $\alpha$ up to and including the symbol $u$.

VT computes two expressions that are attached to all nodes of a discourse structure. Both head and vein expressions are sub-sequences of the maximal sequence of units making up the discourse. The notion of head expression (simply head) in VT is equivalent to that of Marcus’s promotion set (2000). The intention in the head expression of a node of a discourse tree is to capture the sequence of the most prominent units in the span of text covered by the node. It is an ordered sequence of unit labels as follows:

1.1. The head of a terminal node is its label.
1.2. The head of a non-terminal node is the linear concatenation of the heads of its nuclear daughters.

Note that the recursive definition of head induces a bottom-up computation over a VT-tree. Indeed, this computation starts with the terminal nodes and continues, up the tree, until the root is reached and its head expression can be computed. In the following, the whole text is called total context.
The vein expression of a node is intended to give the sequence of edus which are significant for summarizing, in the total context, the span of text covered by the node. In the vein expression of any node in the discourse structure, there are included edus belonging to the span covered by the node, possibly together with edus outside the span. By synthesis (or summary) of a text span we understand a (possibly) shorter text, which can still render the original idea of the text. Irrespective whether it is realized by paraphrasing or by concatenating sub-sequences of the original text (Manni, 2001), any summary should be comprehensible by itself (among other things, this means that it should contain all elements that allow the resolution of anaphors). When the span to be summarized is extracted from a larger span, in order for the summary to be comprehensible, it should contain also elements from outside the span, which belong therefore to the context. We have, in this case, the summary of a text span, in the context of a larger span. Let’s note also that, in many respects, “summarizing” is equivalent to “understanding” because what we are usually left after the reading of a text is a synthesis of it.

In Fig. 1, inner nodes are depicted with rectangles, terminal nodes with circles, and the nodes to which the definition currently applies are depicted in grey. The last category is simultaneously drawn with a rectangle and a circle in order to suggest that they can be either inner nodes or terminal nodes.

Once each node of the tree has received a marking for the head expression, vein expressions can be computed top-down, starting in the root:

(2.1) The vein expression of the root is its head expression.

The vein expression of the root node, conforming to the intention associated to the vein expression of a node, stated earlier, should put in evidence those edus which are necessary to understand/summarize the span covered by the node (in this case – the whole text), in the total context. But, since the covered text span in this case is the whole text, the understanding/summarization of the whole text in the (trivial) total context is provided by the most significant units of the whole text, therefore the very head expression of the root node.

(2.2) For each nuclear node whose parent node has a vein \( v \):

(a) if the node does not have a left non-nuclear sibling, then its vein expression is \( v \) (see Fig. 1a);

(b) otherwise, if the left non-nuclear sibling has the head \( h \), then the vein expression of the nuclear node is \( \text{seq}(	ext{mark}(h), v) \) (see Fig. 1b).

The definitions say that in order to understand/summarize, in the total context, a nuclear span, a right satellite sibling can be ignored, while a left satellite is significant. When positioned at the right of a nuclear unit, a satellite can be ignored, since the same units are necessary to understand/summarize, in the total context, the nuclear span plus the satellite span, or only the nuclear span. When positioned at the left, a satellite helps to understand/summarize its right nucleus, but should be ignored for any other right satellite of this nucleus (case commented in Ex. 4). The marking function \( \text{mark} \) signals the contribution of this left satellite, in order that a subsequent removal is operated in the vein expression of a right satellite (see 2.3b below). On the contrary, twin nuclei cannot be understood/summarized one without the other, meaning that the same units are significant to understand/summarize each one of them as their union span.

(2.3) For each non-nuclear node of head \( h \) whose parent node has a vein \( v \):

(a) if the node is the left daughter of its parent, then its vein expression is \( \text{seq}(h, v) \) (see Fig. 1c);

(b) otherwise, the vein expression is \( \text{seq}(h, \text{simp}(v)) \) (see Fig. 1d).

The definitions express the fact that in the understanding/resuming, in the total context, of a satellite span, one should add to the units that contribute to the understanding/resuming of its parent node the most important units within the satellite span itself (given by the sequence of units in its own head expression). Let’s note that the vein expression of the parent node of this satellite, with one exception, inherits only head expressions of nuclear nodes from its own ancestors, therefore the significant units belonging to the satellite own span cannot be there and must be included explicitly. The exception mentioned refers to exactly the case when a satellite is placed on the left side of the nucleus towards
which this node is itself a satellite, and whose units have been recorded by markings. The *simpl* function will delete this influence (see an example in Fig. 2).

In Figure 2, an example of computation of head and vein expressions is displayed. First, the head expressions are computed, in a bottom-up order, starting in the leaf nodes. So, following (1.1), the leaf nodes 1, 2 and 3 will have the head expressions:

\[
H_1 = 1, \quad H_2 = 2, \quad H_3 = 3
\]

Then, the computation proceeds to node \(b\), which, according to (1.2), has just one nuclear daughter, node 2, will have:

\[
H_b = H_2 = 2
\]

The root node \(a\) has one nuclear daughter as well, node \(b\), such that (again according to (1.2)):

\[
H_a = H_b = 2
\]

The computation of heads being concluded, vein expressions can now be initiated, in a top-down manner, starting in the root. Conforming to (2.1), the vein expression of the root is

\[
V_a = H_a = 2
\]

The interpretation of the vein expressions in this VT-tree is as follows:

- **node a**, the root: the whole text is summarised by unit 2, which is the most salient unit of the whole text;
- **node 1**: to get the meaning of unit 1 in the total context, reading only 1 is not enough; 2 is also needed, because it brings the contribution of the context;
- **node b**: the meaning of the span 2 3 in the total context can be obtained by reading 1 and 2. Unit 2 solely summarizes well the span, but as a separate chunk; to get also the contribution of the context, unit 1, its left satellite, is also needed;
- **node 2**: the meaning of unit 2 (in the total context) is given by the sequence 1 2;
- **node 3**: the meaning of unit 3 (in the total context) is given by the sequence 2 3. Unit 3, only by itself is not relevant; unit 2, its nucleus, is also needed. As we see, unit 1 ceases to influence
any more the meaning at this point, because a nucleus, a very prominent piece of text, is interposed.

4 The Relationship Between Discourse Structure And Referentiality

In section 3 an informal intuition of the notion of vein has been given, followed by formal definitions. If we particularize the conditions of the informal intuitions to apply to a terminal node, we get: the vein expression of a terminal node $u$ reveals the sequence of $e\!d\!u\!s$ that are significant for understanding/summarizing $u$ in the total context. Among other things, such a statement claims that the (non-necessarily contiguous) span which is given by the vein expression of a certain $edu$ should include at least one antecedent for all anaphors belonging to that $edu$. If this would not be true, the recuperation of meaning or the summarisation could be faulty.

The first conjecture of VT (or the cohesion conjecture), defines for any discourse unit a specific domain of accessibility computed in relation with the discourse structure: antecedents of the referential expressions belonging to an $edu$ $u$ are to be found among the discourse entities anchored in the $edus$ which precede $u$ in its vein expression, and including $u$ itself. More formally, a domain of evocative referential accessibility (on short domain of evocative accessibility – $dea$) can be defined as the prefix of the vein expression of the unit the anaphor belongs to:

$$dea(u) = \text{pref}(u, \text{unmark}(V_u)).$$

In other words, this formula says that if unit $u$ includes an anaphor, then an antecedent of this anaphor can be recuperated in the sequence of units included in the prefix of the vein expression of that unit, up to the unit itself. The claim that we can find all relevant antecedents belonging to a discourse unit in a domain that is situated in text only-to-the-left of the unit itself is justified by the common cognitive nature of anaphora and cataphora (as discussed in section 2), which allows for a unique directionality in the search for antecedents, always towards the beginning of the text. The term evocative in this definition will be discussed below.

The cohesion conjecture actually hypothesizes the existence of two main classes of anaphoric references: those which observe the cohesion conjecture, called evocative (or immediate), and those which do not, called post-evocative (or inferential) (see Fig. 3).

![Figure 3. Evocative and post-evocative references. Anaphoric chains are depicted by dotted lines, and deas by thick lines. The anaphor’s unit is the last one to the right.](image)

In an evocative reference, the backward-looking chain of units anchoring the textual expressions that are referentially related with the anaphor intersects the $dea$ of the anaphor’s unit in at least one more unit apart from the anaphor’s unit itself. In post-evocative references this double intersection is missing. In (Cristea, 2000) and (Cristea et al., 2000) the evocative references are further detailed in direct and indirect. In direct references the second intersecting unit (anchoring the same discourse entity as the one referred by the anaphor) is the linearly most recent one (on the backward-looking coreferential chain), counted from the anaphor’s unit. This means that the linearly most recent antecedent can be found on the vein of the anaphor’s unit. In indirect references the two backward-
looking chains intersect in a unit that is not linearly most recent (on the backward coreferential chain) from the anaphor’s unit. There is at least one other interposing unit containing an antecedent which is skipped. However, as all entities in the anaphoric chain are coreferential, the meaning of the anaphor can still be recuperated.

The evocative references appear most frequently, are resolved quickly and can be realized at the surface by any referential material, including the most fragile, as empty subjects and pronouns. They give fluency to the text and make it cohesive. The post-evocative processes are less frequent, need a greater inferential load for their resolution and make use of strong referential material (as proper nouns).

Sometimes an anaphor belonging to the post-evocative class can be understood without even having to make a connection to an antecedent. These are usually called pragmatic references or pseudo-references. The interpretation of res in this class can be made based on knowledge that comes from outside the text, from common knowledge. Although the text contains at least one more re that realizes the same de as the anaphor, the coreferential expressions may not be represented identically in order for the text to be understood.

In the case of functional (or bridge) references, the functional link has usually a length of 1 (for instance, the engine referring back to the car) and therefore it should be consumed as an evocative reference. It could also be possible that the anchor of the bridge reference be part of a coreference chain. In this case, the functional antecedent could be found as an indirect reference, towards any of the coreferential expressions anchoring the functional link, not necessarily the linearly most closest (see Figure 4).

![Figure 4. Evocative functional references.](image)

In the following we will come back to the examples in section 2 in order to show how the deas highlighted in the VT-tree explain the referential phenomena discussed.

![Figure 5: The VT structures of Ex. 1 and Ex. 2](image)
resolution for the other pronouns: he (of unit 2) – as [John] in unit 1, and his (of unit 3) – also as [John] in unit 1.

![Diagram](image)

**Figure 6**: The VT structure of Ex. 3.

Both Figures 7a and 7b could describe the structure of Ex. 4 (they are VT-equivalent, in the sense that all nodes have the same VT expressions). In both trees the dea of unit 3 is 2 3, and this supports the comment presented in section 2 (that it in unit 3 accesses [the company’s bankruptcy] – in unit 2, rather than [Mr. Ross’ mandate as president] – in unit 1). Figure 7c shows the structure of Ex. 6, in which unit 3 has the dea: 1 2 3. The pronoun it can now access both units 1 and 2, but the shorter linear distance to unit 2 as well as the semantic restrictions make readers to prefer as antecedent the entity [Mr. Ross’ mandate as president].

![Diagram](image)

**Figure 7**: Structures for Ex. 4 and Ex. 6.

Finally, let’s look more attentively at the structures of Ex. 8 and 8’. The dialogue 1-5 is displayed in Figure 8a. The main unit of the whole dialogue is 1, the head of the root node and, as such, appearing in all vein expressions of the terminal units. It defines the theme of the dialogue: [the separation of A from Michael]. When turn (unit) 6 is uttered, unit 2, which includes the theme that unit 6 refers, [the visit in Mexico], is closed for weak reference: no node on the right frontier includes unit 2 in the vein expression. The references in 6 of Ex. 8 and 6’ of Ex. 8’ are both of an inference type, therefore, as explained by VT, hard to make. Accordingly, unit 6 uses strong referential means, while unit 6’ does not and this explains the difference.
5 Discussions

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. To make this evident, VT reflects the structure of a discourse or a dialogue as a labelled binary tree, with labels taking values in the set \{nucleus, satellite\}. The nuclearity notion is borrowed from RST, but the resemblance with this theory stops there, since the relation names are disregarded. On such a tree, VT then computes head and vein expressions. For each node of the tree, the head accounts for the saliency within the text span covered, while the vein is intended to copy the influence of the context on the same span. Finally, veins of terminal nodes support the configuration of certain domains of referentiality, as sub-spans of the whole text where referential expressions can find their antecedents.

Common intuition shows that in left-polarised trees inter-unit references are rather to nuclei than to satellites, reflecting the fact that nuclei assert the writer’s main ideas and provide the main “threads” of the discourse (Mann and Thompson, 1988). The referential transparency from satellites to their left nuclei is straightforward in Grosz and Sidner’s (1986) stack-based model and can be transposed in tree-based discourse structures through the mappings outlined by Moser and Moore (1996) and Marcu (1999). VT’s domains of referentiality computed for left-polarized trees are thus consistent with the predictions of the stack-based model.

In cases where 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 the nucleus’ 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 to entities in \(A'\) (its left satellite), but the subsequent right satellite, \(C'\), cannot refer to \(A'\) due to the interposition of the nuclear unit \(B'\). In stack-based approaches to referentiality, such configurations raise problems: as \(B'\) dominates \(A'\), \(B'\) must appear below \(A'\) on the stack, even though it is processed after \(A'\).

The domain of evocative accessibility is defined as part of the vein expression of a terminal node which prefixes the node itself. A domain of a unit is what the reader remembers immediately in the previous text, from the perspective of the given unit. The definition of dea in VT allows for a diversification of inferential references in two main classes: if an antecedent is found on the vein then the reference is called evocative, and if it is outside the vein it is called post-evocative. The existence of a class of references which cannot be satisfied on the vein seems to minimize the importance of the domain of referential accessibility. Indeed, since references can “escape” outside the vein, does the domain, as defined by VT, have a significance any longer?
We claim that the two classes of references have important distinctive features. The evocative references can be solved by fast resolution processes, because they are based on immediate associations with entities which are ‘in focus’. In such references the anaphors can have weak and very weak evoking power, like pronouns and zero-pronouns. When hierarchical adjacency is considered, an anaphor can be resolved to an antecedent which is not the closest linearly. Because co-referential expressions are organized in equivalence classes, it is sufficient if an anaphor is resolved to some member of the set. This is consistent with the distinction between direct and indirect references.

On the other hand, the post-evocative processes are inferential processes that are developed in memory. An attentive reader or listener doesn’t really forget anything, so any entity mentioned sometimes during the discourse unfolding should be reachable by a reference. The matching is realised based on the distinctive features accumulated by the preceding discourse, or by using knowledge outside the text, from the cultural sphere. This is why they require powerful referencing means (like proper nouns, for instance). After exhausting the dea, these inferences should swing the back memory space until they find an antecedent. Therefore they should be slow (computationally and cognitively), because they compel to more inference load, and should be less frequent.

An aspect not described in this paper is the theory’s account on discourse coherence (Cristea et al., 1998). Starting from deas, the notion of segment in a hierarchical sense is introduced, which generalizes the classical notion of segment as used in AST (Grosz and Sidner, 1986) and centering (Grosz et al., 1995). By this, VT generalizes centering from a local theory of coherence to a global one.

Empirical evidences on the VT’s claims on cohesion and coherence have been reported in (Cristea et al., 1998; Cristea et al., 2000) and (Ide and Cristea, 2000), with experiments developed on corpora annotated to discourse structure and coreferentiality in English, French and Romanian. In particular, these studies reveal the following: in most cases the references are direct; in fewer cases the references are indirect; in very few cases the references are inferential; inferential references which are not pragmatic signal a hard-to-make inference or a failed discourse.

Scholars dealing with the interpretation of discourse and reading in connection with the cognitive science (Kintsch and van Dijk, 1975; Schank, 1977; Cornea, 1988; Walker, 1996) generally agree on three types of memory: immediate memory (IM), short term memory (STM) and long term memory (LTM). Usually IM is defined as a sensorial storage of information, which allows the retaining of traces from the last half second. STM keeps information for few seconds. According to Miller (1956), this memory seems to have a length of 7±2 signs (words, figures, letters – depending on the context). In Cristea et al., 2003; Cristea et al., 2005) an incremental discourse parsing model is described in which the developing structure is updated with a new auxiliary tree after the reading of each sentence. The discourse tree becomes bigger and bigger as the text unfolds.

In the human memory, as well as in automatic discourse parsing systems, summarization processes must evolve in parallel with the building of the discourse structure. We believe that the STM should correspond to the dea of the last edu processed: either the last 7±2 edus in this sequence, or the same number of event structures – as representations of edus, or only words picked up from this buffer. When we replace the current unit un with the next unit un+1, actually we replace the STM dea(un) with the STM dea(un+1), bounded to a certain length. Sometimes this means a simple prolongation of the preceding dea, other times it means the shadowing of certain zone of edus and the awakening of other edus. STM is therefore made of a chain of edus (or of microstructures corresponding to edus), which is projected from the dynamically evolving discourse structure. The alterations affecting the STM string reflect the updates of the sub-discourse in focus, while reading. When the interest has moved along another direction, the content of the current vein and, consequently, of the current dea, is updated too. The inclusion and deletion from STM of certain mini-structures, therefore these “recall” and “ oblivion” processes, resemble the calling in attention of Walker’s (1996) cash memory model. The recall processes are possible from the discourse structure that is kept in a summarized form in the LTM. Evocative anaphoric processes are thus developing in the STM, while post-evocative processes are swinging that part of the discourse structure which remained in LTM.

There are as many ways to read a text as there are edus in it. These different readings are given by the edus’ vein expressions. Each vein represents a summary of the text focused on the respective unit. When the reader is focused on a certain episode or entity mentioned by the text s/he can skip entire fragments and look for the manner in which the element of interest integrates in the whole discourse. Summaries focused on different events or entities can contain elements in common, while each of them has also specific elements, although strongly correlated to the main line of the discourse. All these sub-discourses are coherent and, generally, there are no anaphoric references whose interpretation would necessitate elements outside the summary itself.

We believe that the processes of anaphora resolution and discourse structure building are interdependent to such a degree that discourse analysis should make use of them in tandem, and
combine their partial results to acquire the best discourse tree. In the same way that anaphora resolution can benefit from the discourse structure, already solved anaphora can be used in determining the best structure, which in turn contributes to the resolution of further anaphora. The constraints evidenced act as forces that, in a well-interpreted discourse, arouse a sort of state of equilibrium, resembling the minimum potential energy of a physical system. Humans have an innate cognitive mechanism that allows them to obtain naturally the most plausible interpretation of a text. When arrived there, they are invigorated by the reach of a “comfortable” mental state, which should be based on the maximal satisfaction of a constraints system. In (Cristea et al., 2005), a model and an implementation that mimics this behaviour are described. Scores contributed by the cohesion conjecture are combined with scores contributed by the coherence conjecture of VT (the hierarchical generalization of centering) in order to obtain the most “fluid” possible discourse structure (maximum of cohesion and of coherence).

VT’s account on the relationship between discourse structure and referentiality can be exploited in at least the following ways:

- to constrain a simultaneous parsing and anaphora resolution process able to produce that interpretation that requires minimum inferential load in building the structure and in identifying the antecedents of referential expressions (Cristea, 2000; Cristea et al, 2002a; Cristea et al, 2002b Cristea et al, 2005);
- to correct the discourse structure when referential links are known (Serețan and Cristea, 2002);
- to guide a process aimed at producing focused summaries (Cristea et al, 2003; Cristea et al, 2005).

The notice that slightly modified texts can display the same vein structure (although not the same tree structure) can lead to the idea that veins could be seen as a kind of sub-

way to the right frontier constraint revisited. In Proceedings of the Multidisciplinary Approaches to Discourse 2005 (MAD’05), Chorin/Berlin, Germany.


Bibliography


