Motivations and Implications of Veins Theory

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@infoiasi.ro

Abstract. The paper deals with the cohesion part of a model of global discourse interpretation, usually known as Veins Theory (VT). By taking from the Rhetorical Structure Theory the notions of nuclearity and relations, but ignoring the relations’ names, VT computes from rhetorical structures 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. Linguistics and cognitive 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 over anaphora and its resolution in relation with the interpretation of discourse.

1 Introduction

It is rather easy to decide that a text is coherent, and very difficult to risk a statement of the contrary. Not long ago, a friend of mine\(^1\) defied me that I am unable to give him a senseless sentence. So, I uttered the famous Chomskyan sentence “Colorless green ideas sleep furiously.” challenging him to find a sense. And he did, because he explained me that this sentence simply says that one night some ideas (colorless, as all ideas) came, during an agitated sleep, to the mind of a politician, a member of the green party… The example argues for the necessity of a setting (or a context) according to which to give a meaning to a discourse. Often the key to the interpretation of a discourse comes from finding this setting. This is why to some people a novel like The sound and the fury of William Faulkner is obscure and difficult to read, while for others it makes so much a delicious reading. So, the notion of setting, as a dynamic system of conventions, stays at the very basics of our capacity of understanding the world. When we refer to the interpretation of a text we usually call this setting context. Within a defined context certain spans of the text receive meanings and if we enlarge or squeeze this context we might obtain different interpretations of the same piece of text.

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 the reading comes to an end, the discourse also finishes and only a representation of it remains in the reader's memory.

There are three time axes which can be identified in the analysis of the reading or hearing of a text: the real time axis, which includes the process of reading and whatever else the reader might do between two pages or two chapters of a book, the discourse time axis, which takes into account only the linear processing of the text in the reader’s mind and practically includes parts of the real time axis, and the story time axis, which relates only to the events described in the book, not always linear, and with could be measured even on different scales. In our analysis we are interested to model only cognitive processes as projected on the discourse time axis, which means, for instance, that we will deliberately ignore anything related to forgetting in the pauses between readings.

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 existing between structure and referentiality. In Atentional State Theory (AST) \[13\] the discourse is seen a 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 towards the bottom. The Rhetorical Structure Theory (RST) \[19\] gives only a static representation while ignoring any concern on referentiality. Centering Theory \[1, 12\] uses the notion of segment from AST to propose a local theory of discourse coherence.

\(^1\) I thank Dominique Dutoit for this nice interpretation.
We will review in this paper the cohesion part of a model of global discourse interpretation, usually referred to as Veins Theory (VT), while also noticing some new facts about it. By taking from RST its notions of nuclearity and relations, but ignoring the relations’ names, VT [7] reveals a “hidden” structure in the discourse tree, called vein, which enables to determine a domain of evocative accessibility (dea) for each discourse unit, as that space of the discourse where all anaphors belonging to that unit can find an antecedent. 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 (the nucleus to nucleus references and references from nuclei to left satellites).

In the following section we give arguments in favor of an incremental modal in processing discourse. Section 3 enumerates linguistic observations that lead to the formulation of VT. The basic definitions are then revised in section 4. VT’s claim on discourse cohesion is presented in section 5. The last section leaves a synthesis of the theory, exhibits a cognitive argumentation, quickly reviews applications based on the findings of the theory and shows some possible future developments.

2 Incrementality in processing discourse

The same basic principles as in RST apply in VT:
− the principle of sequentiality [11]: 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 [21]: a relation that holds between two spans also holds between the most salient units of the spans involved in that relation.

Resembling more to AST than RST, which is a static theory of discourse structure, VT adopts an incremental processing approach to referentiality, which mimics the activity developed by human brains during reading. According to this view:
a). the discourse structure is build in the rhythm of reading,
b). a referential expression (re) either introduces a discourse entity (de) of refers one already introduced in the discourse (to the left, in left-to-right writing languages, to the right in right-to-left languages²),
c). the resolution of anaphors/cataphors is characterized by an initiation moment and a completion moment.

In the rest of this section we will elaborate on these aspects.

The elementary discourse unit (edu), usually a clause, is considered the minimal text span triggering an operation that elaborates on the developing discourse structure. We ignore any psychological evidence in support of this assumption, but the presupposition is that before the syntactic structure of the clause is united there are few elements on which to anchor a step in the process to build the discourse structure. Moreover, Şoricuţ and Marcu [29] reported that approximately 90% of all sentences in the RST-annotated corpus they have used contain a node covering exactly one sentence. This is evidence that during the elaboration of discourse structure, a parser (as well as a human brain) waits until the whole next sentence is processed in order to integrate its discourse sub-structure into the developing structure (for an automatic approach of this kind, see, [9], [10]).

As shown in [11], two types of operations are sufficient to describe the evolution of any discourse tree in an incremental manner during discourse processing: adjunction and substitution, inspired by (Lexicalized) Tree-Adjoining Grammar (LTAG) [16] [25]. In adjunction, an auxiliary tree is ”inserted” in the developing tree at the level of a certain node of the tree, called adjunction node. The sub-tree of the developing tree down the substitution node is “cut” and re-attached on a special node placed on the terminal frontier of the auxiliary tree, called foot. The substitution is the operation in which an elementary tree (which does not contain a foot) replaces a special node, called substitution node, placed on the frontier of the developing tree. Moreover, it can be proved formally that the principle of sequentiality is observed during an incremental process only if the auxiliary trees are left-footed and the adjunction is restricted to the generalized right-frontier³, result consistent with a similar one based on empirical observations [32].

The left direction in the search for an antecedent can be adopted in case of anaphors as well as cataphors. The first mention of an entity, irrespective of its surface realization, nominal or pronominal,

² 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.
³ The inner right frontier of a developing tree with substitution nodes is the right frontier of the subtree rooted in the left sibling of the most inner substitution node.
introduces a representation of the entity, which is only referred to subsequently, or referred to and supplemented, by subsequent res that are in a referential relation with it. From this point of view, in any referential process, a combination of the following three elementary acts can be put in evidence:

- **creation (C)** of a content-empty de;
- **add (A)** of information into an existent de representation;
- **pure reference (R)**, without information add.

For instance, in the following text:

> "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)

if we concentrate only on the three referential expressions associated with the entity [Lord Watton]\(^4\), the first he pronoun realizes a CA-act since it introduces a new entity, and adorns it with rather poor morpho-semantic features: \{sem=person, gender=masculine, number=singular\}). The second he pronoun accomplishes an R-act since it simply references the already existent entity. The proper noun Lord Henry Wotton refers the entity while also refining the description or supplementing it with new information, therefore performing an RA-act. At this moment the representation of the entity reaches the form: \{sem=lord, name=’Henri Wotton’, gender=masculine, number=singular\].

Anaphoric acts in reference chains can be C, CA and RA. Most of the first mentionings realize CA-acts, since they introduce entities decorated with some features (the case of common nouns, for instance), which can be antecedents for subsequent anaphors. Pure first-mentioning C-acts are rather rare and are always associated with cataphoric references. For instance in English they can be realized only by cataphoric it-pronouns\(^5\), while in some languages which do not have singular neuter pronouns, like Romanian for instance, they can be realized by empty pronouns.

Ex. 1:

\[ Ø_1 A \text{ apărăut deodată. Nimeni nu ştie de unde } \text{Ø}_2 \text{ a venit. } \text{Ø}_3 \text{ Era înalt, brunet şi } \text{Ø}_4 \text{ călărea un cal murg.} \]

\[ \text{Ø}_1 (=he/she/it) \text{ Appeared sudently. Nobody knows where } \text{Ø}_2 (=he/she/it) \text{ came from. } \text{Ø}_3 (=he/she) \text{ Was tall, dark-haired and } \text{Ø}_4 (=he/she) \text{ was riding a dark bay horse.} \]

In Ex. 1 Ø\(_1\) C-acts a new empty de. Nothing can be said about it: could be a person, a comet or an idea, which appeared suddenly in a community. The rest of referential expressions accomplish, in this example, RA-acts. Ø\(_2\) refers this entity while adding also a semantic restriction: a direction for something that comes can be determined for a person, for a comet but not for an idea. Then, Ø\(_2\) and Ø\(_3\) constraint this entity down to a person (only a person can be dark-haired and can ride a horse), therefore we have again RA-acts. Gender of the adjectives indicates in Romanian also this person as being a male. Pure coreferential chains are C- or CA-acts followed by sequences of R-acts. More often, subsequent references add information to the representation of the first mentioning (RA-acts). Only bridge anaphorae can be classified as CAR-acts because they involve des that, at the moment of the first mentioning as res, also refer, through their properties, other des to which they are semantically related. So, in *the computer... its accumulator*, the anaphor *its accumulator* accomplishes an CAR-act, since it simultaneously proposes the entity *the computer’s accumulator* and also refers the entity *the computer*.

In some cases, the anaphora resolution moment may be deferred until other discourse elements intervene to help in the elucidation of the anaphoric reference. This is the case in the following example [30, p. 221]:

Ex. 2:

> Police officer David Cheshire went to Dillard’s home. Putting his ear next to Dillard’s head, Cheshire heard the music also.

The reader can disambiguate the pronoun his, only when s/he reads Dillard’s head, Cheshire heard the music also.

\[^4\text{ Everywhere in this paper we will note, as text put in square brackets and in bold, des, either by conventional names or by collections of pairs attribute=value. The corresponding res in the text are noted by italics.}\]

\[^5\text{ In principle, they could also be realized, as a first mention, by nouns expressing the most general semantic claseses (like entity or think), as in: The entity finally got shape. where the entity could subsequently be revealed to be an idea as well as a cat.}\]
story and a man cannot put his head next to his own head. Therefore, the resolution cannot be accomplished the very moment when the pronoun is read, but later, when new elements appear that can be used as disambiguation constraints. Later reading of the proper noun will strengthen the belief that the referent is [Cheshire], as inferred.

3 VT’s intuitions

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 organization 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 $B$” when we mean “a re belonging to the unit $A$ refers a de introduced or referred in/from unit $B$”):

a). right satellites or nuclei can refer their left nuclear siblings: in combinations $u_1^n R u_2^s$, or $u_1^n R u_2^n$, $u_2$ can refer $u_1$.

Ex. 3:

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 the entity [John] introduced by the referential expression John in the first unit.

b). a right nucleus can refer a left satellite: in combinations $u_1^s R u_2^n$, $u_2$ can refer $u_1$ as in:

Ex. 4:

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

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

b). a right satellite of a nucleus $u$ is not accessible from another right sibling of $u$, nuclear or satellite: in combinations $(u_1^s R_1 u_2^s) R_2 u_3^n$ or $(u_1^s R_1 u_2^s) R_2 u_3^s$, $u_3$ can refer $u_1$ but not $u_2$.

Ex. 5:

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

The sequence 2-3-4 ELABORATES on 1. The sequence 2-3 is in a relation of CONTRAST (a paratactic relation) towards 4, while unit 3 ELABORATES on 2. The structure is therefore: $u_1^s R_1 ((u_2^n R_2 u_3^s) R_3 u_4^s)$. 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, and also 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 the 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 to find its antecedent in the common upper nucleus – unit 1.

c). a nucleus blocks the reference from a right to a left satellite: in combinations $(u_1^s R_1 u_2^s) R_2 u_3^s$, $u_3$ can refer $u_2$ but not $u_1$.

Ex. 6:

1. With one year before finishing his mandate of president of the company,
2. Mr. W. Ross has begun to manipulate its bankrupt.
3. There were rumors that he has obtained it by fraud.

---

6 In the examples in this section we will mark with numbered $u$ – units and with $R$ – relations. An upper $n$ or $s$ at the shoulder of an expression signals the corresponding text span as being a nucleus or a satellite.
In this example the reader is confused on who the referent of the pronoun *he* in unit 3 actually could be. 1 and 3 are both satellites of unit 2: 1 is in a CIRCUMSTANCE relation towards 2, while 3 is intended to give a BACKGROUND for 2, if *it* would be perceived as referring [the mandate of president of Mr. Ross]. However this coreferential link is found with difficulty, which makes the whole discourse to fail. It can be repaired in two ways:

Ex. 7:

1. Mr. W. Ross has begun to manipulate the bankrupt of his company.
2. with one year before finishing his mandate of president.
3. There were rumors that he has obtained it by fraud.

In Ex. 7 unit 2 is a satellite of 1, and 3 – a satellite of 2. The reference *it*=[Mr. Ross’ mandate of president] can be recuperated without difficulty. The motivation for the failing of Ex. 6 compared to the acceptance of Ex. 7 stays not in the linearly longer distance between the anaphor and antecedent in Ex. 6 than in Ex. 7, but in the fact that a nuclear unit is interposed between the unit of the anaphor and unit of the antecedent in Ex. 6, while there is not this case in Ex. 7.

If the reference is eliminated than the discourse is also repaired:

Ex. 8:

1. With one year before finishing his mandate of president of the company
2. Mr. W. Ross has begun to manipulate its bankrupt.
3. There were rumors that he has been elected by fraud.

4 VT’s 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 dea. 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 x’s antecedent 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 dot-to-dot sentence, but sometimes they may be reduced to a clause; 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 [21]; c) terminal nodes of the tree represent elementary 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; d) 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.

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;
- mark(\(\alpha\)) is a function that takes a string of symbols \(\alpha\) and returns each symbol in \(\alpha\) marked in some way (e.g., within brackets);
- unmark(\(\alpha\)) is the reverse function of mark(). It removes all markings attached to symbols in the expression \(\alpha\) (e.g. unmark(\(\alpha\cdot\text{mark}(\beta::\gamma))\) = \(\alpha\cdot\beta::\gamma\));
- simpl(\(x\)) is a function that eliminates all marked symbols from its argument, if they exist, e.g. simpl(mark(\(\alpha\)) = \(\alpha\), the empty string, and simpl(\(\alpha\cdot\text{mark}(\beta::\gamma))\) = \(\alpha::\gamma\);
- 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 markings, if they exist and seq(\(\alpha, \beta\)) = \(\beta\), seq(\(\alpha, \text{seq}(\beta)\) = seq(seq(\(\alpha\)), \(\beta\)) = seq(\(\alpha, \beta\));
- \(H(n)\) and \(V(n)\) are the notations for the head and vein expressions of a node \(n\);
- 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. The notion of head in VT is equivalent to that of Marcu’s promotion set [21]. The intention in the head expression of a node of a discourse tree is to capture the sequence of the most important units in the span of text covered by the node. It is a sequence of unit labels as follows:

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.

Note that the recursive definition of head induces a bottom-up computation over the tree structure.

The vein expression of a node is intended to give the sequence of edus which are significant for summarizing, in the context of the whole text, the span of text covered by the node. In the vein expression of any node in the discourse structure, are included edus belonging to the span covered by the node, possibly together with edus outside the span.

In the following, the whole text is called total context. In Fig. 1, the nodes to which the definition currently applies are depicted in grey. They are 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 is marked for the head expression, vein expressions are computed top-down for each node in the tree:

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 the root node, should be made of the most significant edus that are necessary to understand/resume 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, this gives us the definition of the head expression of the root node.

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 seq(mark(h), v) (see Fig. 1b).

   The definitions say that in order to understand/resume, 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/resume, 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/resume its right nucleus, but should be ignored for any other right satellite of this nucleus (case commented in Ex. 6). The marking function mark signals the contribution of this left satellite for a subsequent removal to be operated in the vein expression of a right satellite (see 3b below). On the contrary, twin nuclei cannot be understood/resumed one without the other, meaning that the same units are significant to understand/resume each one of them as their union span.

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 seq(h, v) (see Fig. 1c);
   b) otherwise, the vein expression is seq(h, simpl(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 nucleus 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).

---

7 By synthesis, or summary, of a text span we understand a shorter text, which still renders the original idea of the text. Irrespective whether it is realized by paraphrasing or by concatenating sub-sequences of the original text [18], 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 resumed 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, “resuming” is equivalent to “understanding” because what we are usually left after the reading of a text is a synthesis of it.
5 The relationship between discourse structure and referentiality

If we particularize the intuition behind the vein expression to a terminal node, we obtain: the vein expression of a terminal node \( u \) gives the sequence of \( \text{edus} \) that are significant for understanding/summarizing \( u \) in the total context. Among other things, which we will not discuss in this paper, this means that, within the material indicated by the vein expression of an \( \text{edu} \), antecedents of all anaphors belonging to that \( \text{edu} \) must be found. More precisely, seen: – the semantic nature of the anaphoric relation [14], – a representation of anaphoric relations in which \( \text{res} \) of a textual layer are linked to representations of \( \text{des} \) on a semantic layer, as the one proposed by Cristea and Dima [4] – and 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, we are lead to the definition of a domain of evocative referential accessibility (on short \( \text{domain of evocative accessibility} – \text{dea} \)):

\[
dea(u) = \text{pref}(u, \text{unmark}(V(u))).
\]

The definition of \( \text{dea} \) formalizes the first conjecture of VT (or the cohesion conjecture), which relates, for all units, a specific domain of accessibility with the discourse structure: antecedents of \( \text{res} \) belonging to an \( \text{edu} \) \( u \) are mostly found, among the \( \text{des} \) anchored in the \( \text{edus} \) which precede \( u \), \( u \) included, in its vein expression.

The first conjecture hypothesizes two types of anaphoric processes: evocative (or \emph{immediate}) and post-evocative (or \emph{inferential}). The evocative processes 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, do need a greater inferential load for their resolution and make use of strong referential material (as proper nouns).

If we transfer this classification to the anaphoric references involved in these processes, we will have evocative and post-evocative references (see Fig. 3). In the evocative references, the backward-looking chain of units anchoring \( \text{res} \) that are referentially related intersects the \( \text{dea} \) of the anaphor’s unit in at least one more unit than the anaphor’s unit itself. In post-evocative references this double intersection is missing. In [3] and [6] the evocative references are further detailed in direct and indirect.
In **direct references** the second intersecting unit (looking backward from the anaphor’s unit) is the linearly most recent one, counting from the anaphor’s unit, anchoring the same *de* as the one referred by the anaphor (in case of coreference), or a *de* that is anaphorically related to the anaphor’s *de* (in case of functional reference). In **indirect references** the two backward looking chains intersect in a unit that is not linearly most recent from the anaphor’s unit.

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 *re* 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.

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

### 6 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. 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 [19]. This is shown in the computation of veins over (binary) left polarized 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.

Given the mapping of Grosz and Sidner’s [13] stack-based model of discourse structure on RST structure trees outlined by Moser and Moore [23] and Marcu [20], 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, as defined by the theory, seem to minimize the importance of the domain of referential accessibility, 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 people 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 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 de that already exists in the dea of the unit the anaphor belongs to [4]. These are fast processes, direct ones being faster and more frequent than indirect ones. 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. 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.

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 the 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 – computationally and cognitively (compel to more inference load), require more powerful referencing means (like proper nouns), and are less frequent.

An aspect not described in this paper is VT’s account on discourse coherence [7]. Starting from deus, the notion of segment in a hierarchical sense is introduced, which generalizes the classical notion of segment as employed in AST [13] and CT [12]. By this, VT generalizes CT 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 [6], [7] and [15] 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 less cases the references are indirect; in very few cases the references are pragmatic; inferential references which are not pragmatic signal a hard-to-make inference or a failed discourse. Moreover, it can be proved that VT’s assumptions regarding the cohesion are stable to the change of granularity (the limit below which material edus are considered) from lower to upper.

A side effect of corpus research motivated by the evaluation of VT claims was the notice that there is a strong relationship between the different kinds 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 (inferential). Four types of inferential references have been discovered: pragmatic, proper nouns, common nouns and pronouns, which revealed to have descending frequencies, in this order. Pragmatic and proper nouns references are easily resolved, which makes their use much less restricted by the placement of an antecedent on a current dea. At the other pole, pronouns are very fragile evoking means, and, as such, a message emitter employs them when s/he is certain that the current structure of the discourse allows for easy recuperation of the antecedent on the dea of the anaphor. The alignment of the evoking power of referential expressions with the percentage of exceptions of references outside the deus shows that the predictions of VT in the cohesion conjecture 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 dea.

Scholars dealing with the interpretation of discourse and reading in connection with the cognitive science [2], [17], [26], [31] generally, agree on three types of memory: immediate memory (IM), short term memory (STM) and long term memory (LTM). IM is 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 [22], the length of this memory seems to be of 7±2 signs (words, figures, letters – depending on the context), while others estimate this “buffer” to an average of 13-15 words [24]. In [9] and [10] 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 be linked to the dea of the last edus 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, actually we replace the STM dea(u) with the STM dea(u+1). Sometimes this means a simple prolongation of the preceding dea, other times it means the deletion of the most distant in time unit and the inclusion of a new unit – the current edu. STM is therefore made of a chain of edus (or of microstructures corresponding to edus), which is projected from the dynamic 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 on
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 [31] 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 of an inferential type, and necessitate greater inference load to recover des from memory or evoke entities kept in the generic cultural sphere of the individual. We believe these evolve also on the developing discourse structure, but leaving the dea when the resolution failed there.

There are as many ways to read a text as many edu it contains. These different lectures are given by the edu’s 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 of entities can have elements in common while each of them contains also specific elements, however strongly correlated to the main line of the discourse. All these sub-discourses are coherent and, generally, there are no anaphoric references for whose interpretation to make use of 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 analysis. 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-understood discourse, give rise to 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 [9] a model and an implementation that mimics this behavior is described. Scores contributed by the cohesion conjecture are combined with scores contributed by the coherence conjecture of VT (hierarchical generalization of CT) 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 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 [3], [5], [8], [9];
− to correct discourse structure when referential links are known [28];
− to guide a process aimed at producing focused summaries [9], [10].

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-specification representation [27], a direction which has not been investigated yet. Also, as trees annotated at discourse structure and veins can facilitate rapid interrogation for the computation of focused summaries on any discourse entity or event mentioned in the text, it would be worth investigating an RDF representation of vein structures, obtained by processes of automatic parsing, with interesting applications in Semantic Web.

Bibliography


