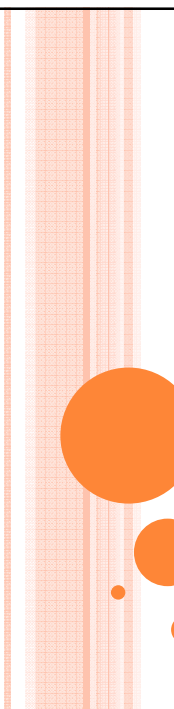




DISCURSO – PARTE 3

SCC5869 Tópicos em Processamento de Língua Natural

Thiago A. S. Pardo



VEINS THEORY
(CRISTEA ET AL., 1998)

TEORIA DAS VEIAS

- Para cada unidade discursiva, identificam-se **veias** na estrutura discursiva que incluem unidades discursivas que formam o **domínio de acessibilidade referencial** da unidade em questão
- Uma vez que os domínios são identificados, centering poderia ser aplicado globalmente

3

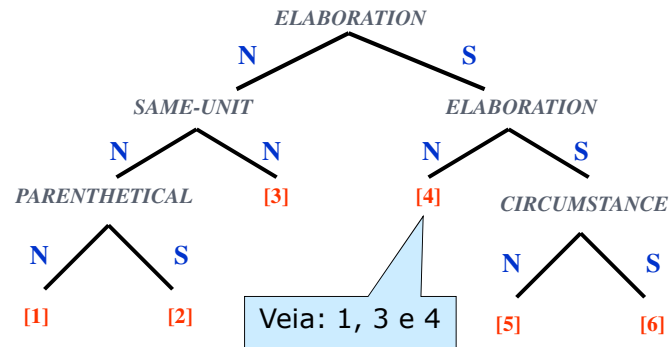
DEFINIÇÃO: VEIAS

- As veias são definidas sobre estruturas retóricas da **RST**
- Para definição das veias, somente topologia da árvore RST e nuclearidade são necessárias
 - **As relações em si não são importantes**

4

EXEMPLO

- Veia da unidade 4
 - Pode “acessar” unidades 1 e 3



DEFINIÇÃO: VEIAS

- *Veia: subsequências de unidades discursivas que compõem a árvore retórica*
- Funções usadas para o cômputo das veias
 - **mark(x)**: recebe uma string x e a coloca entre parênteses (marcando-a, portanto)
 - $\text{mark}(a)=(a)$
 - **simpl(x)**: elimina os elementos marcados de x
 - $\text{simpl}(a(b)cd(e)f)=acdf$
 - **seq(x,y)**: coloca os argumentos em ordem de leitura (na ordem em que ocorrem no texto)
 - $\text{seq}(3,2)=2\ 3$

CÔMPUTO DAS VEIAS

○ 2 passos

- Cálculo das *heads*, de forma *ascendente*, para cada nó da árvore
 - Elementos mais nucleares
- Cálculo das *veias* de forma *descendente*, também para cada nó da árvore

7

CÔMPUTO DAS VEIAS

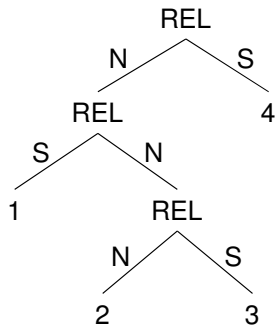
○ Passos para calcular *heads*

- A *head* de um nó terminal é o próprio nó terminal
- A *head* de um nó não-terminal é a concatenação das *heads* de seus nós filhos nucleares

8

CÔMPUTO DAS VEIAS

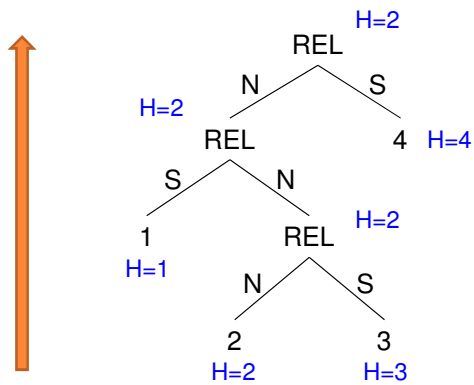
- Exercício: calcule a *head* de cada nó



9

CÔMPUTO DAS VEIAS

- Exercício: calcule a *head* de cada nó



10

CÔMPUTO DAS VEIAS

○ Passos para calcular veias

- A veia da raiz é sua própria *head*
- Para cada nó nuclear cujo pai tem veia v , a veia é
 - Se o nó tem um irmão esquerdo satélite com *head* h_{esq} , então a veia é $\text{seq}(\text{mark}(h_{\text{esq}}), v)$
 - Caso contrário, a veia é v
- Para cada nó satélite de *head* h cujo pai tem veia v , a veia é
 - Se o nó é um filho esquerdo, então a veia é $\text{seq}(h, v)$
 - Caso contrário, a veia é $\text{seq}(h, \text{simpl}(v))$

11

CÔMPUTO DAS VEIAS

○ Passos para calcular veias

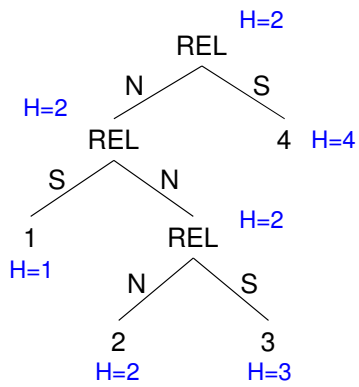
- A veia da raiz é sua própria *head*
- Para cada nó nuclear cujo pai tem veia v , a veia é
 - Se o nó tem um irmão esquerdo satélite com *head* h_{esq} , então a veia é $\text{seq}(\text{mark}(h_{\text{esq}}), v)$
 - Caso contrário, a veia é v
- Para cada nó satélite de *head* h cujo pai tem veia v , a veia é
 - Se o nó é um filho esquerdo, então a veia é $\text{seq}(h, v)$
 - Caso contrário, a veia é $\text{seq}(h, \text{simpl}(v))$

Pouco claro,
mas funciona

12

CÔMPUTO DAS VEIAS

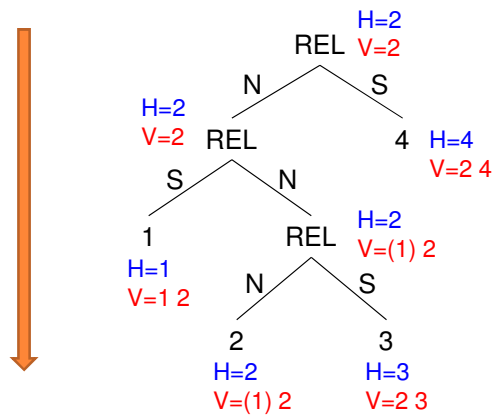
- Exercício: calcule a veia de cada nó



13

CÔMPUTO DAS VEIAS

- Exercício: calcule a veia de cada nó

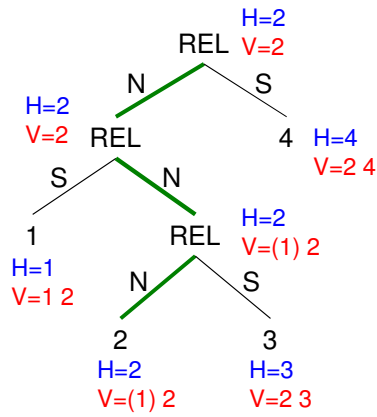


14

CÔMPUTO DAS VEIAS

- Exercício: calcule a veia de cada nó

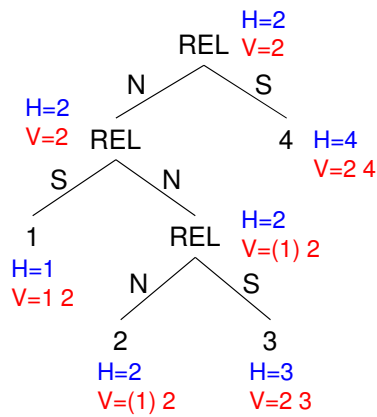
Veia principal



15

CÔMPUTO DAS VEIAS

- Exercício: calcule a veia de cada nó



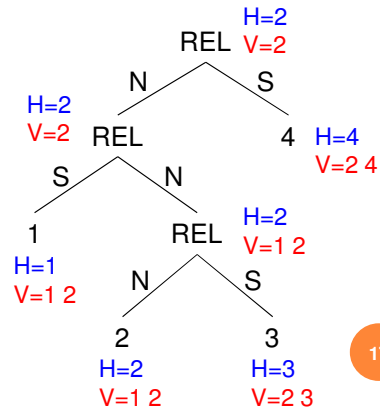
A veia de cada nó define seu **domínio de acessibilidade**: o nó 2 enxerga o nó 1, mas o nó 3 não enxerga o nó 1

16

DOMÍNIO DE ACESSIBILIDADE

- Após desmarcar as veias, o domínio de acessibilidade referencial de uma unidade discursiva u , $\text{acc}(u)$, é composto pelos **elementos de sua veia menores ou iguais a u**
 - E os maiores do que u ?

O domínio de acessibilidade de 3 é composto pelas unidades 2 e 3



17

TEORIA DAS VEIAS

- Conjectura 1:** referências de uma unidade são possíveis somente em seu domínio de acessibilidade
 - Se B é uma unidade e contém uma expressão referencial \mathbf{b} , então \mathbf{b} realiza um centro que aparece pela primeira vez ou se refere a um centro $\mathbf{a} \in A$ realizado antes, tal que $A \in \text{acc}(B)$
 - Se 1 não se aplica, se A , B e C são unidades, $\mathbf{c} \in C$ e se refere a $\mathbf{b} \in B$, mas $B \notin \text{acc}(C)$, então há $\mathbf{a} \in A$, tal que $A \in \text{acc}(B)$ e $A \in \text{acc}(C)$, e \mathbf{b} e \mathbf{c} se referem a \mathbf{a}
 - Se 1 e 2 não se aplicam, então a referência em C pode ser entendida sem seu referente

18

TEORIA DAS VEIAS

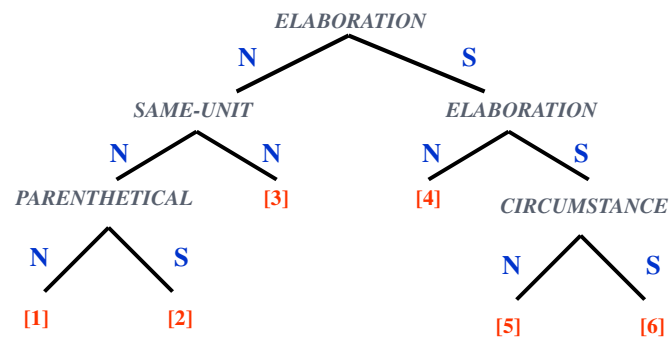
- **Conjectura 1:** referências de uma unidade são possíveis somente em seu domínio de acessibilidade
 1. Se B é uma unidade e contém uma expressão referencial \mathbf{b} , então \mathbf{b} realiza um centro que aparece pela primeira vez ou se refere a um centro $\mathbf{a} \in A$ realizado antes, tal que $A \in \text{acc}(B)$
 2. Se 1 não se aplica, se A, B e C são unidades, $\mathbf{c} \in C$ e se refere a $\mathbf{b} \in B$, mas $B \notin \text{acc}(C)$, então há $\mathbf{a} \in A$, tal que $A \in \text{acc}(B)$ e $A \in \text{acc}(C)$, e \mathbf{b} e \mathbf{c} se referem a \mathbf{a}
 3. Se 1 e 2 não se aplicam, então a referência em C pode ser entendida sem seu referente

19

Não importa a granularidade de segmentação discursiva

EXERCÍCIO

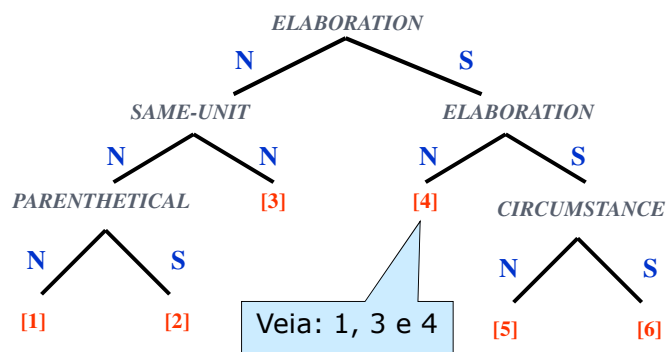
- Em duplas, calculem as veias



20

EXERCÍCIO

- Se a unidade 4 for incluída em um resumo, que outras unidades devem ser incluídas também para evitar anáforas não resolvidas?



21

EXERCÍCIO

[1] A empresa Produtos Pirata Indústria e Comércio Ltda., de Contagem [2] (na região metropolitana de Belo Horizonte), **[3] deverá registrar este ano um crescimento de produtividade nas suas áreas comercial e industrial de 11% e 17%, respectivamente. [4] Os ganhos são atribuídos pela diretoria da fábrica à nova filosofia [5]** que vem sendo implantada na empresa desde outubro do ano passado, **[6]** quando a Pirata se iniciou no Programa Sebrae de Qualidade Total.

22

TEORIA DAS VEIAS

- É possível atribuir uma **nota de coerência** para cada discurso
 - Notas para **segmentos individuais** (segundo *centering*) ou para o **texto todo** (usando as veias)

Tipo de transição	Nota
Continuidade de centro	4
Retenção de centro	3
Mudança de centro suave	2
Mudança de centro abrupta	1
Sem C_b	0

23

TEORIA DAS VEIAS

- Nota de coerência segundo *centering*: para cada segmento, soma das notas das transições / número de transições; em seguida, soma das notas de cada segmento / número de transições entre eles
- Nota de coerência segundo teoria das veias: soma das notas das transições (computadas sobre as veias) / número total de transições

24

TEORIA DAS VEIAS

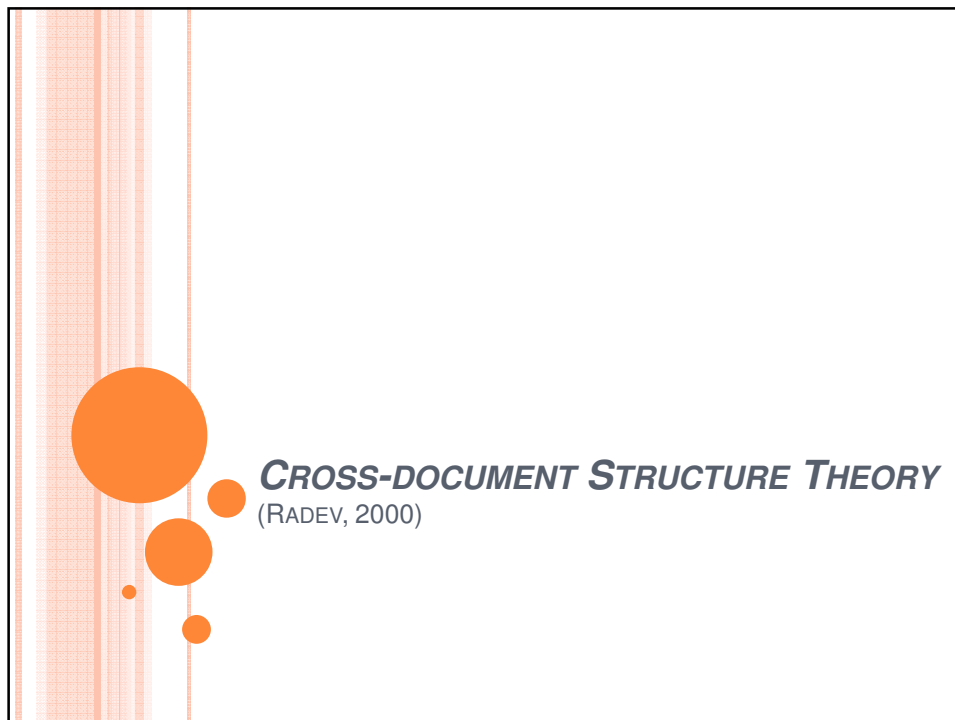
- **Conjectura 2:** a nota de coerência global segundo a teoria das veias é pelo menos tão alta quanto a nota de coerência calculada via *centering*
 - Ou seja... teoria das veias espelha a coerência pelo menos tão bem quanto *centering*

25

ALGUMAS CONSIDERAÇÕES

- Veias ou Veios?
- RST e a relação ATTRIBUTION
 - Relação artificial que atrapalha o processo

26



Um pouco de história

- Principais **pontos no tempo**
 - Trigg e o sistema TextNet (1983, 1986)
 - RST (Mann e Thompson, 1987)
 - Radev e Mckeown (1995): SUMMONS e seus operadores
 - Radev (2000): CST

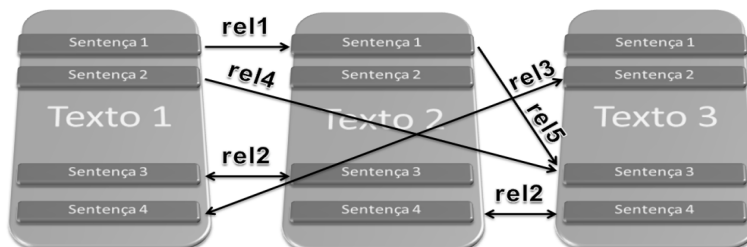
CST

- Teoria **discursiva** multidocumento
- 24 relações para documentos que versam sobre um mesmo assunto
- Motivada por questões de aplicações
 - Sumarização, principalmente

29

CST

- Modelo semântico-discursivo de estruturação multidocumento
 - São definidas relações entre partes (de quaisquer granularidades) dos documentos/textos



30

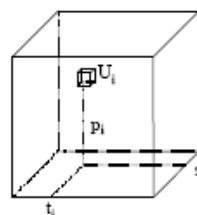
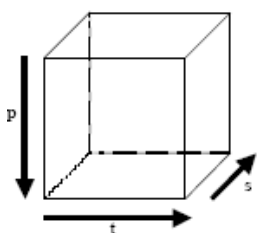
CST

o Estruturas de dados complementares

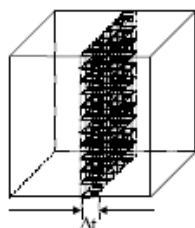
- Cubo multidocumento: fonte (*source*), tempo (*time*) e posição (*position*) dos segmentos textuais
- Grafo multidocumento: relações multidocumento

31

CST: cubo multidocumento



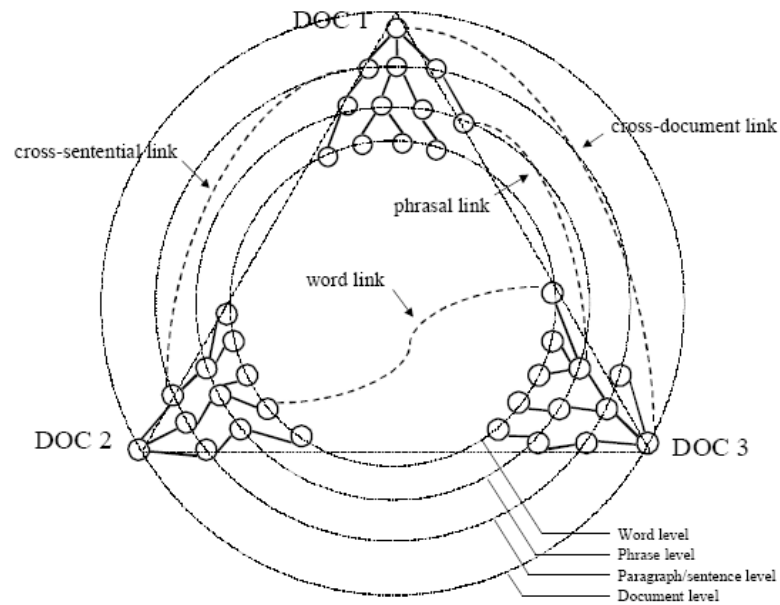
Unidade de documento U_i , sendo que um documento é uma seqüência $U_1 \dots U_n$ projetada nos eixos fonte e tempo



Recorte no tempo

32

CST: grafo multidocumento



33

CST para o inglês

o CSTBank (Radev, 2003)

- Baixa concordância entre anotadores
- *Since it describes relationships that hold across multiple documents rather than across spans of text within the same document, it makes no assumptions about authors' intentions in creating cohesion in texts*

34

CST

○ Relações originais

Identity	Modality	Judgment
Equivalence	Attribution	Fulfillment
Translation	Summary	Description
Subsumption	Follow-up	Reader profile
Contradiction	Elaboration	Contrast
Historical background	Indirect speech	Parallel
Cross-reference	Refinement	Generalization
Citation	Agreement	Change of perspective

Exemplo

○ *Contradiction, overlap, historical background* (←)

D1: Um acidente aéreo na localidade de Bukavu, no leste da República Democrática do Congo (RDC), matou 13 pessoas na quinta-feira à tarde, informou nesta sexta-feira um porta-voz das Nações Unidas.

D2: Ao menos 17 pessoas morreram após a queda de um avião de passageiros na República Democrática do Congo. Segundo um porta-voz da ONU, o avião, de fabricação russa, estava tentando aterrissar no aeroporto de Bukavu em meio a uma tempestade. O Congo tem um histórico de queda de mais de 30 aviões.

EXERCÍCIO

- Parte 1 – estudo de um conjunto de relações

37

- Parte 2 – análise do par de textos

O médico pessoal do argentino Diego Maradona, Alfredo Cahe, revelou nesta segunda-feira que uma recaída da hepatite aguda de que sofre foi o motivo da nova internação do ex-craque.

Maradona havia recebido alta no último dia 11, mas voltou a ser internado na sexta-feira e os boletins médicos não especificaram o que se passava com o ex-jogador --Cahe descartou pancreatite ou úlcera.

"Maradona teve uma recaída na hepatite aguda. Agora está estável. Apesar de ter melhorado no domingo, deverá continuar internado", disse Cahe, em declarações ao jornal "La Nación".

Maradona, 46, desenvolveu um hepatite tóxica por excesso de consumo de álcool, o que já o manteve internado durante 13 dias antes da primeira alta.

Cahe disse ainda que Maradona não voltou a consumir bebidas alcoólicas e que as causas da recaída estão sendo investigadas.

BUENOS AIRES - Maradona voltou a ter problemas de saúde no fim de semana.

Internado em um hospital em Buenos Aires, ele teve uma recaída e voltou a sentir dores devido a hepatite aguda que o atinge, segundo seu médico pessoal, Alfredo Cahe.

"Agora está estável. Mesmo com esta melhora, ele continuará internado", disse o médico, que descartou a possibilidade do ex-jogador ter uma pancreatite (inflamação do pâncreas, órgão situado atrás do estômago e que influencia na digestão).

Cahe reforçou que Maradona ainda tem problemas.

"Os valores hepáticos dele na avaliação não estão equilibrados e ele não está bem. Mas não é nada grave", afirma, em entrevista ao diário La Nación.

No domingo, Maradona assistiu ao empate por 1 a 1 no clássico Boca Juniors e River Plate pela televisão.

Os torcedores do Boca, que compareceram em grande número ao Estádio La Bombonera, levaram muitas faixas e bandeiras com mensagens de apoio ao ídolo argentino.

Sua filha, Dalma, foi ao estádio assistir ao jogo.

E a história continua...

- Afantenos et al. (2004, 2008) e críticas a CST
 - Relações muito genéricas, granularidades diferentes
 - Proposta de relações sincrônicas (mesmo tempo) e diacrônicas (mesma fonte) para domínio específico
- Etoh e Okumura (2005) e o refinamento da CST para o japonês
- Maziero et al. (2010) e a tipologia de relações
- Murakami et al. (2009, 2010) e relações entre opiniões e fatos para o japonês
 - Relações: *agreement*, *confinement*, *conflict* e *evidence*

39

CST

- Para o português
 - Córpus CSTNews
 - 50 grupos de textos jornalísticos
 - Ferramenta semi-automática de anotação de textos
 - CSTTool
 - Parser discursivo em construção
 - Aprendizado de máquina

40

CST E SUMARIZAÇÃO

- Vários trabalhos, método padrão

