

DISCURSO

SCC5908 Introdução ao Processamento de Língua Natural

SCC0633 Processamento de Linguagem Natural

DISCURSO VS. DISCURSOS

- *A discourse is an **instance of language use** whose type can be classified on the basis of such factors as **grammatical and lexical choices** and their distribution in*
 - ***main versus supportive materials***
 - ***theme***
 - ***style, and***
 - *the framework of **knowledge** and **expectations** within which the addressee interprets the discourse*

DISCURSO VS. DISCURSOS

- *A discourse is an **extended sequence of sentences** produced by one or more people with the aim of conveying or exchanging information.*

Mitkov, R. (2004). *The Oxford Handbook of Computational Linguistics*.

DISCURSO VS. DISCURSOS

- *A discourse is any string of language – usually one that is **more than one sentence long**. Textbooks, novels, weather reports and conversations are all discourses.*

Russel, S. and Norvig, P. (2003).
Artificial Intelligence: A Modern Approach.

DISCURSO VS. DISCURSOS

- **Diversas perspectivas de análise**
 - Texto **vs.** discurso
 - Texto **&** discurso
 - Texto-discurso

DISCURSO

- *Um texto é mais do que uma simples sequência de sentenças justapostas*
 - Estrutura altamente elaborada
 - Diversas propriedades em níveis variados
 - Coesão
 - Coerência
 - Tessitura
 - Progressão temática
 - Tópicos
 - Intenções
 - Estilo
 - Etc.
 - Há uma “sintaxe textual”

COESÃO VS. COERÊNCIA

- **Coesão**: uso de mecanismos linguísticos para conectar unidades textuais → “cola” textual
 - Palavras (mesma palavra, sinônimos, hiperônimos, etc.)
 - Expressões referenciais
 - Marcadores discursivos, frases e palavras indicativas
 - Pontuação
 - Paralelismo sintático
 - Etc.
- Superfície textual
- **Coerência**: relação de significado entre unidades textuais, como o significado do discurso pode ser inferido a partir das partes textuais
 - Conteúdo textual

COESÃO VS. COERÊNCIA

○ Fatos

- Há textos coesos e coerentes
- Há textos não coesos e não coerentes
- Apesar da capacidade humana de dar sentido a quase tudo!

João escondeu as chaves do carro de Pedro. Ele estava bêbado.

João escondeu as chaves do carro de Pedro. Ele gosta de espinafre.

○ Questões

- Há textos coesos e não coerentes?
- Há textos não coesos e coerentes?

EXEMPLO: ANÁLISE DISCURSIVA

- *The Tin Woodman went to the Emerald City to see the Wizard of Oz and ask for a heart. After he asked for it, the Woodman waited for the Wizard's response.*
 - *he = Tin Woodman (e não Wizard of Oz)*
 - *it = the heart (e não Emerald City)*
 - *the Wizard = the Wizard of Oz*
 - *the Woodman = the Tin Woodman*
 - *to see the Wizard of Oz and ask for a heart = razão para The Tin Woodman went to the Emerald City*
 - Sequência temporal entre as sentenças
 - Marcadores explícitos: *after*
 - Tópico = *asking for a heart to the wizard*
 - Etc.

EXEMPLO: ANÁLISE DISCURSIVA

- Gracie: Oh, yeah... and then Mr. And Mrs. Jones were having matrimonial trouble, and my brother was hired to watch Mrs. Jones.
- George: Well, I imagine she was a very attractive woman.
- Gracie: She was, and my brother watched her day and night for six months.
- George: Well, what happened?
- Gracie: She finally got a divorce.
- George: Mrs. Jones?
- Gracie: No, my brother's wife.
 - E esse caso?

EXEMPLO: ANÁLISE DISCURSIVA

- Gracie: Oh, yeah... and then Mr. And Mrs. Jones were having matrimonial trouble, and my brother was hired to watch Mrs. Jones.
 - George: Well, I imagine she was a very attractive woman.
 - Gracie: She was, and my brother watched her day and night for six months.
 - George: Well, what happened?
 - Gracie: She finally got a divorce.
 - George: Mrs. Jones?
 - Gracie: No, my brother's wife.
- Referentes de pronomes e nomes, resolução do *my*, mecanismos de “continuidade” (*well*), turnos, expectativas, humor, progressão temática, intenção de manter uma conversa/ desabafar, etc.

DISCURSO

○ Tipos

- “Monólogo”
 - Discurso em um único sentido, do escritor para o leitor
- Diálogo
 - Turnos
 - Humano-humano, humano-máquina

○ Material

- Monodocumento: um único texto
- Multidocumento: vários textos

DISCURSO

○ Nível

- Mais refinado: micro-análise
 - Por exemplo, RST (Mann e Thompson, 1987)
- Blocos textuais: macro-análise
 - Por exemplo, CARS - *Create a Research Space* (Swales, 1990)

○ Escopo

- Local: contexto local de um segmento textual
 - Entidades
- Global: textual, relacional
 - Sentido global
 - Relacionamento entre partes textuais

NÍVEIS DE ANÁLISE DA LÍNGUA

- Entre os níveis mais abstratos
 - Mas considerado por muitos menos complexo do que níveis inferiores
 - Forte interação (nebulosa!) com semântica e pragmática
 - Influência em todos os níveis de análise



NÍVEIS DE ANÁLISE DA LÍNGUA

- Entre os níveis mais abstratos
 - Envolve a situação de comunicação (Koch e Travaglia, 2002): escritor/falante e leitor/ouvinte



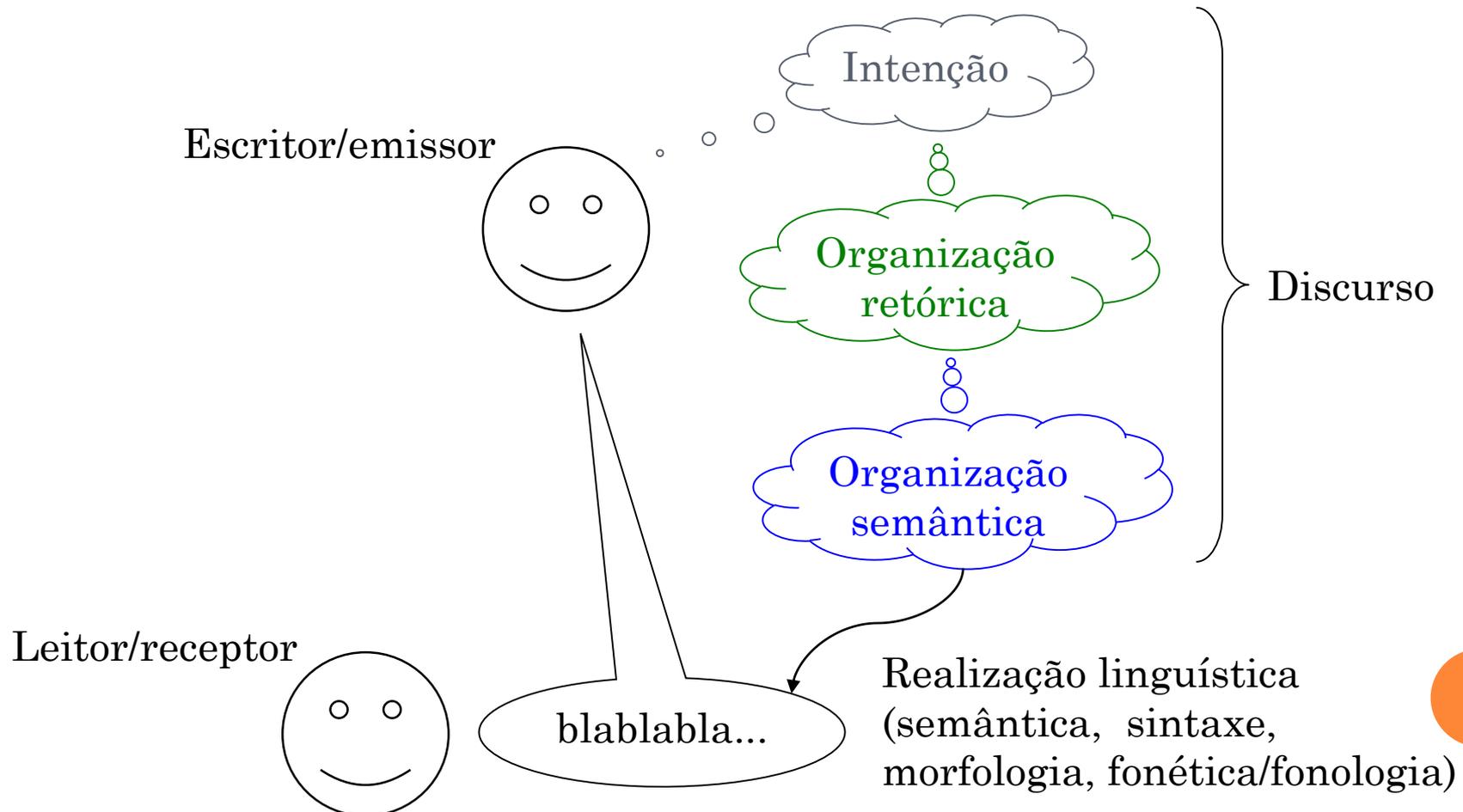
MODELOS DISCURSIVOS

- Grosz e Sidner (1986): *attentions, intentions and the structure of discourse*
 - 2 relações principais entre intenções
- Mann e Thompson (1987): *Rhetorical Structure Theory (RST)*
 - ~23 relações
 - Uma das principais teorias discursivas
- Jordan (1992) e Kehler (2002): relações semânticas, esquemas textuais
 - Jordan: ~70 relações
 - Kehler: 3 relações
 - Relação entre fatos, desprovida de força argumentativa

DISCURSO: UMA VISÃO



- Intenções, retórica e semântica (Moore e Paris, 1993; Korelsky e Kittredge, 1993)



MODELOS DISCURSIVOS

- Hobbs (1979): relações de coerência
- Kamp (1981)
 - *Discourse Representation Theory (DRT)*
- Grosz et al. (1995)
 - *Centering*
- Cristea et al. (1998)
 - *Veins Theory*

MODELOS DISCURSIVOS

- Radev (2000): a popularização multidocumento
 - *Cross-document Structure Theory*
- Asher e Lascarides (2003)
 - *Segmented Discourse Representation Theory (SDRT)*
- Wolf e Gibson (2005): retórica mais livre
- Afantenos et al. (2008)
 - Relações multidocumento sincrônicas e diacrônicas
- Barzilay e Lapata (2008)
 - Grade de entidades e coerência local
- Etc.

DISCURSO

○ **Propósitos** variados

- Descrição textual
- Estudo de fenômenos linguísticos particulares
- Geração de língua natural
- Aplicações computacionais

○ **Origens** variadas

- Linguística
- Computação

DISCURSO E PLN

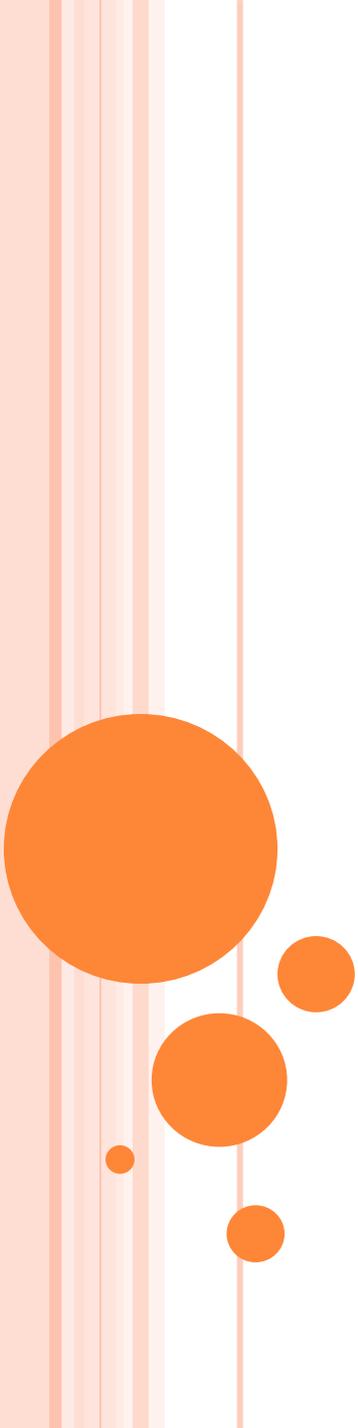
- Duas grandes linhas de trabalho em PLN
 - **Relacionamento** entre partes textuais
 - Questões de coerência, propósito do texto, etc.
 - **Entidades** textuais
 - Correferências, anáforas, etc.

DISCURSO E PLN

○ Muitos usos

- Sumarização automática
- Tradução automática
- Avaliação automática de redações
- Extração de informação
- Auxílio à escrita
- Perguntas e respostas
- Geração textual
- Resolução anafórica
- Etc.

- Como é usado?



*RHETORICAL STRUCTURE
THEORY*

(MANN E THOMPSON, 1987)

RST

- *Como definir retórica?*

RST

○ *Como definir retórica?*

- Parte “palpável” da pragmática (Hovy, 1988)
- Meio pelo qual um texto é organizado para satisfazer um objetivo comunicativo
 - Intenção
- Organização funcional do texto
 - Função de suas partes para o sucesso da comunicação

RST

- Recupera a **estrutura retórica de textos**
 - Descreve **relações** entre o conteúdo de segmentos discursivos em termos funcionais
 - Proposições ou segmentos discursivos: unidades mínimas de significado que compõem um texto
- Premissa
 - Todo texto tem uma estrutura retórica subjacente que permite **recuperar o objetivo comunicativo** que o escritor do texto pretendia atingir ao escrevê-lo

PRINCIPAIS ELEMENTOS

○ Relação retórica

- Relações de significado existentes entre dois ou mais segmentos discursivos
- Conjunto original: 23 relações

○ Cada segmento discursivo representa um núcleo (N) ou satélite (S)

- **Núcleo**, ou unidade nuclear, expressa a informação principal sendo, portanto, mais relevante do que o satélite
- **Satélite** apresenta informação adicional, a qual exerce influência na interpretação do leitor sobre a informação apresentada em N

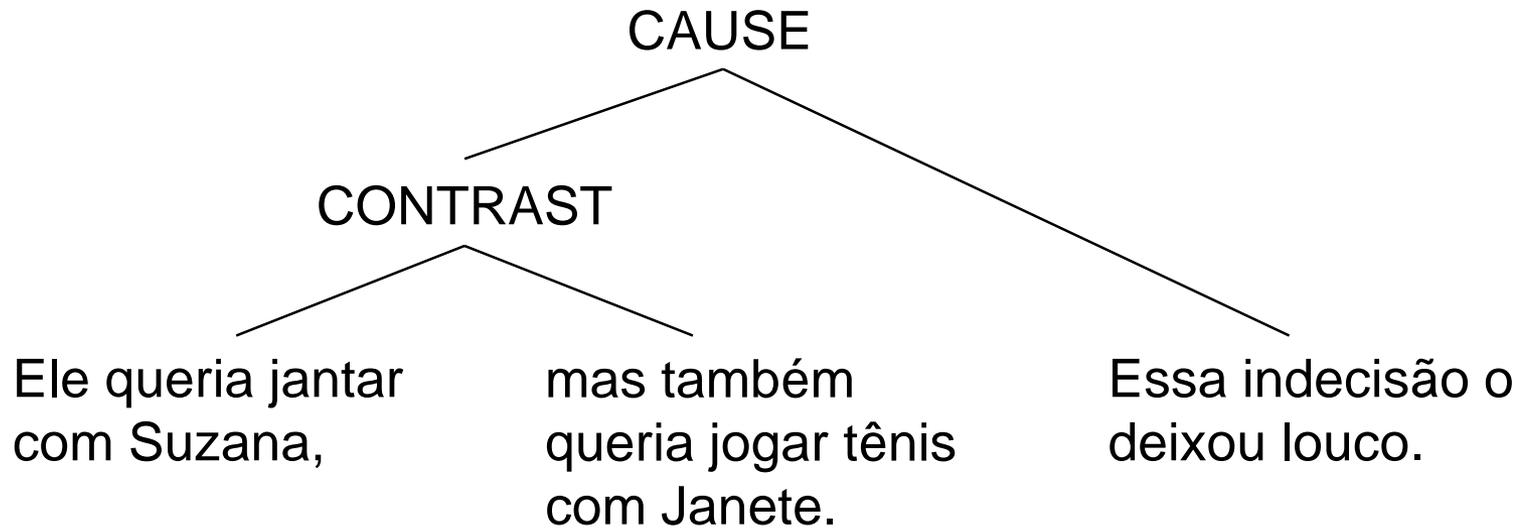
RST: EXEMPLO

Ele queria jantar
com Suzana,

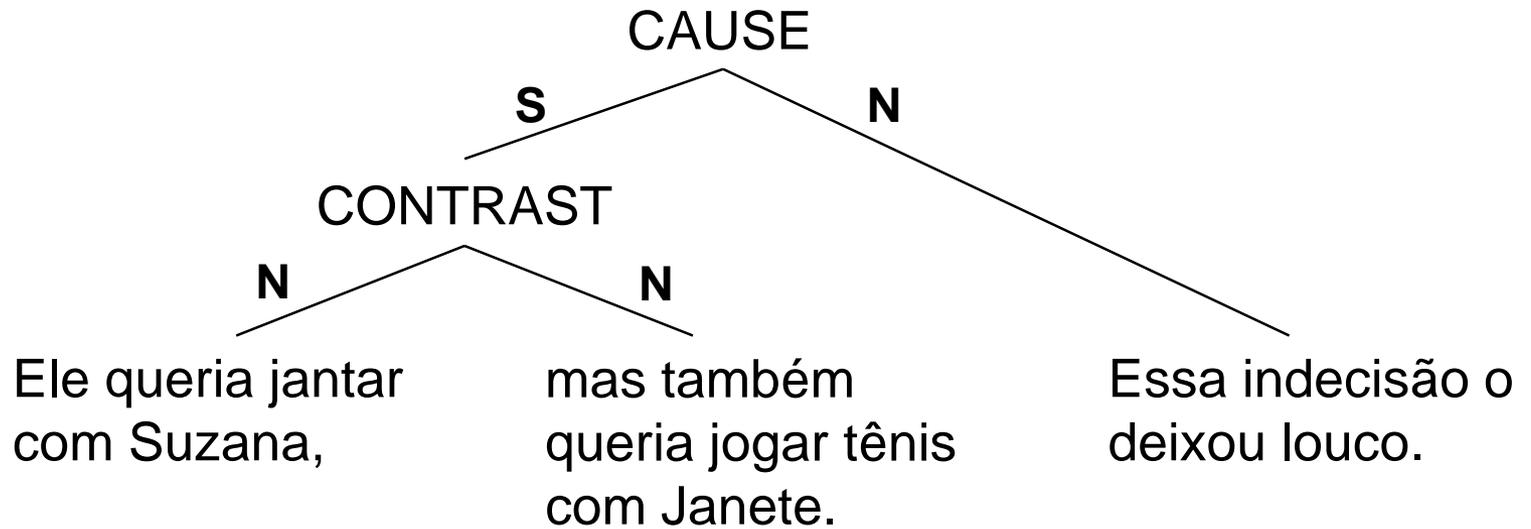
mas também
queria jogar tênis
com Janete.

Essa indecisão o
deixou louco.

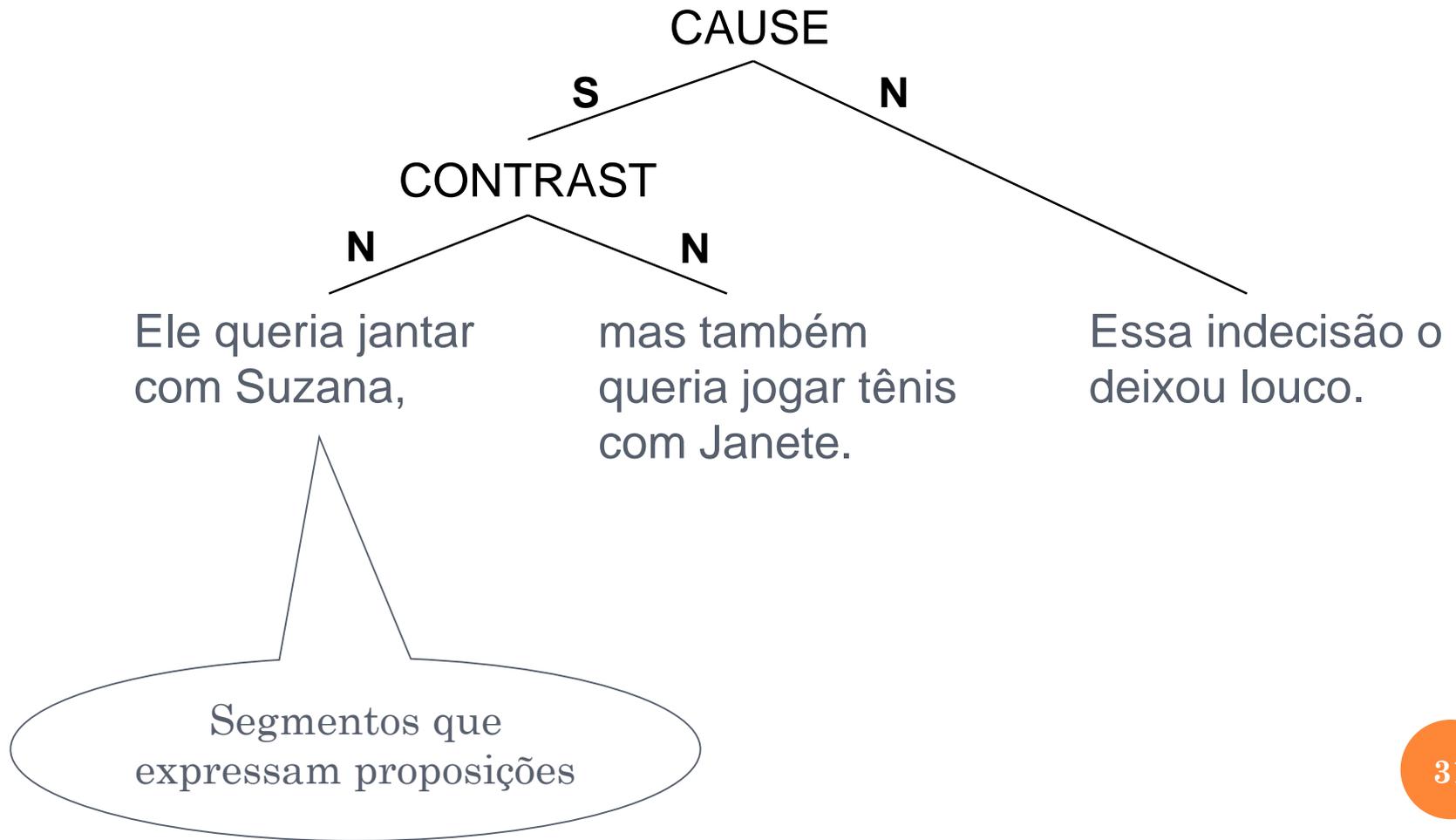
RST: EXEMPLO



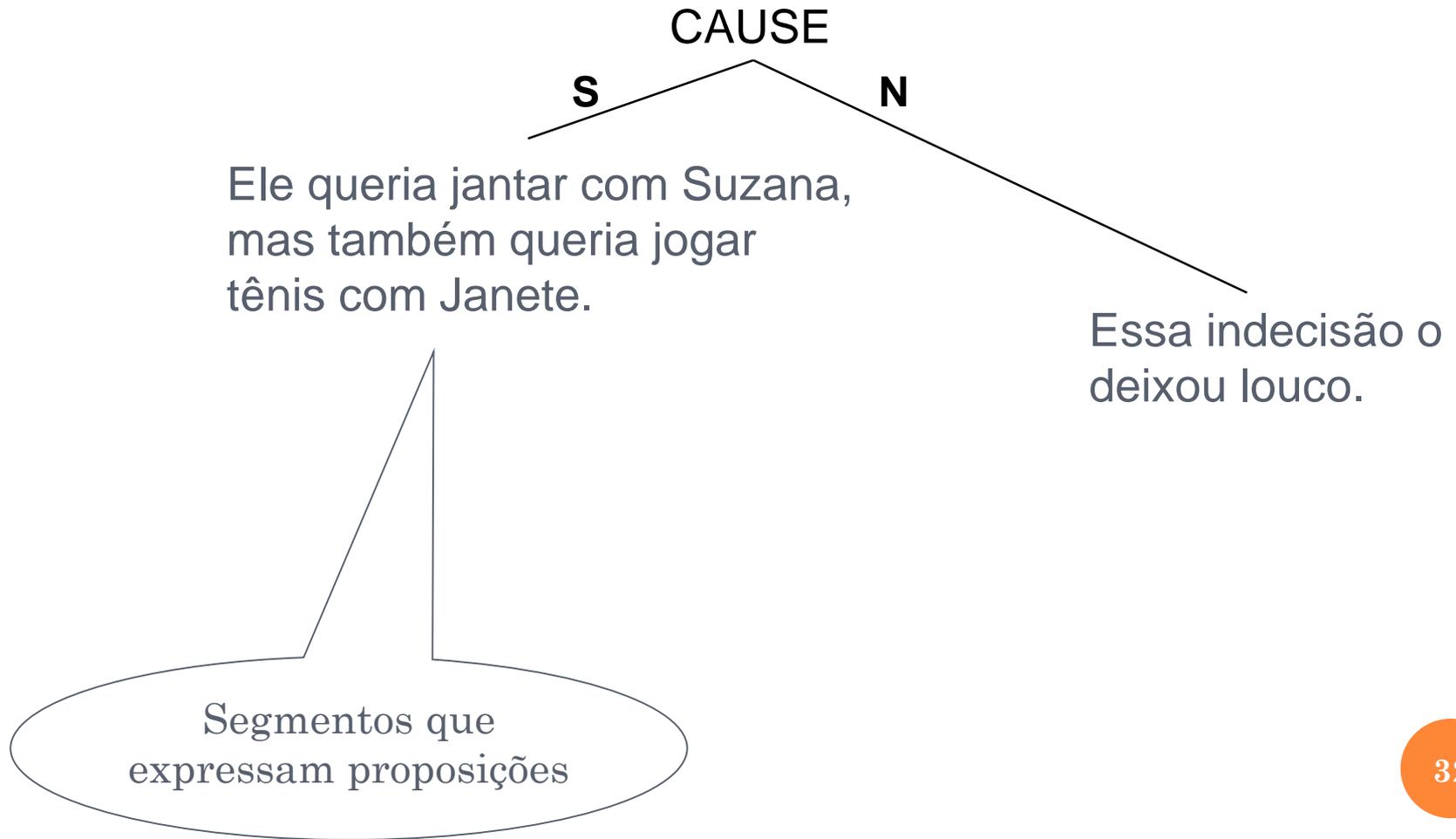
RST: EXEMPLO



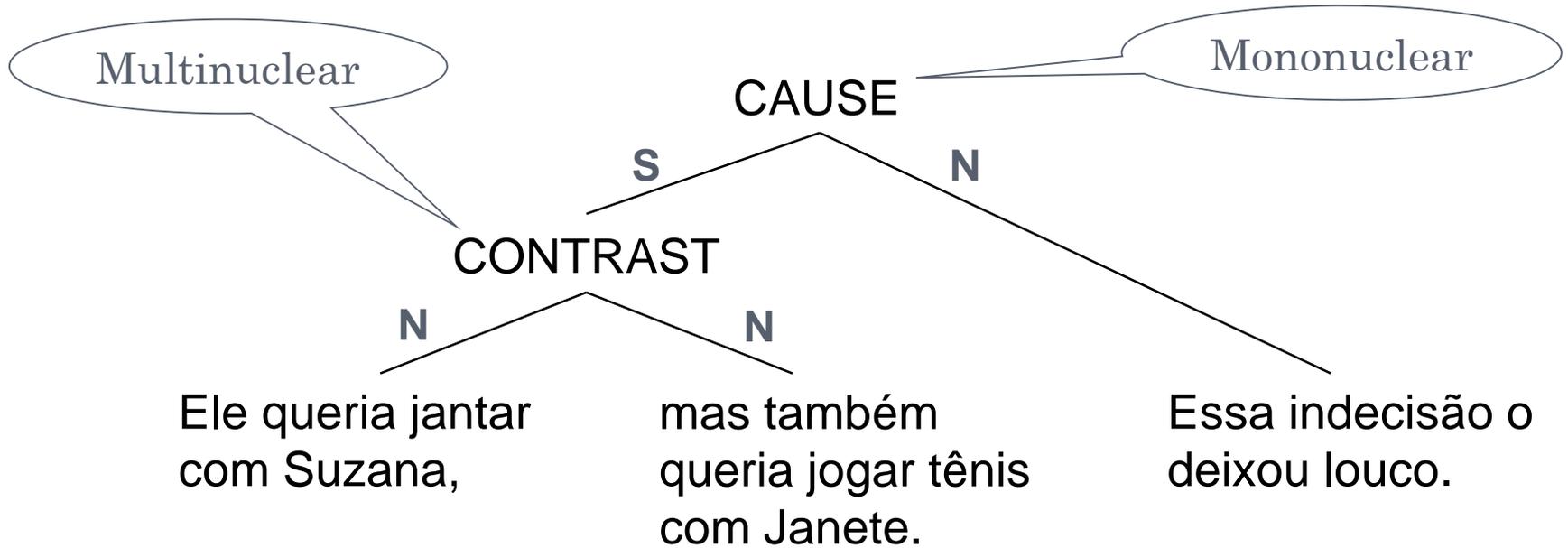
RST: EXEMPLO



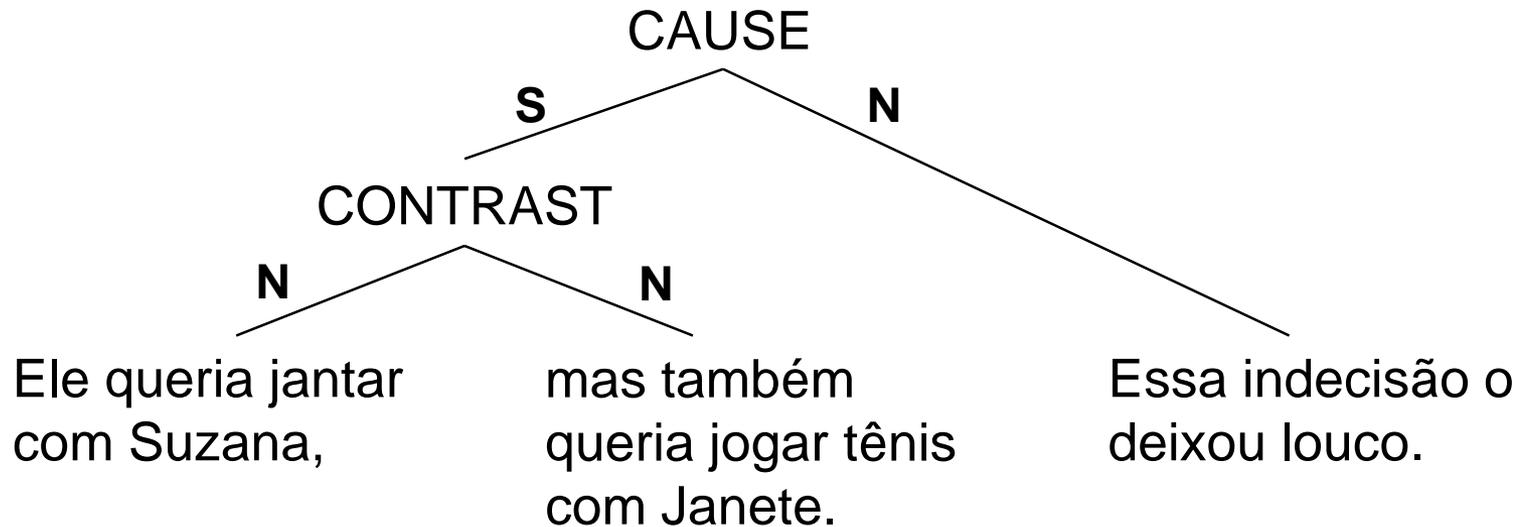
RST: EXEMPLO



RST: EXEMPLO

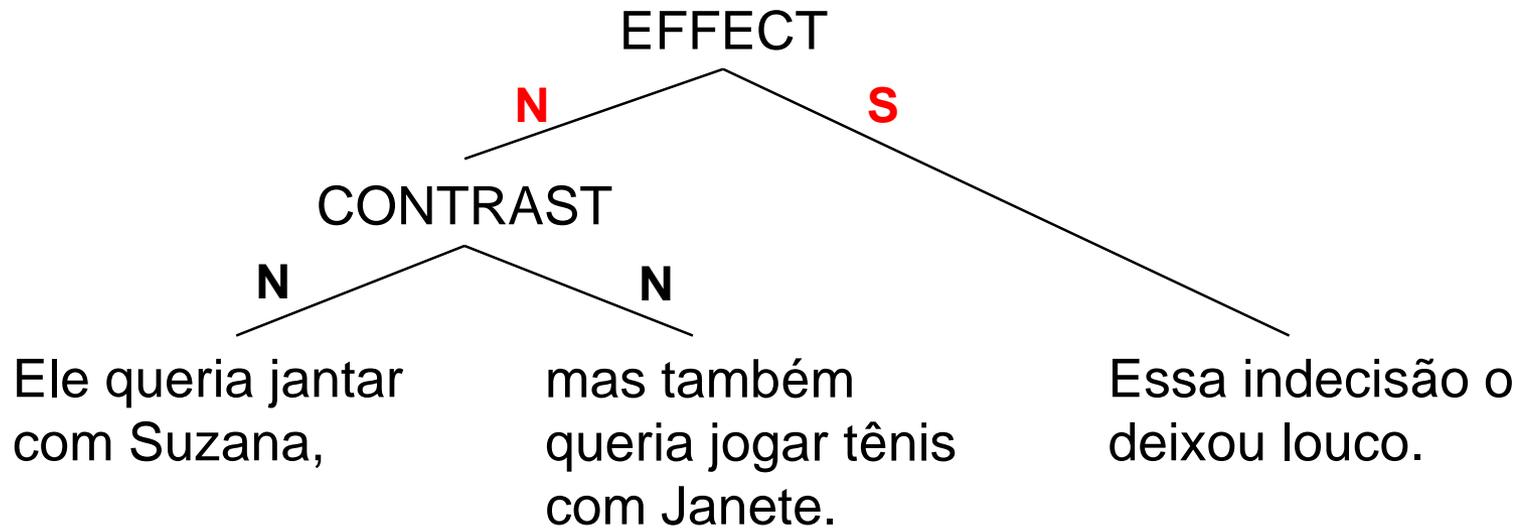


RST: EXEMPLO



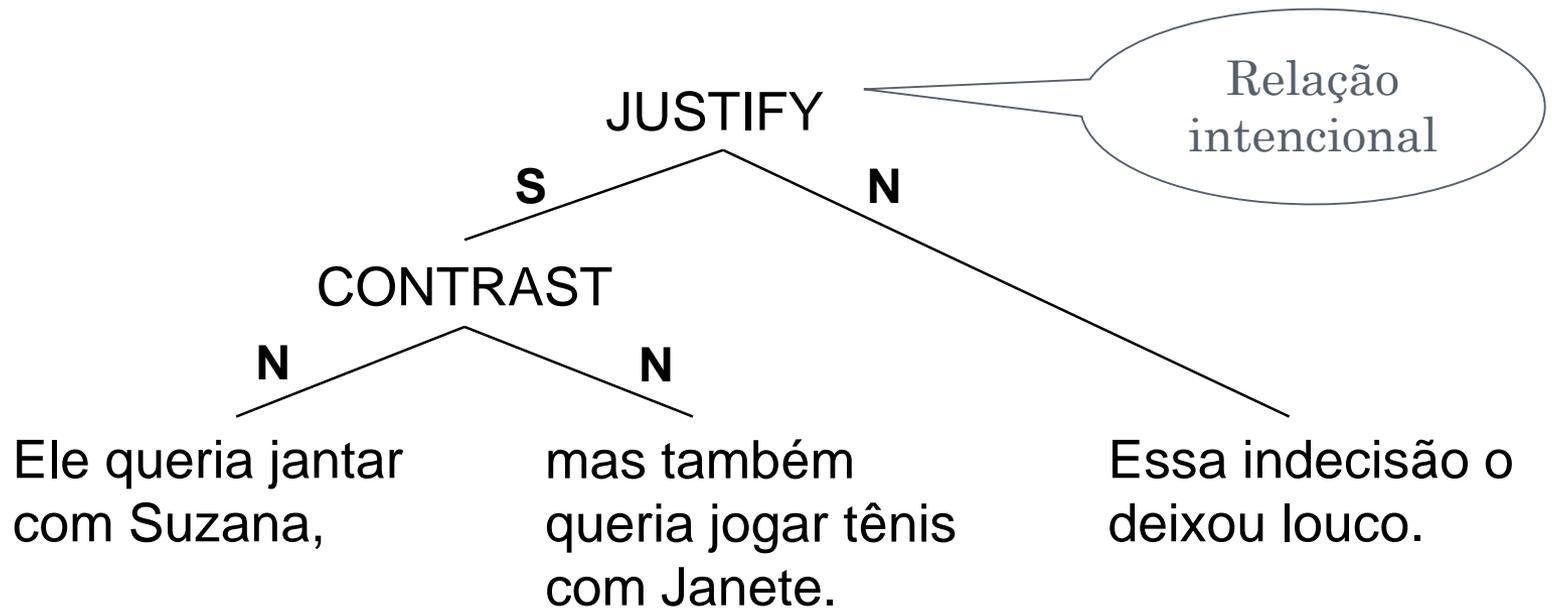
Relações
informativas

RST: EXEMPLO



Relações
informativas

RST: EXEMPLO



RST

- Estilo tradicional de anotação

The screenshot shows the RST Structurer application window. The title bar reads "RST Structurer". The menu bar includes "File", "Display", "Print", "Save", and "Help". Below the menu bar is a toolbar with buttons for "Auto", "Link", "Disconnect", "Modify", "Change Relation", "Change Schema", "Undo", "Reduce", "Enlarge", "Show Text", and "Next Sentence".

The main workspace displays a diagram of a text passage with RST annotations. The text is divided into six numbered segments (1) through (6). The annotations are as follows:

- 1-11**: A vertical line above segment (1).
- 1-5**: A vertical line above segment (1), with a horizontal line extending to the right labeled "elaboration".
- 1-2**: A vertical line above segment (1), with a horizontal line extending to the right labeled "means".
- 3-5**: A vertical line above segment (3), with a horizontal line extending to the right labeled "volitional-result".
- 3-4**: A vertical line above segment (3), with a horizontal line extending to the right labeled "purpose".
- 5**: A vertical line above segment (5), with a horizontal line extending to the right labeled "conc".

The text segments are:

- (1) Este projeto de pesquisa resultará numa ferramenta de monitoramento de dispositivos
- (2) utilizando computação pervasiva e agentes inteligentes.
- (3) A ferramenta permitirá validar as idéias de se integrar o protocolo SNMP e agentes de software inteligentes
- (4) para monitorar dispositivos, com acesso local ou remoto, através do protocolo WAP,
- (5) dando maior flexibilidade e rapidez nas tarefas do administrador.
- (6) Embora já se tenha conhecimento do uso do celular integrado à Internet,

At the bottom of the window, there is a status bar showing the file path: "C:/Arquivos de programas/RSTTool/Rhetalho/Científico/Computação/Conclusao/ArtigoCONGPG_305.txt". Below the status bar is a text area containing the following text:

```
Este projeto de pesquisa resultará numa ferramenta de monitoramento de dispositivos <1>utilizando computação pervasiva e agentes inteligentes.<2> A ferramenta permitirá validar as idéias de se integrar o protocolo SNMP e agentes de software inteligentes <3>para monitorar
```

RELAÇÕES RETÓRICAS

- Organização das relações
 - Relações de Causa
 - *Volitional Cause*
 - *Volitional Result*
 - *Non-Volitional Cause*
 - *Non-Volitional Result*
 - *Purpose*
 - Relações de Evidência e Justificativa
 - *Evidence*
 - *Justify*
 - Relações de Motivação
 - *Enablement*
 - *Motivation*

RELAÇÕES RETÓRICAS

(...CONTINUAÇÃO)

- Relações de Antítese e Concessão
 - *Antithesis*
 - *Concession*
- Relações de Condição
 - *Condition*
 - *Otherwise*
- Relações de Interpretação e Avaliação
 - *Interpretation*
 - *Evaluation*
- Relações de Reiteração e Sumário
 - *Restatement*
 - *Summary*

RELAÇÕES RETÓRICAS (...CONTINUAÇÃO)

- Outras
 - *Background*
 - *Circumstance*
 - *Contrast*
 - *Elaboration*
 - *Sequence*
 - *Solutionhood*
 - *Joint*

RST

○ Relações

- **Intencionais:** alteram a inclinação do leitor para algo
- **Informativas:** informam o leitor sobre algo
 - Podem co-ocorrer? É natural que isso aconteça? Ou é um problema da teoria?

RST

- Exemplo de **ambiguidade**

George Bush supports big business.

He's sure to veto House Bill 1711.

- Relação de **evidência** (intencional) ou de **causa** (informativa)?
 - O que fazer?

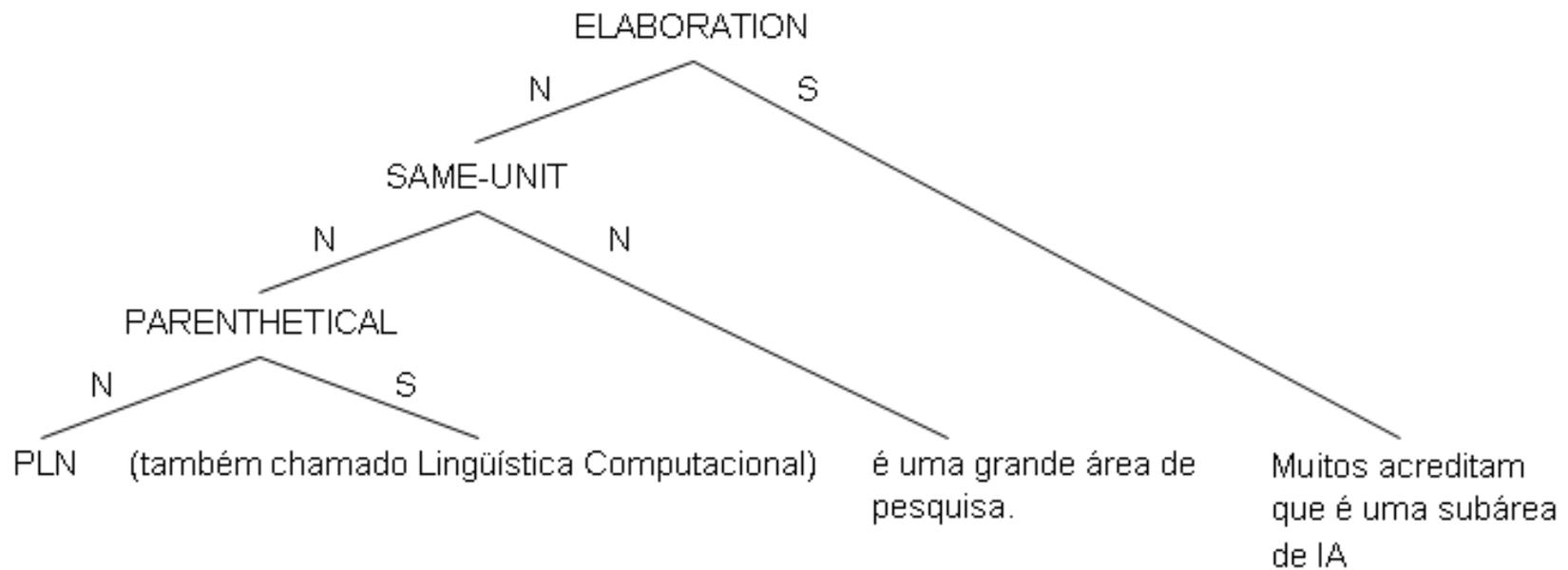
RST

○ Relações

- **Estruturais:** sem significado, apenas mantêm a integridade da estrutura textual
 - Um terceiro tipo, introduzido posteriormente

RST

- Exemplos de **relações estruturais**



DETERMINAÇÃO DE UMA RELAÇÃO ENTRE DUAS PROPOSIÇÕES

- Definição com 4 campos
 - Restrições sobre o núcleo (N)
 - Restrições sobre o satélite (S)
 - Restrições sobre o núcleo e o satélite
 - Efeito
- Verificação das restrições para as proposições envolvidas
 - Escrita ou leitura de um texto

DEFINIÇÃO DE RELAÇÕES

Relação: JUSTIFY

Restrições sobre N: não há

Restrições sobre S: não há

Restrições sobre N+S: a compreensão de S pelo leitor aumenta sua prontidão para aceitar o direito do escritor de apresentar N

Efeito: a prontidão do leitor para aceitar o direito do escritor de apresentar N aumenta

DEFINIÇÃO DE RELAÇÕES

Relação: CAUSE

Restrições sobre N: apresenta uma ação volitiva ou uma situação que poderia surgir de uma ação volitiva

Restrições sobre S: não há

Restrições sobre N+S: S apresenta uma situação que pode ter acarretado o fato do agente da ação volitiva em N ter realizado a ação; sem S, o leitor poderia não reconhecer a motivação da ação; N é mais central para a satisfação do objetivo do escritor do que S

Efeito: o leitor reconhece que a situação apresentada em S como a causa da ação apresentada em N

DEFINIÇÃO DE RELAÇÕES

Relação: CONCESSION

Restrições sobre N: o escritor julga N válido

Restrições sobre S: o escritor não afirma que S pode não ser válido

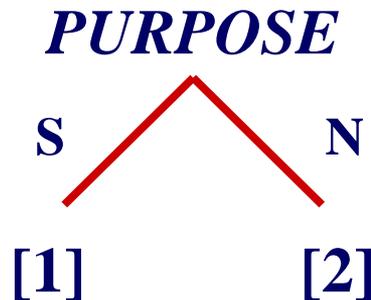
Restrições sobre N+S: o escritor mostra uma incompatibilidade aparente ou em potencial entre N e S; o reconhecimento da compatibilidade entre N e S melhora a aceitação de N pelo leitor

Efeito: o leitor aceita melhor N

TIPOS DE RELAÇÕES

- Relações Mononucleares X Relações Multinucleares
- Exemplo (Mononuclear): **PURPOSE**

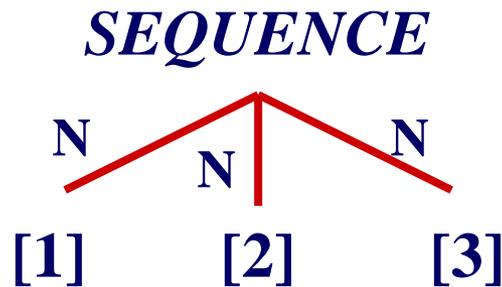
[1] A fim de passar no vestibular de medicina, [2] Carla se dedicou aos estudos durante o ano todo.



TIPOS DE RELAÇÕES (...CONTINUAÇÃO)

- Exemplo (Multinuclear): **SEQUENCE**

[1] A menina caiu da bicicleta, [2] quebrou o braço [3] e foi levada para o hospital.



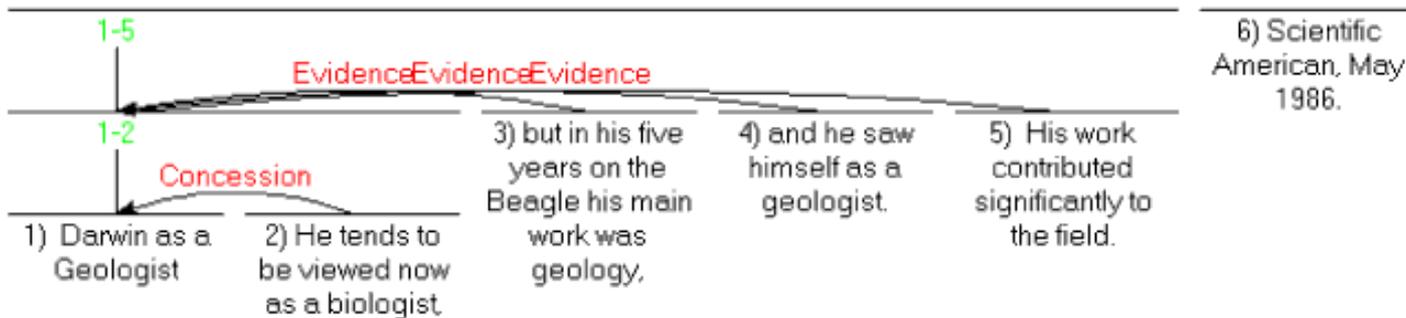
ALGUMAS LIMITAÇÕES

- **Subjetividade** da análise e ambiguidade
 - Determinação dos **segmentos**
 - Orações, sentenças ou parágrafos?
 - Fragmentos com significado?
 - Determinação da **nuclearidade**
 - NS, SN ou NN?
 - Escolha das **relações**
 - CAUSE, EFFECT ou JUSTIFY?
 - CONCESSION, CONTRAST, ANTITHESIS ou OTHERWISE?
 - Conjunto de relações?
 - Forma de **representação** da análise
 - Árvore ou grafo

ANÁLISE RETÓRICA

○ Restrições na análise

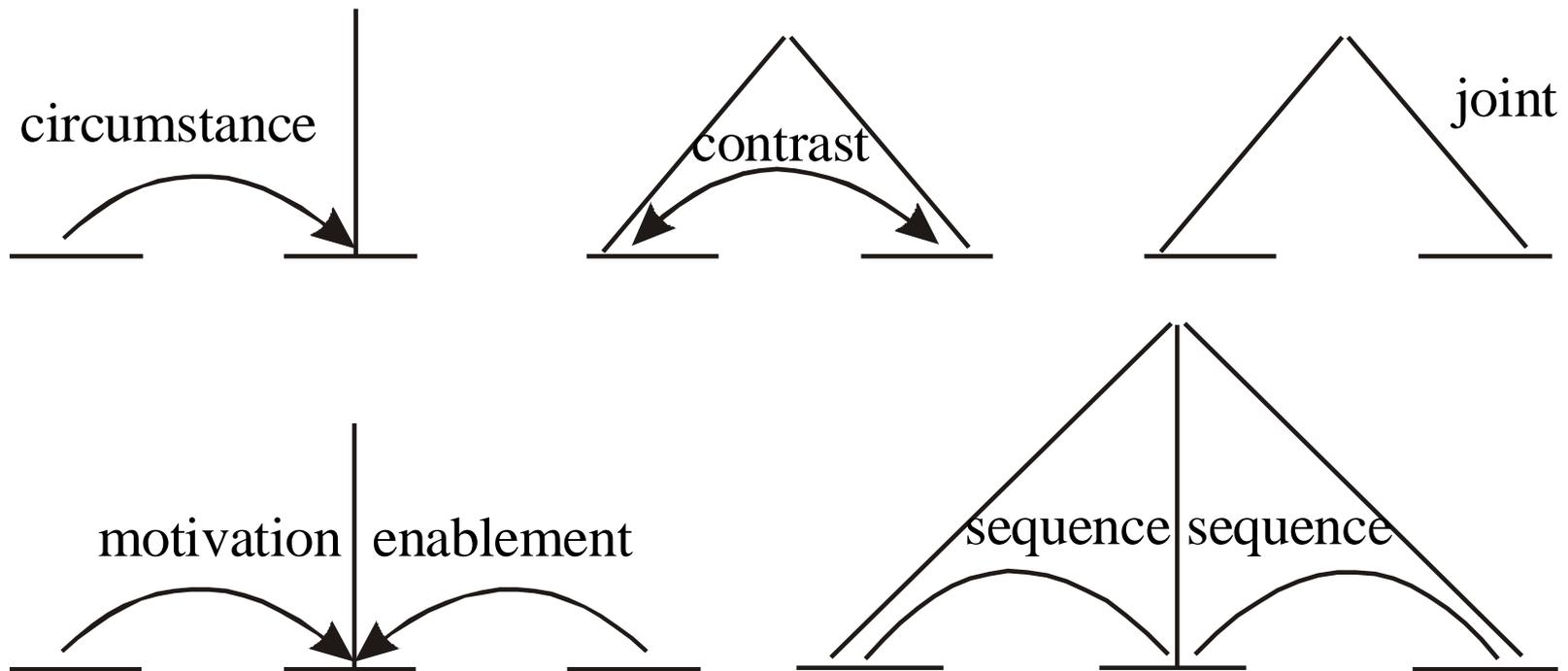
- Adjacência de segmentos, com algumas exceções



- Não compartilhamento de **satélites**
- Não sobreposição de segmentos
- Relações entre **núcleos** de segmentos maiores

ANÁLISE RETÓRICA

- Esquemas possíveis de análise



ALGUMAS REGRAS PARA SEGMENTAÇÃO (CARLSON E MARCU, 2001)

○ Segmentar

- Orações principais
- Orações sinalizadas por marcadores discursivos como *Porque, Apesar de, Conforme, Segundo, Em consequência de*, entre outros
- Orações subordinadas introduzidas por marcadores discursivos
- Orações coordenadas
- Orações relativas, apositivas e parênteses como segmentos encaixados

ALGUMAS REGRAS PARA SEGMENTAÇÃO (...CONTINUAÇÃO)

○ Não segmentar

- Orações complementares, exceto quando introduzirem complemento de um verbo de atribuição
 - Por exemplo: *[1] A companhia disse que [2] fechará a fábrica*
- Orações subordinadas substantivas e objetivas

CONSTRUÇÃO DA ESTRUTURA RETÓRICA

- Possível procedimento para análise: **análise incremental**
 - Relacionamento de todos os segmentos discursivos que compõem uma sentença, dois a dois
 - Relacionamento de todas as sentenças que compõem um parágrafo, duas a duas
 - Relacionamento de todos os parágrafos que compõem o texto, dois a dois

EXEMPLO DE ANÁLISE RETÓRICA

○ Texto

Uma explosão neste domingo em Pristina, capital do Kosovo, deixou três crianças feridas e causou estragos no prédio onde fica a sede do partido de oposição ORA, informaram policiais e testemunhas.

Um porta-voz da polícia do país disse que as crianças, que vivem no andar inferior ao do escritório do partido, sofreram ferimentos leves por vidros estilhaçados.

"Foram levadas a um hospital, mas já foram liberadas", disse Refki Morina. Segundo ele, a explosão causou grandes estragos no prédio.

O local foi isolado e membros da força de paz liderada pela Otan (Organização do Tratado do Atlântico Norte) se dirigiram ao local.

SEGMENTAÇÃO (ORACIONAL)

PARÁGRAFO 1: [1] Uma explosão neste domingo em Pristina, capital do Kosovo, deixou três crianças feridas e causou estragos no prédio [2] onde fica a sede do partido de oposição ORA, [3] informaram policiais e testemunhas.

PARÁGRAFO 2: [4] Um porta-voz da polícia do país disse que [5] as crianças, [6] que vivem no andar inferior ao do escritório do partido, [7] sofreram ferimentos leves por vidros estilhaçados.

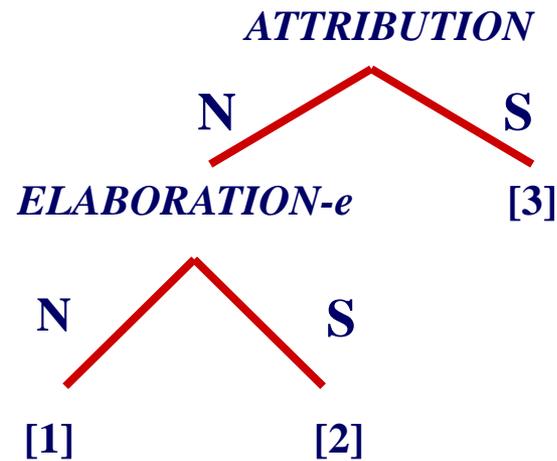
PARÁGRAFO 3: [8] "Foram levadas a um hospital, mas já foram liberadas", [9] disse Refki Morina. [10] Segundo ele, a explosão causou grandes estragos no prédio.

PARÁGRAFO 4: [11] O local foi isolado, [12] e membros da força de paz liderada pela Otan (Organização do Tratado do Atlântico Norte) se dirigiram ao local.

CONSTRUÇÃO DA ESTRUTURA RETÓRICA (...CONTINUAÇÃO)

○ PARÁGRAFO 1: **uma sentença**

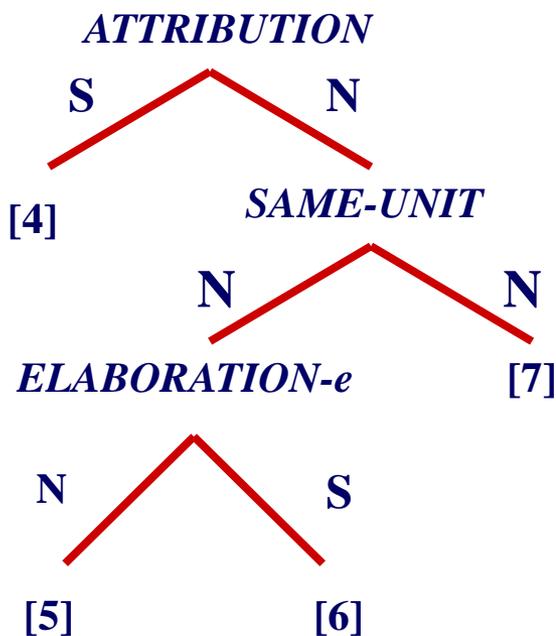
[1] Uma explosão neste domingo em Pristina, capital do Kosovo, deixou três crianças feridas e causou estragos no prédio [2] onde fica a sede do partido de oposição ORA, [3] informaram policiais e testemunhas.



CONSTRUÇÃO DA ESTRUTURA RETÓRICA (...CONTINUAÇÃO)

○ PARÁGRAFO 2: uma sentença

[4] Um porta-voz da polícia do país disse que [5] as crianças, [6] que vivem no andar inferior ao do escritório do partido, [7] sofreram ferimentos leves por vidros estilhaçados.



CONSTRUÇÃO DA ESTRUTURA RETÓRICA (...CONTINUAÇÃO)

○ PARÁGRAFO 3: duas sentenças

• Sentença 1

[8] "Foram levadas a um hospital, mas já foram liberadas", [9] disse Refki Morina.



• Sentença 2

[10] Segundo ele, a explosão causou grandes estragos no prédio.

CONSTRUÇÃO DA ESTRUTURA RETÓRICA (...CONTINUAÇÃO)

- PARÁGRAFO 3: duas sentenças

- Relacionamento: sentença 1 e sentença 2



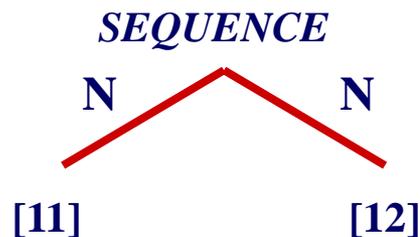
[10]

Relacionar ao
parágrafo 3 ou 4?

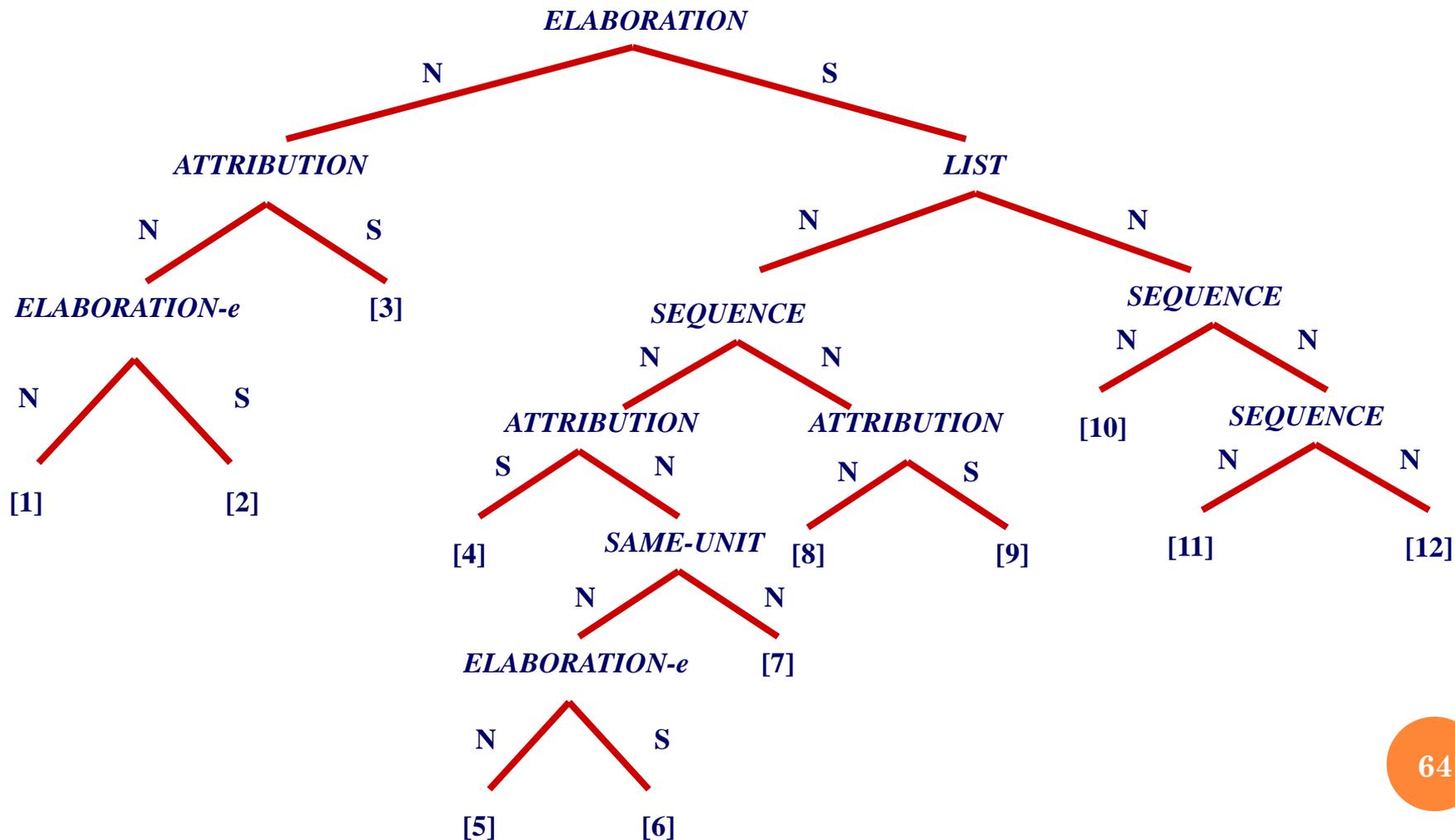
CONSTRUÇÃO DA ESTRUTURA RETÓRICA (...CONTINUAÇÃO)

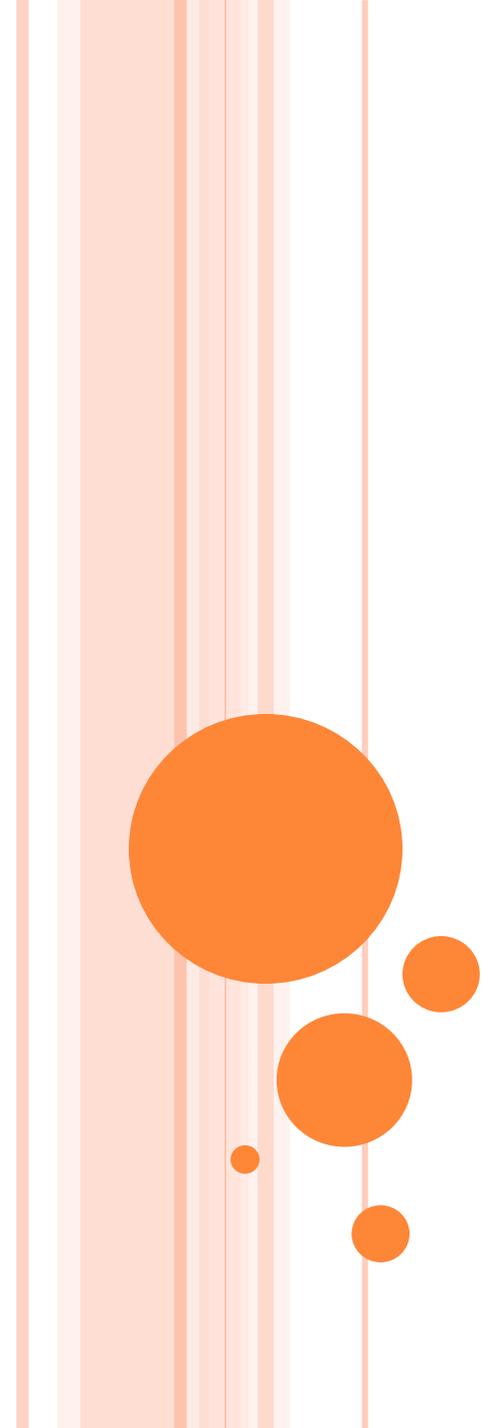
- **PARÁGRAFO 4: uma sentença**

[11] O local foi isolado, [12] e membros da força de paz liderada pela Otan (Organização do Tratado do Atlântico Norte) se dirigiram ao local.



ESTRUTURA RETÓRICA DO TEXTO





EXERCÍCIO

EXERCÍCIO

- Parte 1 – análise de um conjunto de relações

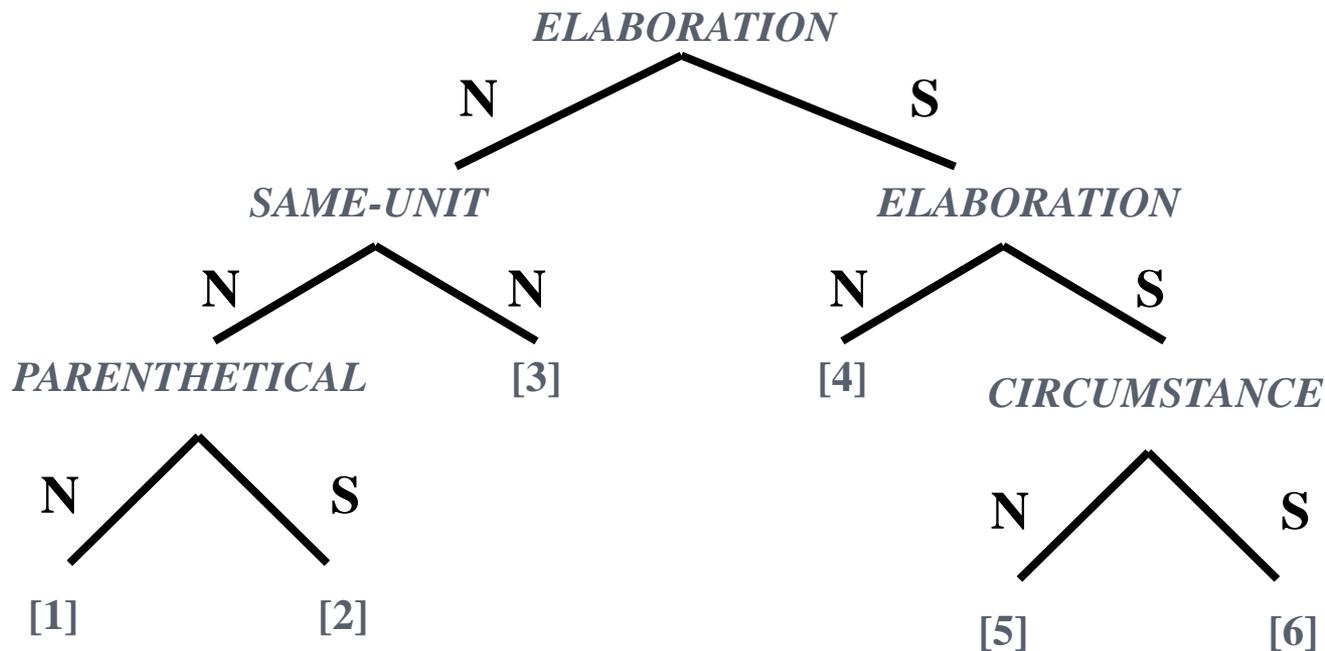
ANTITHESIS	CONTRAST	JUSTIFY	PURPOSE
ATTRIBUTION	ELABORATION	LIST	RESTATEMENT
BACKGROUND	ENABLEMENT	MEANS	SAME-UNIT
CIRCUMSTANCE	EVALUATION	MOTIVATION	SEQUENCE
COMPARISON	EVIDENCE	NON-VOL-CAUSE	SOLUTIONHOOD
CONCESSION	EXPLANATION	NON-VOL-RESULT	SUMMARY
CONCLUSION	INTERPRETATION	OTHERWISE	VOL-CAUSE
CONDITION	JOINT	PARENTHETICAL	VOL-RESULT

EXERCÍCIO

- Parte 2 – análise do texto abaixo

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

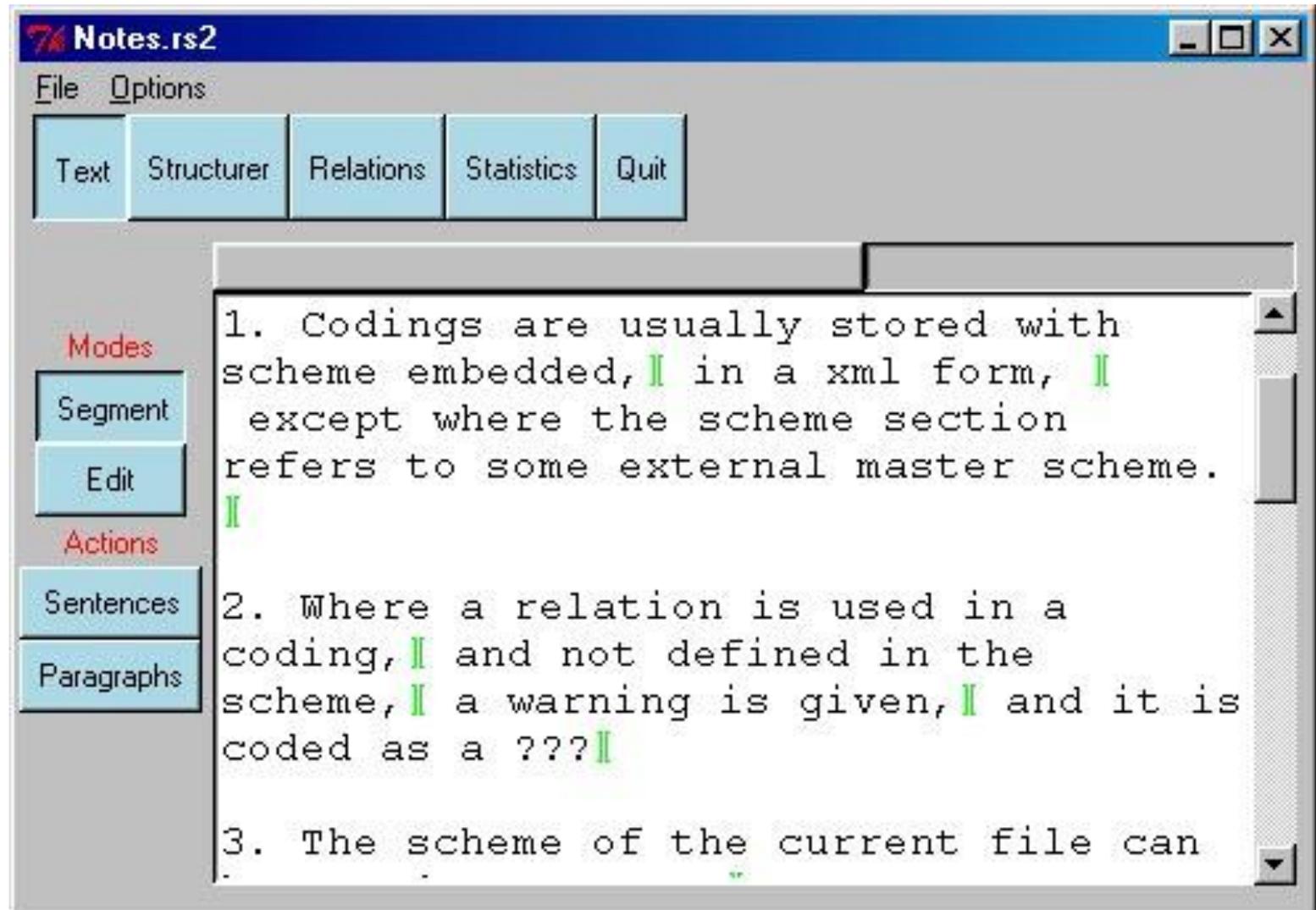
[1] A empresa Produtos Pirata Indústria e Comércio Ltda., de Contagem [2] (na região metropolitana de Belo Horizonte), [3] deverá registrar este ano um crescimento de produtividade nas suas áreas comercial e industrial de 11% e 17%, respectivamente. [4] Os ganhos são atribuídos pela diretoria da fábrica 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.



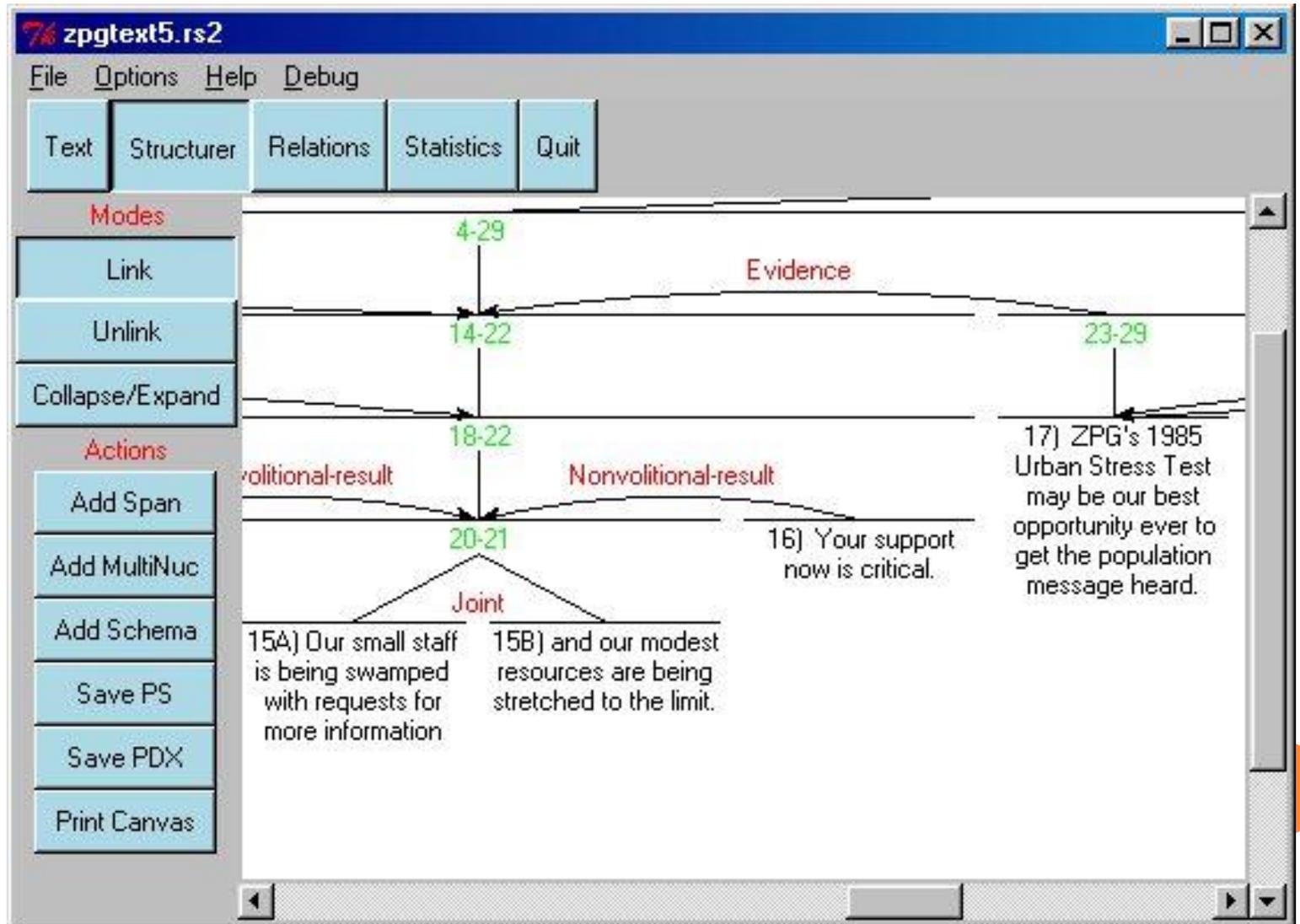
FERRAMENTAS DE AUXÍLIO

- *RSTTool* (O' Donnell, 1997, 2000)
- *RST Annotation Tool*: extensão da *RSTTool* (Marcu, 1997)
- *rstWeb* (Zeldes, 2016)
- Etc.

RSTTOOL



RSTTOOL



PARSING DISCURSIVO

- Como fazer?

PARSING DISCURSIVO

- Tarefas envolvidas
 - **Determinação dos segmentos**
 - Deve ser sistemática
 - Uso de indicações lexicais, morfossintáticas, sintáticas, aprendizado de máquina
 - **Determinação das relações e nuclearidade**
 - Marcadores discursivos, palavras e frases indicativas
 - Relações lexicais
 - Sintaxe
 - Aprendizado de máquina
 - **Construção de árvores RST válidas**
 - Critérios da teoria

APLICAÇÕES

- Como a RST pode auxiliar?
 - Sumarização automática?
 - Tradução automática?
 - Correção automática de redações?
 - Auxílio à escrita?
 - Perguntas e respostas?
 - Etc.