

Automatização de um Método de Avaliação de Estruturas Retóricas

Erick Galani Maziero (erickgm@grad.icmc.usp.br)
Thiago Alexandre Salgueiro Pardo (tasparado@icmc.usp.br)

Núcleo Interinstitucional de Linguística Computacional (NILC)
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo
Av. Trabalhador São-carlense, 400 - Centro
Caixa Postal: 668 - CEP: 13560-970 - São Carlos - SP

Resumo

Nesse artigo, aborda-se a questão da avaliação automática de estruturas retóricas produzidas segundo os princípios da RST (*Rhetorical Structure Theory*). O método automatizado apresentado é um método proposto na literatura e visa, principalmente, à comparação de estruturas retóricas produzidas automaticamente com estruturas de referência produzidas por humanos especialistas em RST. Tal comparação produz resultados em termos de cobertura e precisão das estruturas automáticas, levando em consideração os aspectos envolvidos em uma estrutura retórica, isto é, identificação de segmentos discursivos, suas importâncias em termos de nuclearidade e as relações retóricas estabelecidas entre eles.

1. Introdução

Uma das mais expressivas e desafiadoras tarefas no mundo do Processamento de Línguas Naturais (PLN) é o tratamento pragmático-discursivo, nível em que o texto veicula uma mensagem e incorpora as intenções do autor do mesmo. As partes do texto não agem independentemente para atingir o objetivo do texto, mas se inter-relacionam numa estrutura bem definida, chamada de estrutura retórica ou discursiva. A RST (*Rethorical Structure Theory*) (Mann e Thompson, 1987) é uma das teorias discursivas mais difundidas e visa justamente representar esse nível de análise e descrição textual.

A ADA (Análise Discursiva Automática) é um processo de PLN que, dado um texto, produz sua estrutura discursiva de forma automática. Na RST, essa estrutura é representada tradicionalmente na forma de uma árvore, em que suas folhas são as partes constituintes do texto, ou seja, os segmentos discursivos, sendo que seus conteúdos informativos são relacionados por relações retóricas. A cada segmento é atribuído o caráter de núcleo ou satélite, dependendo de sua importância: o segmento mais importante da relação é núcleo e o segmento complementar ao núcleo é tido como satélite. Também é possível que existam somente segmentos nucleares, se os segmentos forem considerados igualmente importantes.

A avaliação da qualidade de uma estrutura retórica, produzida automaticamente ou por humanos, é um grande desafio diante de toda a subjetividade envolvida na análise RST. A subjetividade provém do fato de que cada leitor de um texto pode ter impressões diferentes e produzir estruturas retóricas diversas que tenham variações no que se considera segmento discursivo, nas relações estabelecidas e nos níveis da árvore retórica em que acontecem, e até mesmo na importância dos segmentos. Portanto, é difícil quantificar o quão boa uma estrutura retórica é, e isso é especialmente importante

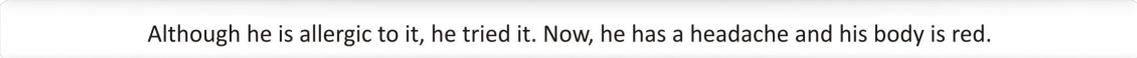
em ADA, já que vários sistemas que automatizam a tarefa têm sido propostos e desenvolvidos. É importante que se tenha um método de avaliação consistente e reproduzível para que se saiba qual a melhor abordagem a se seguir e o real estado da arte.

Marcu (2000) propôs um método de avaliação de estruturas retóricas produzidas automaticamente. Seu método compara uma estrutura automática com uma estrutura dita de referência, ou seja, produzida por um humano especialista em RST, e quantifica a cobertura e precisão da árvore para cada um de seus aspectos: identificação de segmentos, das relações retóricas e a nuclearidade dos segmentos. Apesar de o método proposto ser voltado para a avaliação de sistemas de ADA, ele também poderia ser usado, por exemplo, para comparar duas estruturas humanas (o quão diferente elas são) ou duas estruturas automáticas. A grande vantagem do método de Marcu reside no fato de ele ser sistemático e determinístico, sendo passível, portanto, de ser automatizado, tornando o processo de avaliação rápido e reproduzível.

Neste artigo, apresenta-se o método de Marcu e sua implementação em uma ferramenta com interface web de uso público. Na Seção 2, demonstra-se o método e, na Seção 3, a interface web construída. Por fim, algumas considerações finais são feitas na Seção 4.

2. Método de Avaliação

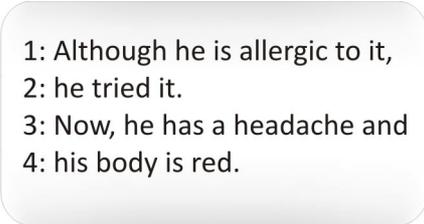
Para apresentar o método utilizado na ferramenta de avaliação automática de estruturas retóricas, será utilizada a análise de um fragmento de texto, apresentado na Figura 1.



Although he is allergic to it, he tried it. Now, he has a headache and his body is red.

Figura 1. Texto-exemplo para análise discursiva

Para produzir sua estrutura retórica, tem-se que primeiramente segmentar o texto em porções menores; essas partes são categorizadas em núcleo ou satélite de acordo com a relação que se escolherá para os pares de segmentos. Para que se tenham segmentos que possam ser relacionados, os mesmos devem ter algum sentido isoladamente; desta forma, cada segmento deve conter uma idéia ou conceito básico do texto (ou seja, cada segmento deve expressar uma proposição). Considere a segmentação dada na Figura 2.



1: Although he is allergic to it,
2: he tried it.
3: Now, he has a headache and
4: his body is red.

Figura 2. Segmentação do texto-exemplo

O próximo passo da análise retórica é relacionar os segmentos para produzir a estrutura retórica. Por exemplo, o segmento 1 relaciona-se com o segmento 2 pela relação *concession*.

A ordem de relacionamento dos segmentos, comumente, é relacionar os segmentos dentro de uma mesma sentença e, após isso, relacionar as sentenças dentro de um mesmo parágrafo e, por fim, relacionar os parágrafos entre si, sempre dois a dois. Tal estratégia é proposta e implementada por Pardo e Nunes (2008).

A atribuição do caráter núcleo ou satélite é outro passo que depende, em muitos casos, da interpretação do analisador. Na relação acima estabelecida, temos o segmento 2 como sendo o segmento principal (núcleo - N) da sua relação com o segmento 1 (satélite - S).

Feita a análise retórica, obtemos a estrutura ilustrada na forma de uma árvore na Figura 3. Essa estrutura será a estrutura utilizada para exemplificar todo o método de avaliação apresentado.

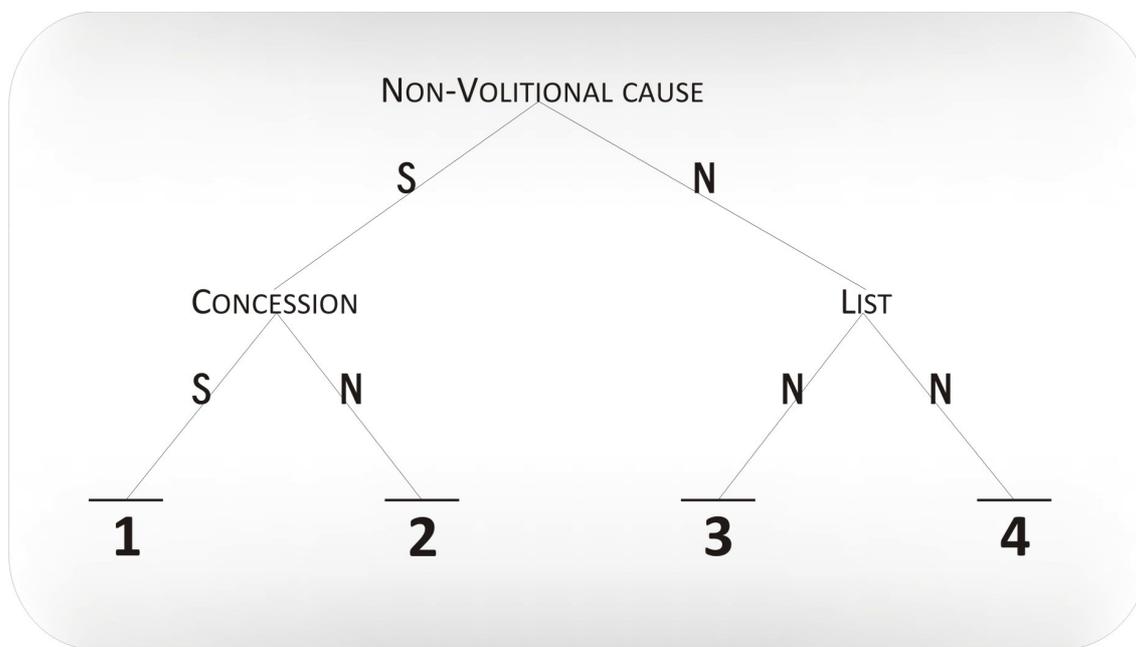


Figura 3. Estrutura retórica do texto-exemplo

Na avaliação de uma estrutura retórica produzida por algum sistema de análise automática, utiliza-se uma estrutura retórica de referência feita manualmente por um especialista da área. As medidas de cobertura e precisão são utilizadas para expressar o grau de similaridade entre a estrutura produzida automaticamente e a produzida manualmente, que é tida como ideal. Optou-se por estas duas medidas devido a sua grande utilização no estudo de desempenho de várias ferramentas de PLN, principalmente analisadores sintáticos, que também produzem estruturas na forma de árvore, similares às retóricas.

Apesar do emprego de um método bem definido, a subjetividade da interpretação deve ser levada em conta na avaliação, pois um mesmo texto pode ser corretamente interpretado por dois analistas, mas com estruturas produzidas que difiram inteira ou parcialmente. Detalhes sem importância, como um sinal de pontuação ou uma palavra considerada *stopword* que esteja em segmentos diferentes nas duas análises, não pode diminuir o desempenho relativo de um método automático. Neste ponto, o método de avaliação implementado difere do método proposto por Marcu (2000). Pode-se dizer, portanto, que o método descrito neste artigo é um pouco mais flexível do que o de Marcu.

A medida cobertura reflete o número de elementos corretamente identificados pelo analisador automático em relação ao número identificado na estrutura construída manualmente. Precisão é o número correto de elementos identificados pelo analisador automático em relação ao total identificado nesta mesma estrutura.

No método de avaliação, aplicam-se as medidas de cobertura e precisão sobre cada parte da estrutura retórica apresentada na Tabela 1.

Tabela 1. Elementos verificados na avaliação

Itens que são avaliados
Segmentos textuais
<i>Spans</i>
Relação entre os segmentos textuais
Nuclearidade dos segmentos textuais

Segmentos textuais são as partes disjuntas do texto sob análise. Todo segmento textual, simples ou composto (pela junção por alguma relação retórica em algum nó interno na árvore), é considerado um *span*.

A avaliação é realizada em etapas bem definidas. Primeiramente, para comparar a similaridade dos segmentos textuais, retiram-se sinais de pontuação e *stopwords*, numerando cada palavra resultante seqüencialmente e colocando um rótulo para cada segmento como sendo a junção da numeração de sua palavra inicial e final. Retirando-se as *stopwords* e pontuações do texto-exemplo da Figura 1, obtém-se o texto da Figura 4, em que cada palavra resultante é numerada seqüencialmente.

allergic[1] tried[2] headache[3] body[4] red[5]

Figura 4. Texto-exemplo após exclusão de stopwords e numeração

Como próximo passo, de acordo com a estrutura retórica, rotula-se cada nó da árvore com os números das palavras inicial e final de cada nó. Assim, o segmento 4 recebe o rótulo referente aos números das palavras inicial e final, ou seja, 4..5, referente aos números das palavras *headache* e *red*, respectivamente. Quando se deseja rotular os nós internos da árvore retórica, utiliza-se o mesmo procedimento, utilizando os números das palavras inicial e final dos segmentos contidos no nó interno da estrutura.

Após a rotulação dos nós da estrutura, gera-se uma tabela contendo, em cada linha, um nó da árvore, seja um nó folha ou nó interno (união de segmentos textuais). Para cada nó, coloca-se sua nuclearidade. Assume-se, neste método, que as árvores retóricas são binárias, desta forma, coloca-se o nome da relação estabelecida entre dois nós apenas na linha do nó satélite da relação (seguindo-se a mesma nomenclatura de Marcu, o nome da relação associado ao núcleo de uma relação mononuclear é *span*); caso a relação seja multinuclear, isto é, as duas porções de texto relacionadas tenham igual importância, o nome da relação é posto para os dois nós envolvidos na relação. A tabela gerada para a estrutura da Figura 3 é apresentada na Tabela 2.

Tabela 2. Tabela para avaliação da estrutura

Nós	Nuclearidade	Relação
1..1	Satélite	Concession
2..2	Núcleo	Span
3..3	Núcleo	List
4..5	Núcleo	List
1..2	Satélite	Non-volitional cause
3..5	Núcleo	Span

Como a avaliação é realizada tendo-se duas estruturas retóricas, realizam-se os mesmos passos descritos para a outra estrutura retórica. Considere outra análise do mesmo texto-exemplo da Figura 1, apresentada na Figura 5 (essa figura também apresenta a segmentação que foi utilizada para a análise).

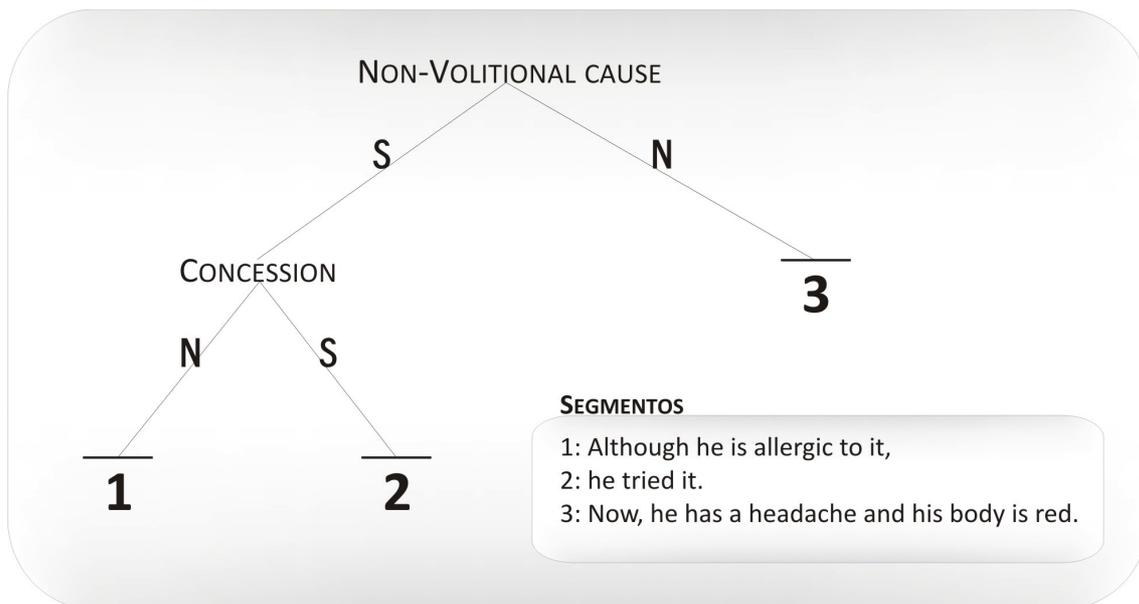


Figura 5. Outra análise retórica para o texto-exemplo

Obtém-se para essa análise uma tabela semelhante à Tabela 2, apresentada na Tabela 3. Observe que nesta análise o segmento 3 equivale aos segmentos 3 e 4 da primeira análise apresentada; outra diferença entre as estruturas apresentadas acontece na definição de nuclearidade na relação *concession* entre os segmentos 1 e 2; essa relação acontece nas duas estruturas, entre os mesmo segmentos, porém o núcleo da primeira análise é o segmento 2 enquanto que na segunda análise esse segmento é satélite na relação.

Tabela 3. Tabela para avaliação da estrutura da segunda estrutura retórica

Nós	Nuclearidade	Relação
1..1	Núcleo	Span
2..2	Satélite	Concession
3..5	Núcleo	Span
1..2	Satélite	Non-volitional cause

Para o cálculo das medidas de cobertura e precisão mesclamos as tabelas, onde cada linha representa um possível nó nas estruturas retóricas. Suponhamos que a primeira estrutura apresentada (Figura 3) seja a estrutura manual (ideal, utilizada como referência na avaliação) e a estrutura da Figura 5 seja a estrutura produzida por algum sistema de análise automática (estrutura que se deseja avaliar). Quando há correspondência de linhas entre as tabelas, apenas uma delas é mantida, porém indica-se que a linha ocorreu tanto na análise automática quanto na manual. Para o exemplo apresentado, obtém-se a Tabela 4.

Tabela 4. Tabela final para cálculo das medidas de avaliação

Limites de Palavras / Nós	Segmentos Textuais		<i>Spans</i>		Nuclearidade		Relação	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
1..1	X	X	X	X	S	N	Concession	Span
2..2	X	X	X	X	N	S	Span	Concession
3..3	X		X		N		List	
4..5	X		X		N		List	
1..2			X	X	S	S	Non-vol...	Non-vol...
3..5		X	X	X	N	N	Span	Span
1..5			X	X	Raiz	Raiz	Span	Span

Pela Tabela 4, calculamos as medidas de cobertura e precisão para os quatro itens listados na Tabela 1. Por exemplo, considere Segmentos Textuais o item a avaliar: na análise manual, há 4 segmentos, sendo que há 2 correspondências na análise automática, levando a uma cobertura de 2/4 dos segmentos identificados manualmente. Dos 3 segmentos considerados na análise automática, 2 estão corretos, obtendo-se a precisão de 2/3. A Tabela 5 mostra os cálculos de cobertura e precisão para todos os itens.

Tabela 5. Cobertura e precisão finais

Item avaliado	Cobertura	Precisão
Segmentos textuais	2/4	2/3
<i>Spans</i>	5/7	5/5
Relações	3/7	3/5
Nuclearidade	3/7	3/5

3. A Ferramenta de Avaliação Automática

Disponível na web, a ferramenta desenvolvida implementa o método de avaliação apresentado na seção anterior. Necessita-se apenas que se forneçam as estruturas retóricas, seguindo-se, nos arquivos, as especificações de formato.

Na Figura 6, apresenta-se a tela inicial, onde se localizam as entradas para os arquivos que contenham as análises discursivas.

RSTc | Tool for discourse parsing evaluation

This tool provides a automatic method to compare two RST structures, one made by a human being (the ideal structure) and another made by a automatic system.

1 Evaluation ID:

2 **Manual Analysis:**

TREE (prolog-like file with tree):

EDU (File with the segments:):

3 **Automatic Analysis:**

TREE (prolog-like file with tree):

EDU (File with the segments:):

4 Language: ▾

5

Figura 6. Tela inicial da ferramenta de avaliação

A ferramenta permite que se armazene o histórico de diversas avaliações realizadas anteriormente (na forma de uma tabela). Para isso, o usuário deve informar o mesmo *Evaluation ID* a cada avaliação que realizar. Para cada identificação fornecida, é armazenado um histórico de todas as avaliações realizadas.

O método de avaliação não considera a língua em que foi feita a análise, exceto na etapa de remoção das *stopwords*. Atualmente, a ferramenta conta com as línguas Português, Inglês e Espanhol, sendo fácil a adição de uma nova língua pela colocação de uma lista de *stopwords* na ferramenta.

Seguindo o formato utilizado pelo DiZer (Pardo e Nunes, 2008), cada análise é constituída de dois arquivos. Um dos arquivos contém a estrutura arbórea da análise através de uma representação parentizada, onde os nós da árvore são indicados por números que identificam os segmentos textuais. O outro arquivo contém os segmentos textuais, um por linha, identificados seqüencialmente por números, os mesmos do arquivo com a estrutura arbórea.

Considerando o texto-exemplo apresentado na Figura 1 e as análises apresentadas nas Figuras 3 e 5 (sendo a primeira considerada ideal e a segunda provinda de alguma ferramenta automática), a Figura 7 apresenta as estruturas retóricas das respectivas análises citadas. A Figura 8 apresenta os arquivos contendo os segmentos conforme o formato requerido pela ferramenta.

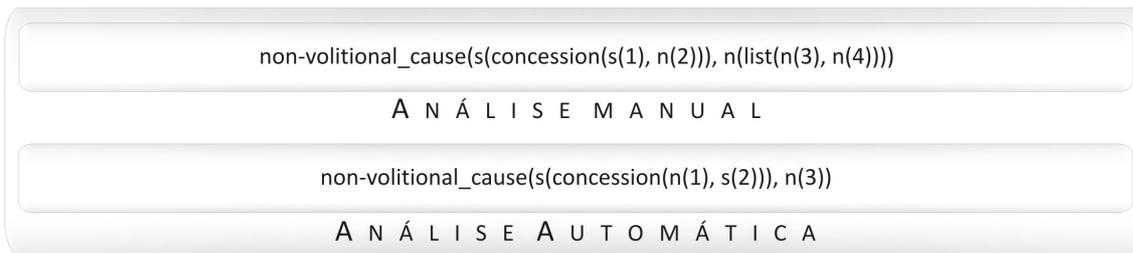


Figura 7. Exemplos de arquivos de estruturas retóricas utilizados pela ferramenta

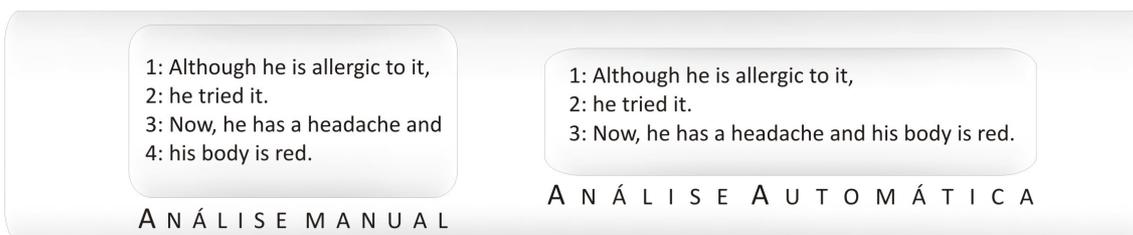


Figura 8. Exemplos de arquivos de segmentos utilizados pela ferramenta

Ao se submeter os arquivos, o método de avaliação é aplicado e as medidas são calculadas. A Figura 9 mostra a tela que informa os resultados da avaliação, assim como resultados anteriores que foram realizados sob o mesmo *Evaluation ID*, que é indicado na tela inicial. Para a última avaliação realizada, uma tabela semelhante à Tabela 4 é exibida, permitindo ao usuário verificar os resultados. Como na Tabela 4, os segmentos e *spans* são identificados apenas pelos números das palavras iniciais e finais; acrescentaram-se, na tabela final exibida pela ferramenta, quais são essas palavras limites dos segmentos, para facilitar a análise dos resultados.

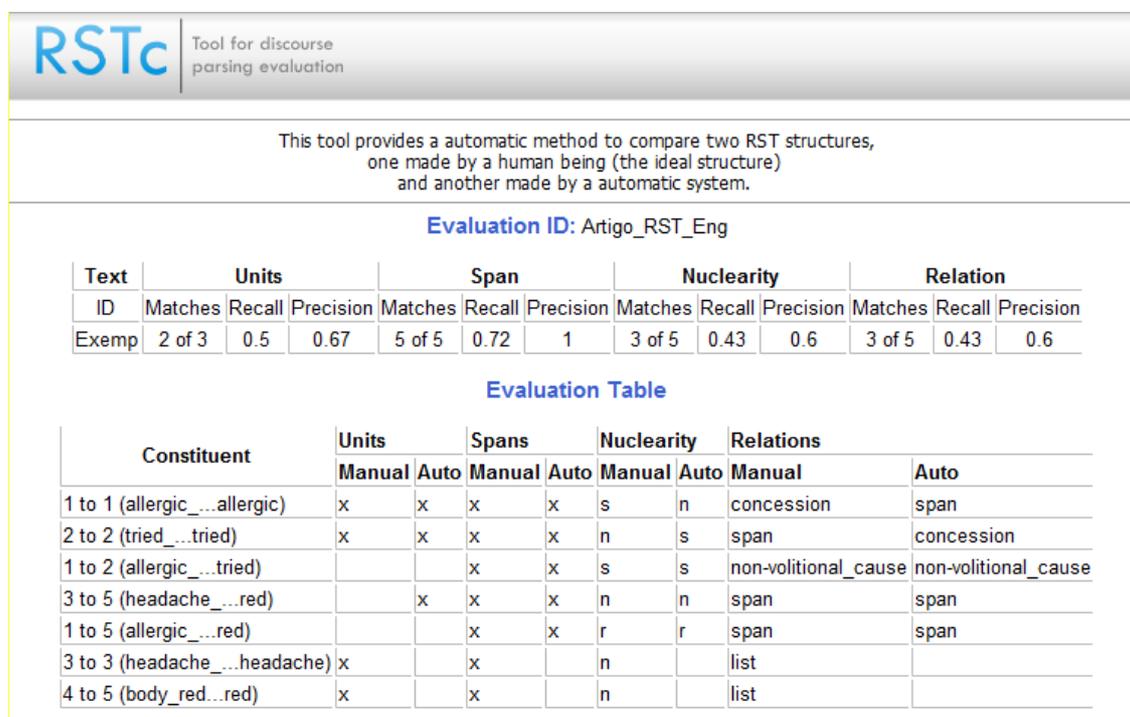


Figura 9. Resultado exibido pela ferramenta na avaliação do exemplo apresentado

4. Considerações Finais

Espera-se, com esse método implementado na ferramenta de avaliação automática, que a evolução da ADA aconteça mais rapidamente, pois a etapa de avaliação de sistemas de análise discursiva automática pode ser realizada automaticamente. Sistemas como o DiZer, desenvolvido por Pardo e Nunes (2008), beneficiam-se dessa ferramenta de avaliação automática, pois evoluções no analisador retórico automático podem ser validadas automaticamente.

A ferramenta desenvolvida encontra-se disponível para uso na página do NILC, mais especificamente, no endereço <http://www.nilc.icmc.usp.br/rsteval/>.

Agradecimentos

À FAPESP, pelo apoio a este trabalho, e a Iria da Cunha Fanego, pelas sugestões e auxílio no design e avaliação da interface produzida.

Referências Bibliográficas

- Mann, W.C. e Thompson, S.A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization*. Technical Report ISI/RS-87-190.
- Marcu, D. (2000). The Rhetorical Parsing of Unrestricted Texts: A Surface-based Approach. *Computational Linguistics*, Vol. 26, pp. 396-448.
- Pardo, T.A.S. and Nunes, M.G.V. (2008). On the Development and Evaluation of a Brazilian Portuguese Discourse Parser. *Journal of Theoretical and Applied Computing*, Vol. 15, N. 2, pp. 43-64.