

DiZer: Uma Proposta de Análise Discursiva Automática para o Português do Brasil¹

Thiago Alexandre Salgueiro Pardo, Lucia Helena Machado Rino,
Maria das Graças Volpe Nunes

Núcleo Interinstitucional de Linguística Computacional (NILC)
CP 668 – ICMC-USP, 13560-970 São Carlos, SP, Brazil
<http://www.nilc.icmc.usp.br>
{thiago@nilc.icmc.usp.br; lucia@dc.ufscar.br;
gracan@icmc.usp.br}

***Abstract.** This paper presents a proposal for automatic discourse analysis of texts written in Brazilian Portuguese. The corresponding system, named DiZer, takes as input a full text and yields its rhetorical, semantic, and intentional structures. Based on corpus analysis, the underlying research is aimed at verifying the contribution of morphology, syntax, semantics, and statistics to discourse analysis.*

***Resumo.** Este artigo apresenta uma proposta de análise discursiva automática de textos em português do Brasil. O sistema correspondente, chamado DiZer, tem como entrada um texto e deve construir suas estruturas retórica, semântica e intencional. Baseando-se em análise de corpus, essa pesquisa deverá verificar a contribuição do conhecimento morfológico, sintático, semântico e estatístico para a análise discursiva proposta.*

1. Introdução

Considerando que um texto não é uma simples seqüência de sentenças desconexas, a análise de discurso contempla sua estruturação altamente elaborada, identificando relações entre suas partes, as quais lhe atribuem sentido, tornando-o coerente e passível de ser entendido. Há vários modelos de estruturação de textos propostos tanto na área de Linguística quanto na área de Processamento de Línguas Naturais (PLN), todos considerando relações interproposicionais como determinantes de sua coesão e coerência. As proposições, unidades mínimas de significado (Mann e Thompson, 1988), são expressas na superfície textual por segmentos textuais. O Texto 1 da Figura 1 (com seus segmentos numerados, para referência), por exemplo, sugere que há uma relação de oposição entre o significado dos segmentos 1 e 2 e uma relação de causa-efeito entre o significado da primeira sentença (formada pelos segmentos 1 e 2) e da segunda (segmento 3), como mostra a Figura 2. Essas relações podem, assim, fazer parte do elenco de relações discursivas que um sistema automático deve buscar ao analisar um texto.

¹ Este trabalho é financiado pela FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo.

[1] Ele queria jogar tênis com Janete, [2] mas também queria jantar com Suzana. [3] Sua indecisão o deixou louco.

Figura 1 - Texto 1

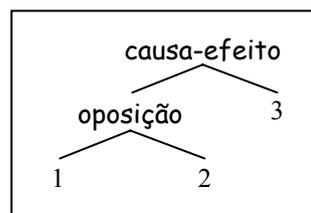


Figura 2 - Estrutura discursiva do Texto 1

Relações discursivas podem ser de diversas naturezas e ocorrer em diferentes níveis de análise. Podem, ainda, ser interpretadas diferentemente por diversos leitores, dependendo de suas intenções ou do conhecimento utilizado durante a compreensão do conteúdo. De forma análoga, elas podem ser usadas ou selecionadas também de modos diversos, dependendo das intenções de escrita do autor.

A análise discursiva automática é muito importante devido a suas inúmeras aplicações. Na Sumarização Automática, por exemplo, a estrutura discursiva pode ajudar a identificar segmentos supérfluos do texto, os quais devem ser omitidos para a produção de um sumário (Spark Jones, 1993). Na Resolução Anafórica, pode-se identificar o contexto de referência de uma anáfora pela estrutura discursiva do texto (Cristea et al., 2000). Na Categorização de Textos, relações discursivas podem indicar o gênero dos textos. A análise discursiva é essencial, sobretudo, para o desenvolvimento de técnicas fundamentais de PLN, isto é, de técnicas que manipulam conhecimento profundo em suas aplicações.

Neste trabalho, o conhecimento profundo considerado envolve tanto o conhecimento lingüístico quanto o discursivo, em uma proposta de análise discursiva automática para textos em português do Brasil. É sugerida a construção do DiZer-PBr (*Discourse analyZER for BRazilian Portuguese*), um sistema que produz as estruturas retórica, semântica e intencional de um texto. No nível retórico, utiliza-se a RST (*Rhetorical Structure Theory*) (Mann e Thompson, 1988); no nível semântico, as relações de Jordan (1992); no nível intencional, as relações da GSDT (*Grosz e Sidner Discourse Theory*) (Grosz e Sidner, 1986). Essas teorias são apresentadas na próxima seção. O conhecimento morfológico, sintático, semântico e estatístico necessário para o DiZer executar uma análise discursiva é extraído de um corpus de textos científicos. Essa abordagem é apresentada na Seção 3. A Seção 4 ilustra uma possível aplicação para o DiZer. Na Seção 5, algumas considerações finais são apresentadas.

2. Teorias discursivas

Como já foi dito, as relações discursivas ocorrem em diferentes níveis de análise. Os trabalhos descritos abaixo contemplam os níveis retórico, intencional e semântico.

2.1. Retórica

No âmbito da RST (Mann e Thompson, 1988), a retórica é o meio pelo qual um texto é organizado para satisfazer o objetivo comunicativo que o escritor pretende atingir ao escrever um texto. São as relações retóricas que delineiam como seu conteúdo é relacionado e em que medida o significado de cada segmento textual contribui para a satisfação das intenções do escritor. Segundo os autores, a RST estabelece relações retóricas aplicáveis a uma grande gama de textos. No caso padrão, essas relações se estabelecem entre duas proposições simples, geralmente expressas por segmentos adjacentes no texto, sendo uma nuclear (N) e outra complementar (S, de satélite). O núcleo e o satélite indicam, respectivamente, a informação principal para a satisfação da intenção subjacente à relação e a informação adicional, a qual influencia de alguma

forma a interpretação que o leitor faz da informação nuclear. Quando ambas as informações relacionadas são igualmente importantes, diz-se que se tem uma relação multinuclear, isto é, com mais de um núcleo e nenhum satélite. A relação retórica de oposição CONCESSION sugere no texto “Embora você não goste, trabalhar é importante.”, indica que a primeira proposição é o satélite e a segunda é o núcleo. Por sua vez, em “O garoto chegou da escola e fez sua lição de casa. Depois, foi brincar com os amigos.”, a relação RST SEQUENCE (que indica uma seqüência de eventos) se estabelece entre as proposições correspondentes aos segmentos de texto “O garoto chegou da escola”, “fez sua lição de casa.” e “foi brincar com os amigos.”, sendo que todas são consideradas núcleos da relação, pois possuem a mesma importância.

É importante notar que, ao organizar as proposições em núcleos e satélites das relações, a estrutura RST resultante será uma árvore hierárquica de proposições. A Figura 3 mostra uma possível estrutura retórica para o Texto 1 (arcos com rótulos N indicam núcleos, enquanto o arco com rótulo S indica o satélite da relação), na qual (i) a relação multinuclear CONTRAST indica oposição entre as proposições 1 e 2 e (ii) a relação NON-VOLITIONAL-CAUSE indica que 1 e 2 causam 3.

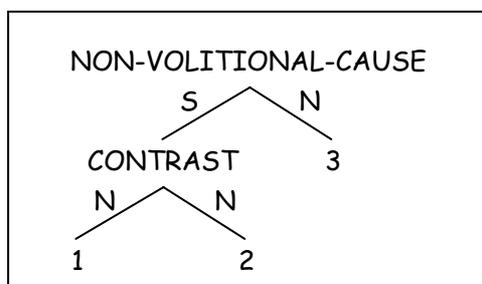


Figura 3 - Estrutura retórica do Texto 1

2.2. Intenções

Segundo a GSDT (*Grosz e Sidner Discourse Theory*) (Grosz e Sidner, 1986), todo discurso é essencialmente produzido com a finalidade de satisfazer uma ou mais intenções, sendo elas as responsáveis por individualizar e tornar o discurso coerente. Assim, quando um escritor escreve seu texto, ele seleciona e estrutura seu conteúdo em função de suas intenções. Pode-se dizer, então, que todo segmento discursivo possui uma intenção que contribui para a escolha dos demais segmentos do discurso e, de modo geral, para a satisfação da intenção do escritor. É o reconhecimento dessas intenções que permite ao leitor recuperar o que o escritor pretendia comunicar. Como as intenções possíveis em um discurso são teoricamente infinitas, a GSDT organiza o discurso por meio de relações de contribuição e satisfação entre as intenções, que, por sua vez, são finitas e em número de duas:

- *dominance* (DOM): se a intenção subjacente a um segmento Y contribui para a satisfação da intenção subjacente a um segmento X, então a intenção subjacente a X domina (*dominates*) a intenção subjacente a Y, ou seja, DOM(X,Y);
- *satisfaction-precedence* (SP): se a intenção subjacente a um segmento X deve ser satisfeita antes da intenção subjacente a um segmento Y, então a satisfação da intenção subjacente a X deve preceder (*satisfaction-precedes*) a satisfação da intenção subjacente a Y, ou seja, SP(X,Y).

No Texto 1, por exemplo, existe uma relação de dominância entre a intenção subjacente ao segmento discursivo formado pela proposição 3 e a intenção subjacente ao segmento discursivo formado pelas proposições 1 e 2, ou seja, DOM(3,[1-2]).

2.3. Semântica

De acordo com Jordan (1992), uma relação semântica constitui uma “noção semântica textual de conexão binária entre quaisquer duas partes de um texto”. Pode-se dizer que as relações semânticas relacionam as proposições subjacentes ao texto, atribuindo-lhe sentido. Embora relações semânticas pareçam similares às relações retóricas, elas são desprovidas de força argumentativa, por não considerarem sua intenção subjacente (Hovy, 1991, 1993; Moore e Pollack, 1992; Moser e Moore, 1996; Rino, 1996; Marcu, 1999). Por exemplo, na estrutura retórica da Figura 3, considerou-se que a proposição 3 é mais importante do que as proposições 1 e 2 e, por isso, ela foi identificada como núcleo da relação NON-VOLITIONAL-CAUSE. Por outro lado, caso as proposições 1 e 2 fossem consideradas mais importantes, elas formariam o núcleo da relação retórica NON-VOLITIONAL-RESULT, enquanto a proposição 3 seria o satélite. Ou seja, dependendo da informação que se deseja ressaltar para a satisfação da intenção subjacente, diferentes relações retóricas podem ser utilizadas. Diferentemente dessas conotações, o modelo semântico considera apenas a relação factual entre as proposições que, nesse exemplo, é a relação única de causa-efeito. A Figura 2 é um exemplo de estrutura semântica possível para o Texto 1.

Com base nesses modelos de organização do discurso, a metodologia de automação da análise discursiva no DiZer é descrita a seguir.

3. DiZer: análise discursiva automática para o português do Brasil

O DiZer visa produzir, a partir de um texto-fonte, suas estruturas retórica, semântica e intencional. O modelo de análise discursiva proposto é baseado em conhecimento, abrangendo as informações morfológicas, sintáticas, semânticas e estatísticas de textos em português. No estágio atual, as bases de conhecimento são produzidas a partir da análise de um corpus de textos científicos.

3.1. Relações discursivas e seu inter-relacionamento no DiZer

Diferentemente das pesquisas atuais, as quais abordam principalmente a automação da análise retórica (por exemplo, Marcu 1997, 2000a; Corston-Oliver, 1998; Schilder, 2002), este trabalho aborda também a automação das análises semântica e intencional, seguindo a perspectiva de que esses três níveis são representativos do discurso e, portanto, devem ser úteis para a resolução de problemas de PLN (Moore e Pollack, 1992; Hovy, 1993; Moser e Moore, 1996).

No DiZer, a estrutura retórica é a principal fonte de informação, que originará as estruturas semântica e intencional, caracterizando como vital a modelagem do mapeamento das relações retóricas em relações semânticas e intencionais. Seguindo outras propostas (Maier e Hovy, 1991; Hovy, 1991; 1993; Maybury, 1992; Moore e Pollack, 1992; Dale, 1993; Korelsky e Kittredge, 1993; Maier, 1993; Moore e Paris, 1993; Moore, 1995; Moser e Moore, 1996; Rino, 1996; Marcu, 1999; 2000b; Pardo, 2002), o DiZer assume a co-existência de vários níveis de representação do discurso: a intenção, que dá origem a qualquer discurso, é expressa por meio de relações retóricas; por sua vez, estas são definidas sobre proposições semanticamente relacionadas. Especificamente em relação à modelagem do mapeamento entre as relações, são adotados o mapeamento entre intenções e retórica sugerido por Moser e Moore (1996) e Marcu (1999, 2000b) e o mapeamento entre retórica e semântica sugerido por Korelsky e Kittredge (1993).

Moser e Moore foram os primeiros a propor um inter-relacionamento entre a GSDT e a RST. Apesar dos pressupostos teóricos dessas teorias diferirem, Moser e Moore afirmam que o núcleo N de uma relação retórica domina seu satélite S, caracterizando a relação GSDT DOM(N,S). Complementando essa pesquisa, Marcu sugere que, em relações multinucleares, nas quais não há satélite, pode haver relações de precedência entre os núcleos, caracterizando a relação GSDT SP(N₁,N₂). Assim, segundo esse modelo de mapeamento, a partir de uma estrutura retórica é possível derivar automaticamente as relações intencionais de dominância e, pelo menos parcialmente, as relações intencionais de precedência. Em relação ao mapeamento entre retórica e semântica, Korelsky e Kittredge mostram que uma mesma relação retórica pode ser definida entre proposições relacionadas por diferentes relações semânticas e que uma mesma relação semântica pode originar diferentes relações retóricas. Os autores sugerem algoritmos que derivam relações semânticas a partir das relações retóricas, fazendo uso de restrições associadas às últimas. A Figura 4 mostra um algoritmo para mapear a relação RST EVIDENCE na relação semântica mais adequada (Korelsky e Kittredge, 1993, p. 54). Korelsky e Kittredge, entretanto, utilizam outras denominações de relações: suas relações semânticas representam um subconjunto das relações RST – as chamadas relações informativas. O DiZer trata essa diferença univocamente: as relações RST informativas continuam sendo retóricas, porém, seu mapeamento em relações semânticas é direto, com correspondentes únicos no modelo de Jordan (1992). Quando as relações retóricas não forem informativas, mas pragmáticas (segundo a terminologia da RST), o mapeamento não necessariamente será único.

Se a relação retórica EVIDENCE for observada entre duas proposições P1 e P2, em que P1 for o núcleo e P2 for o satélite, então

- 1) se houver um agente consciente de tal forma que P1 e P2 fazem referência a suas ações, então a relação semântica *volitional-cause* se estabelece entre as proposições
- 2) se não há um agente consciente, então a relação semântica *non-volitional-cause* se estabelece entre as proposições
- 3) se P2 é uma proposição genérica, então a relação semântica *elaboration* se estabelece entre as proposições

Figura 4 - Determinação da relação semântica correspondente a EVIDENCE

O conhecimento utilizado nesse mapeamento, assim como o conhecimento empregado no mapeamento entre intenções e retórica, será compilado a partir de análise de corpus, como descrita a seguir.

3.2. Metodologia de desenvolvimento do DiZer

Um corpus de 54 introduções de teses e dissertações do domínio da Computação, chamado CorpusTCC (CORPUS de Textos Científicos da Computação), foi montado a partir de textos do Corpus NILC (Pinheiro e Aluísio, 2003). Esse corpus ainda se encontra em processo de marcação retórica, sendo usada, para isso, a ferramenta RSTTool (O'Donnel, 1997). Além da marcação retórica, esse corpus será marcado com as relações semânticas e intencionais consideradas neste projeto. Toda essa marcação contempla as relações interproposicionais. O corpus também será anotado com

informações intra-sentenciais morfológicas, sintáticas e semânticas. Serão utilizados, para isso, um etiquetador morfossintático (*tagger*) (Aires, 2000; Aires et al., 2000) e um *parser* (Martins et al., 2002) do NILC. Para a marcação semântica, será investigado o uso da UNL (*Universal Network Language*) (Uchida, 2000), formalismo que incorpora outras teorias de representação semântica (por exemplo, Jackendoff, 1990; Fillmore, 1968), relativamente simples para o tratamento intra-sentencial.

A partir do conhecimento indicado pelas marcações do corpus, serão extraídos os padrões de análise (retórica e intencional), assim como as regras de mapeamento de relações retóricas em relações semânticas e as estatísticas de estruturação discursiva. Esse conhecimento servirá de base para derivar automaticamente as estruturas discursivas de um novo texto. Para o Texto 2 da Figura 5, supondo-se que as relações RST VOLITIONAL-RESULT e CONTRAST existam entre as proposições expressas pelos segmentos 1 e 2 e as proposições expressas pelos segmentos [1-2] e 3, respectivamente, os padrões de análise retórica das Figuras 6 e 7 podem ser produzidos.

A Figura 6 indica que a relação VOLITIONAL-RESULT com o núcleo realizado antes do satélite no texto (NS) ocorre se (a) não há marcador discursivo, (b) os elementos de conexão entre os segmentos correspondentes às duas proposições relacionadas retoricamente são uma vírgula seguida por um verbo no gerúndio, (c) esses segmentos são orações adjacentes, sendo que o segmento cuja proposição é o satélite é uma oração subordinada adjetiva explicativa reduzida de gerúndio e (d) o agente das duas proposições é o mesmo. Os itens (a) e (b) correspondem, basicamente, ao conhecimento morfológico; o item (c), ao conhecimento sintático e o item (d), ao conhecimento semântico. O padrão de análise da Figura 7 pode ser interpretado de forma similar, exceto pelo fato de se referir a uma relação multinuclear (NN).

Esses padrões englobam, a princípio, todo o conhecimento necessário para a análise retórica automática do DiZer, que será realizada pelo simples processo de casamento de padrões. Por exemplo, o padrão da Figura 7 permite a análise automática do Texto 3 (Figura 8), pois todas as suas restrições são satisfeitas.

[1] Sistemas de exploração livre permitem que o aluno tenha progresso de acordo com os seus interesses e objetivos, [2] criando um ambiente favorável para a busca de informações. [3] Entretanto, a falta de orientação pode retardar o alcance dos objetivos e o desvio do mesmo.

Figura 5 - Texto 2

| Relação retórica: VOLITIONAL-RESULT-NS |
|---|
| a) marcador discursivo: Ø |
| b) elementos de conexão: vírgula (,) seguida por um verbo no gerúndio |
| c) restrições sintáticas: duas orações A e B adjacentes; oração A é a oração principal; oração B é uma oração subordinada adjetiva explicativa reduzida de gerúndio |
| d) restrições semânticas/conceituais: o agente da proposição A é o mesmo da B |

Figura 6 - Padrão de análise para a relação retórica VOLITIONAL-RESULT

| |
|---|
| Relação retórica: CONTRAST-NN |
| a) marcador discursivo: "Entretanto" |
| b) elementos de conexão: ponto (.) seguido pelo marcador discursivo, que é seguido por uma vírgula (,) |
| c) restrições sintáticas: duas sentenças A e B adjacentes; marcador discursivo pertencente à sentença B |
| d) restrições semânticas/conceituais: \emptyset |

Figura 7 - Padrão de análise para a relação retórica CONTRAST

O desenvolvimento do hardware sempre foi sustentado e motivado pelo desenvolvimento da eletrônica. Entretanto, o desenvolvimento do software não acompanha o do hardware na mesma velocidade.

Figura 8 - Texto 3

De forma similar, pode-se extrair padrões de análise intencional para as relações de precedência. Não serão construídos padrões de análise para as relações de dominância, pois estas podem ser derivadas diretamente da estrutura retórica produzida anteriormente. A Figura 9 mostra o padrão de análise intencional do Texto 2. É importante notar que, além das restrições para a verificação da intenção subjacente, acrescenta-se ainda a restrição sobre a relação retórica associada, necessária para o mapeamento entre as relações em foco.

| |
|--|
| Relação intencional: SP(A,B) |
| a) marcador discursivo: "Entretanto" |
| b) elementos de conexão: ponto (.) seguido pelo marcador discursivo, que é seguido por uma vírgula (,) |
| c) restrições sintáticas: duas sentenças A e B adjacentes; marcador discursivo pertence à sentença B |
| d) restrições semânticas/conceituais: \emptyset |
| e) relação retórica: CONTRAST-NN |

Figura 9 - Padrão de análise intencional

Para extrair a estrutura semântica, o algoritmo da Figura 4 pode ser usado. Por fim, estatísticas sobre a estruturação retórica serão coletadas do corpus e utilizadas para resolver casos de ambigüidade na análise automática e/ou escolher a estrutura retórica mais provável, dentre várias. Por exemplo, para a estrutura retórica do Texto 2, uma possível estatística seria: "uma relação retórica VOLITIONAL-RESULT seguida por uma relação CONTRAST e precedida por uma relação... ocorre com probabilidade P".

Segundo as etapas de processamento antes descritas, a arquitetura do DiZer é ilustrada na Figura 10. As bases de conhecimento contêm todas as informações resultantes da análise de corpus. É importante notar que, para que o casamento de padrões seja possível, qualquer texto de entrada deve ser marcado com as informações (intra-sentenciais) morfossintáticas, sintáticas e semânticas. A construção de estruturas retóricas válidas a partir das relações retóricas entre as proposições identificadas pelos padrões de análise será baseada no algoritmo já proposto por Marcu (1997) em seus trabalhos de análise retórica automática.

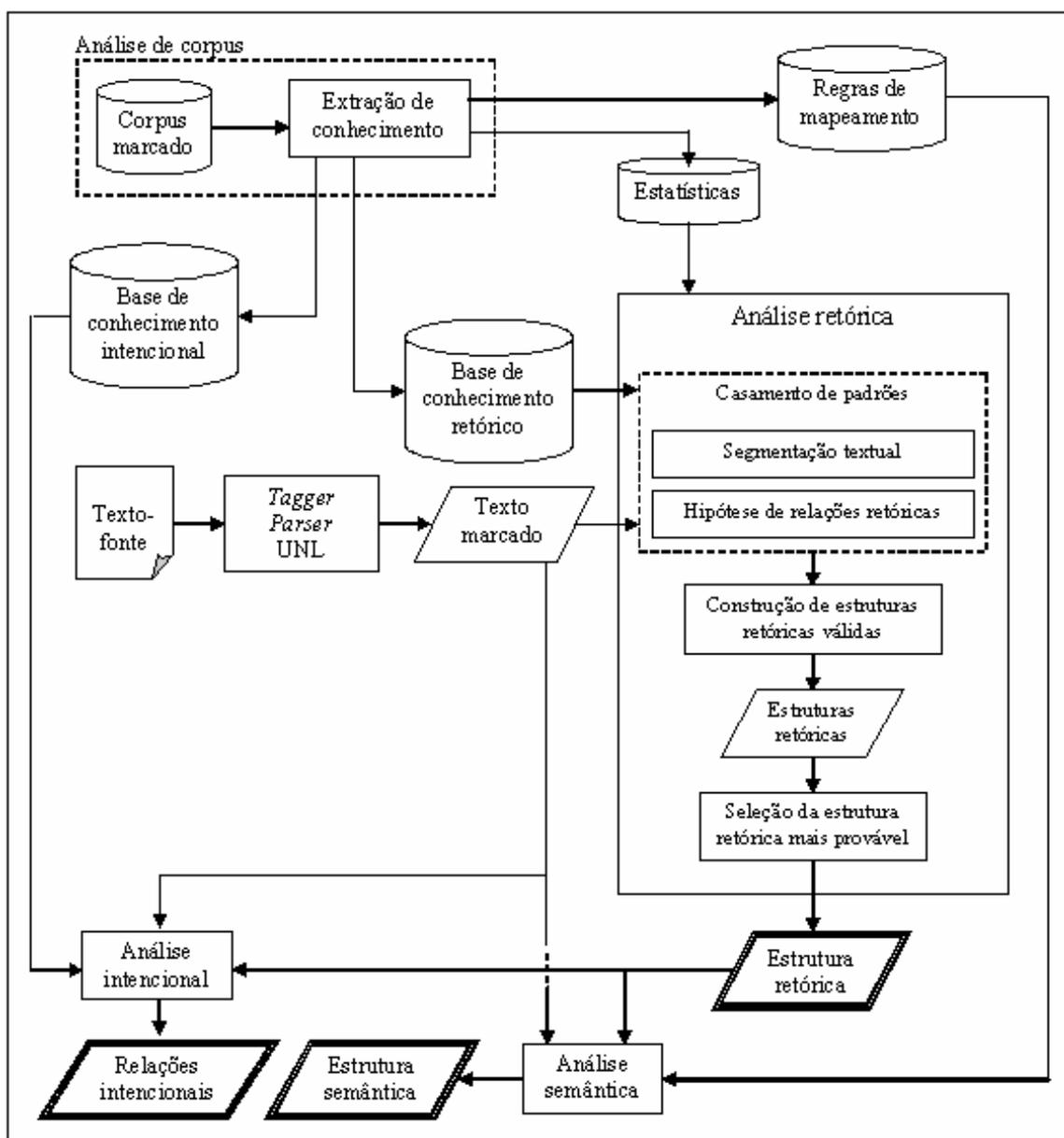


Figura 10 - Arquitetura do DiZer

Embora ainda em estágio preliminar, o DiZer será avaliado para verificar a contribuição de cada nível de conhecimento para a análise discursiva proposta. Ele também será validado na Sumarização Automática de Textos, como se discute na próxima seção.

4. Validação do DiZer no ambiente do DMSumm

O DMSumm (*Discourse Modeling SUMMARizer*) (Pardo, 2002) é um gerador automático de sumários baseado em modelagem discursiva cuja entrada não é um texto em língua natural, mas já sua representação discursiva, composta de suas estruturas semântica e intencional. Até o momento, essa representação tem sido obtida manualmente, devido à inexistência de um analisador discursivo automático de textos. O DiZer pretende suprir essa tarefa interpretativa, fornecendo a entrada ao DMSumm.

Uma avaliação preliminar do DMSumm já foi apresentada (Pardo e Rino, 2002), devendo ser reproduzida para o mesmo corpus de teste a ser adotado para o DiZer. Uma

vez garantida a qualidade do DMSumm para o processamento desse corpus, a validação do DiZer consistirá na avaliação de seu desempenho, quando acoplado ao DMSumm, a qual indicará sua validade para a Sumarização Automática. Ao validar o DiZer, será possível, então, ter um sumarizador automático de textos completo.

5. Considerações finais

A proposta fundamental do DiZer é original, quer em relação à análise discursiva automática de textos em português, quer em relação à metodologia de desenvolvimento, pois agrega propostas isoladas originalmente voltadas ao processamento do inglês, utilizando conhecimentos de diversas naturezas. A extração das bases de conhecimento discursivo (envolvendo, aqui, os três níveis de representação do conhecimento: semântico, retórico e intencional) é a tarefa mais complexa do DiZer, pois depende de conhecimento especializado para a marcação manual dos textos que servem de base para a modelagem da análise discursiva. Além disso, essa modelagem é dependente da representatividade do corpus de análise: a) ao se alterar seu gênero, será preciso reconstruir as bases de conhecimento a partir da marcação manual; b) mesmo mantendo-se o gênero, ainda não se sabe o volume ideal de textos para a extração dessas bases: atualmente, em uma metodologia empírica, utilizam-se 54 textos científicos. Vislumbra-se, a partir dessa primeira etapa de análise, a possibilidade de se explorar técnicas de Aprendizado de Máquina para extrair mais conhecimento ou assegurar que o sistema se mantenha estável para determinado volume de textos.

Referências bibliográficas

- Aires, R.V.X. (2000). *Implementação, Adaptação, Combinação e Avaliação de Etiquetadores para o Português do Brasil*. Dissertação de Mestrado. Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo. São Carlos – SP.
- Aires, R.V.X.; Aluísio, S.M.; Kuhn, D.C.S.; Andreetta, M.L.B.; Oliveira Jr., O.N. (2000). Combining Multiple Classifiers to Improve Part of Speech Tagging: A Case Study for Brazilian Portuguese. In the *Proceedings of the Brazilian AI Symposium (SBLA'2000)*, pp. 20-22.
- Corston-Oliver, S. (1998). *Computing Representations of the Structure of Written Discourse*. PhD Thesis, University of California, Santa Barbara, CA, USA.
- Cristea, D.; Ide, N.; Marcu, D.; Tablan, V. (2000). Discourse Structure and Co-Reference: An Empirical Study. In the *Proceedings of the 18th International Conference on Computational Linguistics*. Luxembourg.
- Dale, R. (1993). Rhetoric e Intentions in Discourse. In the *Proceedings of the Intentionality e Structure in Discourse Relations Workshop*, pp. 5-6. Ohio, USA.
- Fillmore, C. (1968). The case for case. In Bach, E. and Harms, R.T. (orgs.), *Universals in linguistic theory*, pp. 1-88. Rinehard and Winston, New York.
- Grosz, B. e Sidner, C. (1986). Attention, Intentions, e the Structure of Discourse. *Computational Linguistics*, Vol. 12, N. 3.
- Hovy, E. (1991). Approaches to the planning of coherent text. In C. Paris, W. Swartout e W. Mann (eds.), *Natural Language Generation in Artificial Intelligence e Computational Linguistics*, pp. 83-102. Kluwer Academic Publishers, Boston.
- Hovy, E. (1993). In Defense of Syntax: Informational, Intentional, e Rhetorical Structures in Discourse. In the *Proceedings of the Intentionality e Structure in Discourse Relations Workshop*, pp. 35-39. Ohio, USA.
- Jackendoff, R. (1990). *Semantic Structures*. The MIT Press, Cambridge.
- Jordan, M.P. (1992). An Integrated Three-Pronged Analysis of a Fund-Raising Letter. In W.C. Mann e S.A. Thompson (eds), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, pp. 171-226.

- Korelsky, T. e Kittredge, R. (1993). Towards stratification of RST. In the *Proceedings of the Intentionality e Structure in Discourse Relations Workshop*, pp. 52-55. Ohio, USA.
- Maier, E. (1993). The Representation of Interdependencies between Communicative Goals e Rhetorical Relations in the Framework of Multimedia Document Generation. In the *Proceedings of the Intentionality e Structure in Discourse Relations Workshop*, pp. 70-73. Ohio, USA.
- Maier, E. e Hovy, E. H. (1991). A Metafunctionally Motivated Taxonomy for Discourse Structure Relations. In the *Proceedings of the 3rd European Workshop on Language Generation*. Innsbruck, Austria.
- Mann, W.C. e Thompson, S.A. (1988). Rhetorical structure theory. Toward a functional theory of text organization. *Text*, Vol. 8, N. 3, pp. 243-281.
- Marcu, D. (1997). *The Rhetorical Parsing, Summarization, e Generation of Natural Language Texts*. PhD Thesis, Department of Computer Science, University of Toronto.
- Marcu, D. (1999). A formal e computational synthesis of Grosz e Sidner's e Mann e Thompson's theories. In *Proceedings of the Workshop on Levels of Representation in Discourse*, pp. 101-108. Edinburgh, Scotland.
- Marcu, D. (2000a). *The Theory e Practice of Discourse Parsing e Summarization*. The MIT Press. Cambridge, Massachusetts.
- Marcu, D. (2000b). Extending a Formal e Computational Model of Rhetorical Structure Theory with Intentional Structures à la Grosz e Sidner. In the *Proceedings of the 18th International Conference on Computational Linguistics (COLING'2000)*, Saarbrueken.
- Martins, R.T.; Hasegawa, R.; Nunes, M.G.V. (2002). *Curupira: um parser funcional para o português*. Série de Relatórios do NILC. NILC-TR-02-26.
- Maybury, M.T. (1992). Communicative Acts for Explanation Generation. *Int. Journal of Man-Machine Studies* 37, pp. 135-172.
- Moore, J.D. (1995). *Participating in Explanatory Dialogs: Interpreting e Responding to Questions in Context*. The MIT Press. Cambridge, Massachusetts.
- Moore, J.D. e Paris, C. (1993). Plannig Text for Advisory Dialogues: Capturing Intentional e Rhetorical Information. *Computational Linguistics*, Vol. 19, N. 4, pp. 651-694.
- Moore, J. D. e Pollack, M. E. (1992). A problem for RST: the need for multi-level discourse analysis. *Computational Linguistics*, Vol. 18, N. 4, pp. 537-544.
- Moser, M. e Moore, J. D. (1996). Toward a synthesis of two accounts of discourse structure. *Computational Linguistics*, Vol. 22, N. 3, pp. 409-419.
- O'Donnell, M. (1997). Variable-Length On-Line Document Generation. In the *Proceedings of the 6th European Workshop on Natural Language Generation*, Gerhard-Mercator University, Duisburg, Germany.
- Pardo, T.A.S. (2002). *DMSumm: Um Gerador Automático de Sumários*. Dissertação de Mestrado. Departamento de Computação. Universidade Federal de São Carlos. São Carlos – SP.
- Pardo, T.A.S. and Rino, L.H.M. (2002). DMSumm: Review and Assessment. In E. Ranchhod and N. J. Mamede (eds.), *Advances in Natural Language Processing*, pp. 263-273 (Lecture Notes in Artificial Intelligence 2389). Springer-Verlag, Germany.
- Pinheiro, G.M. e Aluísio, S.M. (2003). *Corpus NILC: Descrição e Análise Crítica com Vistas ao Projeto Lacio-Web*. Série de Relatórios Técnicos do Instituto de Ciencias Matemáticas e de Computação – ICMC, Universidade de São Paulo, N. 190.
- Rino, L.H.M. (1996). *Modelagem de Discurso para o Tratamento da Concisão e Preservação da Idéia Central na Geração de Textos*. Tese de Doutorado. IFSC-Usp. São Carlos - SP.
- Schilder, F. (2002). Robust discourse parsing via discourse markers, topicality e position. In J. Tait, B.K. Boguraev e C. Jacquemin (eds.), *Natural Language Engineering*, Vol. 8. Cambridge University Press.
- Sparck Jones, K. (1993). *Discourse Modelling for Automatic Summarising*. Technical Report No. 290. University of Cambridge.
- Uchida, H. (2000). *Universal Networking Language: An Electronic Language for Communication, Understanding e Collaboration*. UNL Center, IAS/UNU, Tokyo.