



IMPLEMENTAÇÃO DE UM *CORPUS* ANOTADO DISCURSIVAMENTE DO GÊNERO RESPOSTA ARGUMENTATIVA

Camila Cristiane Moreschi (PIBIC/CNPq/Uem), Juliano Desiderato Antonio (Orientador), e-mail: prof.jdantonio@gmail.com

Universidade Estadual de Maringá / Linguística, Letras e Artes/
Departamento de Teorias Linguísticas e Literárias/ Maringá, PR.

Palavras-chave: Funcionalismo, Teoria da Estrutura Retórica, Resposta Argumentativa.

Resumo: Este projeto teve como objetivo criar um *corpus* discursivo anotado com textos do gênero resposta argumentativa tendo em vista a futura implementação de aplicações de avaliação automática da coerência de textos desse gênero. O *corpus* formado é formado por 100 redações do gênero resposta argumentativa produzidas para o vestibular de verão/2013 da UEM. Inicialmente, os textos foram digitalizados para, na sequência, serem segmentados em sentenças. Após a segmentação, realizou-se a identificação da unidade central de cada texto (entendida como o núcleo principal do texto, ou seja, o núcleo ao qual estão associados todos os demais núcleos).

Introdução

O estudo do Processamento de Línguas Naturais (doravante PLN) é responsável pela criação de aplicações como sumarizadores automáticos, sistemas de geração automática de textos e sistemas de tradução automática (Iruskieta *et al.*, 2013). Um outro campo de estudos que vem se mostrando profícuo é a avaliação automática da coerência de textos (Lin e Kan, 2011).

Para que essas aplicações sejam desenvolvidas, é necessário o trabalho de linguistas que anotem um *corpus* com o(s) gênero(s) textual(is) a ser(em) processado(s) em diferentes níveis, inclusive no nível discursivo. A avaliação automática da coerência de textos depende da identificação das estruturas relacionais associadas ao estabelecimento das relações de coerência.

Materiais e métodos

O *corpus* da pesquisa é composto por 100 redações do gênero resposta argumentativa produzidas para o vestibular de verão/2013 da UEM



(as redações foram disponibilizadas pela Comissão Central do Vestibular Unificado). Os candidatos atenderam ao seguinte comando: “Como vestibulando, redija, em até 15 linhas, uma resposta argumentativa à pergunta ‘Qual o segredo do vestibular: inteligência, esforço ou sorte?’. Você pode basear-se nas informações dos textos de apoio, mas não deve apresentar cópia deles”.

O primeiro passo para o tratamento dos textos foi digitalizá-los. Na sequência, os textos foram segmentados em sentenças. Após a segmentação, realizou-se a identificação da unidade central de cada texto. A unidade central é, segundo Stede (2008), o núcleo principal do texto, ou seja, o núcleo ao qual estão associados todos os demais núcleos. Após a identificação da unidade central, o *corpus* foi disponibilizado *online* no link <http://ixa2.si.ehu.es/rst/pt/index.php>.

Resultados e Discussão

A identificação da unidade central foi possível a partir dos seguintes parâmetros: posição da unidade central no texto; retomada da pergunta; frequência de ocorrência de determinados substantivos, advérbios e verbos associados com um determinado padrão de resposta.

Verificou-se que a unidade central ocorre em 90,7% dos textos do *corpus* e que a afirmação inicial com retomada da pergunta, que se espera no gênero argumentativa, ocorreu em 82,7% das unidades centrais, conforme a tabela 1.

	Inicial		Não-inicial	
	N	%	N	%
Com retomada da pergunta	62	82,7%	4	5,3%
Sem retomada da pergunta	6	8%	3	4%
	68	90,7%	7	9,3%

Tabela 1- Posição da unidade central e retomada da pergunta para iniciar a resposta

Devido à falta de espaço não é possível apresentar todos os 55 substantivos utilizados na unidade central dos textos. Os 5 substantivos responsáveis por 75% das 317 ocorrências de substantivos foram: “vestibular” e “segredo”, que fazem parte da pergunta fundadora da resposta, e os substantivos “esforço”, “sorte” e “inteligência”, que representam as respostas dadas pelos candidatos.

Um grupo de substantivos também apresentou 13 ocorrências. Vários candidatos elegeram mais de um fator para o sucesso no vestibular, e utilizaram algum substantivo do campo semântico da junção: junção, combinação, conjuntura, miscigenação (este substantivo foi utilizado de forma equivocada pelo candidato), mistura, mix, soma, união.



A maioria dos verbos encontrados na unidade central era de cópula, como os verbos “ser” e “estar”, que tiveram 94,7% das ocorrências. Isso se deve à pergunta que motivou a seguinte estrutura: “O segredo do vestibular + *verbo de cópula* + [junção de] + **fator**”.

Verbos evidenciais como “acredito”, “creio” e “penso” (Hattner, 2007) foram utilizados para encaixar a resposta. Dessa forma, o padrão de resposta passa a ter a seguinte estrutura: [Eu + *verbo evidencial* + que] o segredo do vestibular + *verbo de cópula* + [junção de] + **fator**.

Advérbios e sintagmas adverbiais asseverativos também foram utilizados para aumentar a credibilidade das respostas, como pode ser observado na tabela 2.

		N	%
Epistêmicos (Neves, 2000)	Sem dúvida	5	26,2%
	Na verdade	1	5,3%
	Indubitavelmente	1	5,3%
	Certamente	1	5,3%
Intensificadores (Neves, 2000)	Muito	3	15,7%
	Pouco	2	10,5%
Inclusão com exclusividade (Neves, 2000)	Apenas	2	10,5%
	Só	1	5,3%
Focalizadores (Castilho, 2010)	Principal-mente	1	5,3%
	Também	1	5,3%
Frequência (Neves, 2000)	Geralmente	1	5,3%
Total		19	100%

Tabela 2- Advérbios e sintagmas adverbiais

O padrão de resposta encontrada com maior frequência no corpus apresenta, portanto, a seguinte estrutura: [Eu + *verbo evidencial* + que] o segredo do vestibular + *verbo de cópula* + [advérbio] + [junção de] + **fator**.

Conclusões

Por meio da pesquisa foi possível identificar o padrão de resposta das unidades centrais dos textos do *corpus*, o que possibilitou a segmentação dos textos e a criação do banco de dados online.

Agradecimentos

Agradeço ao CNPq por subsidiar esta pesquisa, ao Professor Dr. Juliano Desiderato Antonio pelas orientações e ao grupo de pesquisa FuncPar (Grupo de Pesquisas Funcionalistas do Norte/Nordeste do Paraná)



pelas oportunidades de melhoramento e aprendizado na minha jornada acadêmica.

Referências

CASTILHO, A.T. *Nova Gramática do Português Brasileiro*

IRUSKIETA, M. et al. **The RST Basque TreeBank: an online search interface to check rhetorical relations**. Anais do IV Workshop on RST and Text Studies. Fortaleza: Biblioteca Digital Brasileira de Computação: 40-49 p. 2013.

LIN, H. T. N. Z.; KAN, M.-Y. **Automatically Evaluating Text Coherence Using Discourse Relations**. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland: Association for Computational Linguistics: 997–1006 p. 2011.

NEVES, M.H.M.N. *Gramática de usos do português*. S. Paulo: Editora Unesp, 2000.

STEDE, M. RST revisited: disentangling nuclearity. In: FABRICIUS-