

UM CORPUS DO PORTUGUÊS DO BRASIL: VARIAÇÃO ENTRE GÊNEROS DISCURSIVOS

**Alunas: Rubiane Guilherme Valério
Márcia Gonzaga de Brito
Orientadora: Lúcia Pacheco de Oliveira**

Introdução

Este projeto visa a compilação de um *corpus* representativo do português do Brasil (CORPOBRAS-PUC-Rio) através da coleta de vários gêneros do discurso oral e escrito. A ampliação desse corpus tem sido feita através da compilação e incorporação de textos de gêneros como: artigos científicos, cartas, crônicas, circulares, contos, editoriais, e-mails, discursos políticos, notícias de jornal, romances, roteiros de cinema; entrevistas, reuniões de negócios, conversas face a face e atendimentos de serviço. Quanto à seleção de textos, visa-se obter uma amostra equilibrada em cada gênero, sobre temas diversos, sendo cada texto produzido por um escritor/falante diferente. Nesta etapa do projeto, o corpus atingiu 660 mil palavras, a serem utilizadas para o estudo da variação entre gêneros discursivos.

Objetivos

O objetivo principal desta fase do projeto foi a compilação e análise de artigos científicos das áreas de Linguística e Nutrição, visando o aumento significativo do corpus e o estudo da variação de um gênero em duas áreas disciplinares. Ainda com o propósito de compilação e estudo do discurso acadêmico, redações do vestibular foram também incorporadas ao corpus.

Outro objetivo do projeto concentrou-se na leitura de textos teóricos sobre a metáfora gramatical (Halliday, 1994) e análise de nominalizações e processos nos textos compilados (Heyvaert, 2003).

Metodologia

Nesta etapa do projeto, a compilação de textos foi feita através da coleta de artigos científicos das áreas de Linguística e Nutrição, a partir do *site* 'Periódicos CAPES'. Ao todo, foram compilados e incorporados ao corpus 24 artigos, de 12 diferentes periódicos, o que corresponde aproximadamente a um acréscimo de 130 mil palavras¹. As redações de vestibular, que correspondem a cerca de 30.000 palavras, foram coletadas em uma universidade pública e duas particulares. No que diz respeito à organização do corpus, seguiu-se as convenções internacionais de codificação de textos, e os artigos científicos e as redações foram codificados por uma sigla que indica o gênero, o número do texto e a língua nos quais os textos foram produzidos.

Para a análise dos artigos científicos, fez-se uma seleção de amostras, cada uma com cerca de 1.000 palavras, distribuídas em introdução, metodologia, resultados e discussão. A análise aplicada aos textos concentrou-se em identificar e quantificar nominalizações, ou seja, formas nominais transformadas a partir de verbos. Nominalizações com sufixos em -

¹ Esta coleta incluiu também textos em inglês (N=12), os quais foram contrastados aos textos em português (N=12).

ção,ssão,-mento,-cia,-dor foram identificadas nos textos em português com auxílio de *software* de buscas em contexto, MonoConc Pro. Nos textos em inglês, foram identificadas nominalizações em *-tion,-ment,-a-ence, -er*. Processos, ou seja, formas verbais, também foram identificados nos textos. As frequências foram normatizadas em 1.000 palavras, médias calculadas e comparadas para contrastar os textos. Nas redações do vestibular também foram identificadas e quantificadas as nominalizações e os processos.

Com a finalidade de tornar acessíveis os dados dos textos para possíveis consultas futuras, foi redigido um relatório com informações relevantes sobre os mesmos, tais como número de palavras, data de publicação e área de pesquisa.

Conclusão

Nessa fase do projeto acrescentou-se ao corpus um total de 160 mil palavras em textos dos gêneros artigo científico e redações do vestibular. A compilação dos textos trouxe, conseqüentemente, a ampliação do CORPOBRAS PUC-Rio.

Os resultados da pesquisa contrastiva de artigos científicos em Português e Inglês apontam para uma variação significativa (* $p < .05$) quanto à língua, havendo mais nominalizações nos textos em português, tanto na área de Linguística como de Nutrição. Nos textos em língua inglesa, as nominalizações são mais frequentes em Nutrição do que em Linguística. Quanto aos processos, estudos estão sendo desenvolvidos, a fim de se comparar, futuramente, o seu uso com o das nominalizações em artigos científicos. Nas redações de vestibular, onde os processos já foram analisados, foram identificadas mais nominalizações do que processos, indicando que estes textos são mais caracterizados por ações do que abstrações.

Referências

- 1- HALLIDAY, M.A.K (1994). *An introduction to functional grammar*. London: Edward Arnold. 2ª ed.
- 2- HEYVAERT, L. (2003). Nominalization as grammatical metaphor: On the need for a radically systemic and metafunctional approach. In A. Simon-Vandenberg, M. Taverniers & L. Ravelli, (Eds.) *Grammatical metaphor: Views from systemic functional linguistics*, pp. 66-99. John Benjamins: Amsterdam.
- 3- SIMON-VANDENBERGEN, A., Taverniers, M & Ravelli, L. (Eds.) (2003). *Grammatical metaphor: Views from systemic functional linguistics*. John Benjamins: Amsterdam
- 4- BIBER, D., Conrad, S. e Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press
- 5- SARDINHA, T. (2004). *Introdução à Linguística de Corpus*. São Paulo. Manole.