

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 e análise de um corpus representativo do português do Brasil (CORPOBRAS PUC-Rio) através da coleta de vários gêneros do discurso oral e escrito e do estudo da variação entre eles. 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 atendimento de serviço. Quanto à representatividade e extensão do corpus, procura-se obter uma amostra de textos 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, que poderão servir de base para o estudo da variação entre gêneros discursivos.

Objetivos:

O objetivo principal desta fase do projeto foi a compilação, organização 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. Além disso, buscou-se, através da compilação de um corpus paralelo de artigos científicos em inglês, estudar a variação deste gênero em duas línguas (Moraes, 2005). Ainda com o propósito de compilação e estudo do discurso acadêmico, redações de vestibular foram acrescentadas ao corpus (Caldeira, 2006).

Pressupostos teóricos:

O embasamento teórico para a análise destes gêneros discursivos foi desenvolvido a partir da leitura de textos sobre a metáfora gramatical (Halliday, 1994), nominalizações e processos (Heyvaert, 2003, Ravelli, 2003, Simon-Vandenberg, 2003). Em fases anteriores do projeto tinham sido iniciadas leituras teóricas sobre estes aspectos, visando análises de textos do corpus. Atualmente, com base nesse instrumental teórico, investigou-se o uso do que Halliday denomina 'nominalizações dos processos' e verificou-se a frequência dessa metáfora nominal no discurso acadêmico.

Visando uma maior inserção acadêmica das bolsistas na pesquisa e a integração Graduação e Pós-Graduação, as duas alunas assistiram aulas em disciplinas de Pós-Graduação, 'Linguística Aplicada' e 'Linguística de Corpus', ministradas pela orientadora em 2006, onde foram lidos e discutidos textos que dão suporte teórico ao projeto (Sardinha, 2004, Teubert, 1996, Biber, Conrad & Reppen, 1998) e/ou mostram resultados de pesquisas

baseadas no corpus compilado, o CORPOBRAS (Oliveira & Dias, 2006, Caldeira, 2006, Moraes, 2005).

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, em textos completos, de 12 diferentes periódicos e 139 redações de vestibular, coletadas em uma universidade pública e duas privadas.

No que tange 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 e o número do texto, a língua na qual o mesmo foi produzido e se foram redigidos por um falante nativo ou não.

Para análise dos artigos científicos, fez-se uma seleção de amostras de aproximadamente 1.000 palavras cada uma, 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 sufixo em *-ção/ssão, -mento, -cia, -dor* foram identificadas nos textos em português com o auxílio de software de buscas em contexto, MonoConc Pro.

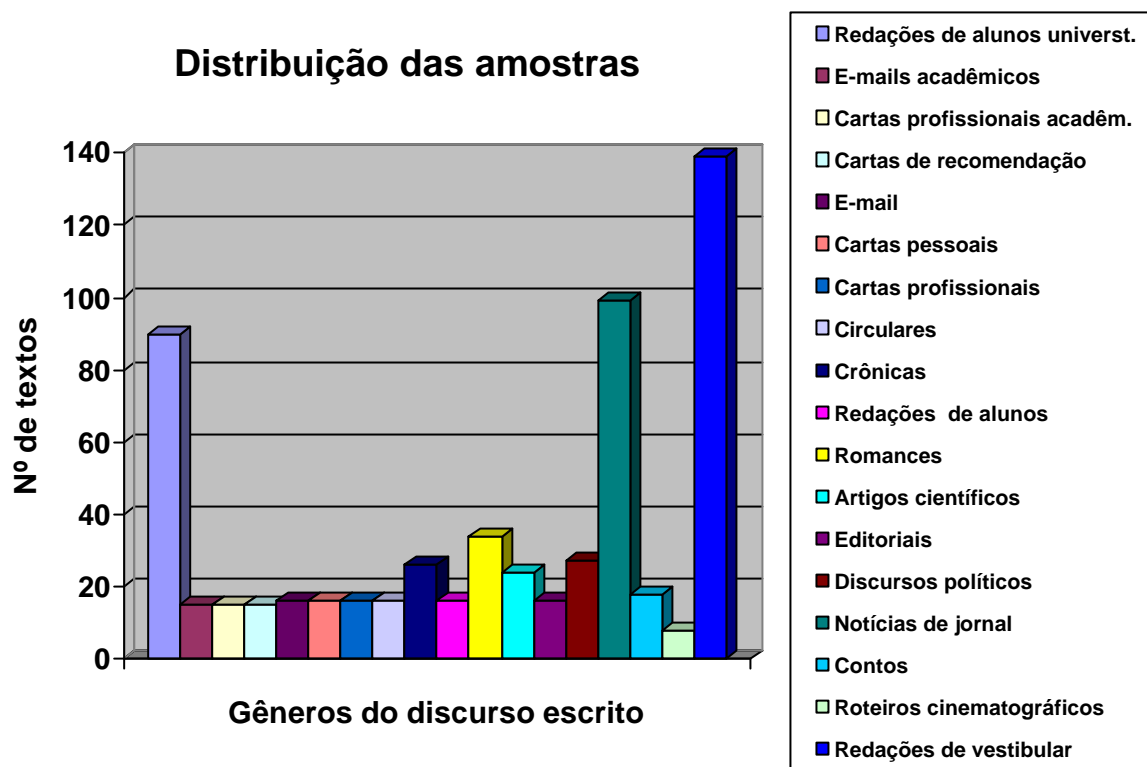
Por se tratar de uma análise contrastiva, nos textos em inglês foram identificadas nominalizações com sufixos paralelos aos selecionados em português, a saber: *-tion, -ment, -a/ence, -er*. Processos, ou seja, formas verbais, também foram identificados nos textos. Todas as frequências foram normatizadas em 1.000 palavras, médias calculadas e comparadas para contrastar os textos. Nas redações de vestibular também foram identificadas e quantificadas as nominalizações e os processos.

Com a finalidade de tornar acessíveis os dados dos textos para consultas futuras, foram redigidos dois relatórios, um para os artigos científicos, contendo as seguintes informações: autor, área de pesquisa, periódico, língua, código, data de publicação e número de palavras, referentes tanto aos textos completos quanto às amostras selecionadas para análise; e outro relatório para as redações de vestibular, com informações sobre a instituição de ensino, o ano do concurso, curso, código do texto e nota.

Resultados:

Nessa fase do projeto acrescentou-se ao corpus um total de 160 mil palavras em textos dos gêneros artigo científico e redações de vestibular. A compilação dos textos levou, conseqüentemente, à ampliação do CORPOBRAS PUC-Rio quanto ao discurso escrito. O Gráfico 1, abaixo, mostra a distribuição dos textos no corpus.

Gráfico 1: Distribuição dos textos do discurso escrito



Os resultados da pesquisa contrastiva de artigos científicos em Português e Inglês apontam para uma variação significativa (* $p < 0.5$) quanto à língua, conforme a Tabela 1, a seguir:

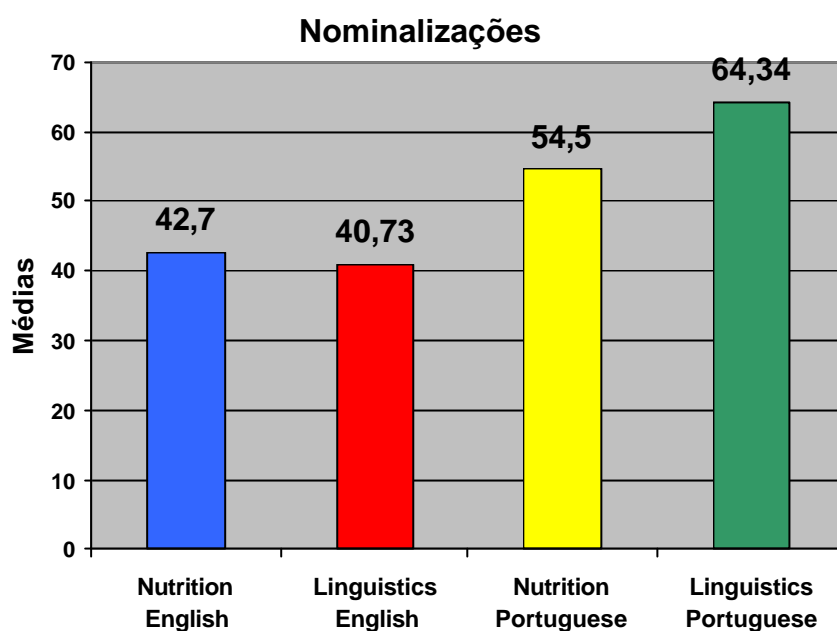
Tabela 1: Análise da Variação : Testes Multivariados (MANOVA)

<i>Efeito</i>	<i>Teste</i>	<i>Sig de F</i>
Língua	Pillais	.048*
Área	Pillais	.201
Área E Língua	Pillais	.102

Valores significativos * $p < .05$

A comparação das médias de frequência de nominalizações mostra que houve mais nominalizações nos textos em português, tanto na área de Linguística como de Nutrição. Nos textos em língua inglesa, as nominalizações são ligeiramente mais frequentes em Nutrição do que em Linguística, conforme mostra o Gráfico 2, abaixo:

Gráfico 2: Nominalizações nas línguas e nas áreas disciplinares



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, em que os processos já foram analisados, foram identificadas mais nominalizações do que processos, indicando que estes textos são mais caracterizados por ações do que abstrações.

Os resultados das pesquisas referentes a esta fase do projeto (2005-2006) foram apresentados em congressos internacionais: American Association for Applied Linguistics/Canadian Association for Applied Linguistics (Oliveira, 2006a), International Systemic-Functional Congress (Oliveira, 2006b, Caldeira & Oliveira, 2006); Rede Panlatina de Terminologia (Oliveira & Dias, 2006).

Referências:

- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Caldeira, J. R. (2006). A redação de vestibular como gênero: Configuração textual e processo social. Dissertação de Mestrado, Departamento de Letras, Rio de Janeiro, PUC- Rio.
- Caldeira, J. R. & Oliveira, L. P. (2006). O mundo criado em um corpus de redações de vestibular: Nominalizações vs. processos em uma perspectiva sistêmico-funcional. Trabalho apresentado no International Systemic-Functional Linguistics Congress, julho, São Paulo, PUC-SP.
- Halliday, M.A.K.(1994). *An introduction to functional grammar*. London: Edward Arnold. 2ª ed.
- Heyavert, 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.
- Moraes, L. S. B. (2005). O metadiscorso em artigos acadêmicos: Variação intercultural, interdisciplinar e retórica. Tese de Doutorado, Departamento de Letras, Rio de Janeiro, PUC-Rio
- Oliveira, L. P. (2006a). Grammatical metaphor in research articles: Linguistic and disciplinary contrasts. Trabalho apresentado na American Association for Applied Linguistics and the Canadian Association for Applied Linguistics Conference (AAAL/CAAL), Montreal, Canadá.
- Oliveira, L.P.(2006b). Grammatical metaphors in academic texts: Cross-linguistic and cross-disciplinary contrasts'. Trabalho apresentado no International Systemic-Functional Linguistics Congress, julho, São Paulo, PUC-SP.
- Oliveira, L.P.& Dias, M.C.P. Representatividade na compilação de um corpus: CORPOBRAS PUC-Rio. Rede Panlatina de Terminologia: Jornada sobre 'Metodologia para a recolha e sistematização de corpora para fins dicionarísticos', maio, PUC-Rio.
- Ravelli, L. J. (2003). Renewal of connection: Integrating theory and practice in an understanding of grammatical metaphor. In A. Simon-Vandenberg, M. Taverniers & L. Ravelli, (Eds.) *Grammatical metaphor: Views from systemic functional linguistics*, pp. 37-64. John Benjamins: Amsterdam.
- Sardinha, Tony B. (2004). *Linguística de Corpus*. Manole: São Paulo.
- Simon-Vandenberg, A., Taverniers, M. & Ravelli, L. (Eds.)(2003). *Grammatical metaphor: Views from systemic functional linguistics* . John Benjamins: Amsterdam.
- Teubert, W. (1996). Editorial. *International Journal of Corpus Linguistics*, Vol.1, No. 1. iii-x.