

**CORPOBRAS PUC-Rio:
Desenvolvimento e análise de um corpus representativo do português**

**Aluna: Rubiane Guilherme Valério¹
Orientadora: Lúcia Pacheco de Oliveira**

Introdução

Este estudo, desenvolvido no âmbito do PIBIC, está vinculado ao projeto de pesquisa 'Compilação de um corpus do português do Brasil e análise multidimensional da variação dos gêneros discursivos'(CORPOBRAS PUC-Rio), do Programa de Pós-Graduação em Letras, tendo contado com apoio do Edital Universal CNPq n°019/2004, até 2007.

A presente pesquisa, que se insere na área de Lingüística de Corpus (Biber et al, 1998), tomou como base para estudos lingüísticos a teoria sistêmico-funcional (Halliday, 1994). Nesta etapa do projeto, ênfase especial foi dada às atividades de análise de gêneros do discurso acadêmico, enfocando-se o estudo do uso da metáfora gramatical (Simon-Vanderbergen et al, 2003) em textos produzidos por pós-graduandos. A metáfora gramatical implica, dentre outros aspectos, na transformação de idéias mais concretas em mais abstratas, através do uso de nominalizações em lugar de processos verbais e, por esta razão, pode ser considerada como um elemento fundamental para a elaboração de textos produzidos em contextos pedagógicos, podendo constituir-se em dificuldade específica para o domínio da escrita em contextos universitários e escolares.

Visando aprofundar os estudos sobre a escrita em português, na universidade, e caracterizar seu processo de transformação através do uso da metáfora gramatical, este projeto enfocou a compilação e análise de dissertações e teses produzidas por alunos de pós-graduação, em diferentes fases do seu desenvolvimento. A fim de investigar como mestrandos e doutorandos, em sua escrita, transformam idéias mais concretas em mais metafóricas, buscou-se responder as seguintes perguntas de pesquisa: 1) Os alunos de mestrado/doutorado usam a metáfora gramatical na escrita acadêmica? 2) O uso de uma linguagem gramaticalmente mais metafórica aumenta durante o processo de escrita da dissertação/tese de mestrado? 3) Os textos acadêmicos de pós-graduandos tornam-se mais adequados através do uso de nominalizações? (Heyvaert, 2003, Martin, 1993).

Aspectos Metodológicos

Nesta etapa do projeto, foram selecionadas para o corpus e analisadas 4 dissertações de mestrado e 2 teses de doutorado da área de Estudos da Linguagem produzidas por alunos do Departamento de Letras da PUC-Rio, que cursaram a graduação em universidades particulares e públicas, de diferentes estados do país. A análise foi baseada em 32 amostras de textos correspondentes a três diferentes versões das introduções e conclusões dessas dissertações, respectivamente, uma versão inicial, uma intermediária e uma final, produzidas durante o processo de escrita do trabalho de pesquisa.

¹ No período de janeiro a dezembro de 2007, o projeto contou também com a participação de uma bolsa de Iniciação Científica da FAPERJ, concedida para a aluna Márcia Gonzaga de Brito.

As diferentes versões foram comparadas através da análise de traços linguísticos que representam nominalizações. Para calcular a frequência média dessas nominalizações em cada versão das dissertações e teses, foram identificados os seguintes sufixos com o auxílio do software *WordSmith Tools* (Scott, 2004): -ção, -ções, -ssão, -ssões, -mento, -mentos, -cia, -cias, -dor e -dores. A análise automática dos textos foi desenvolvida através da ferramenta *Concordancer*, que faz buscas em contexto de palavras, ou partes de palavras, como sufixos, e indica a frequência das mesmas em um corpus selecionado. A figura 1, abaixo, ilustra uma janela de concordância para um texto, com sufixo de busca -ção. A frequência dos sufixos foi calculada em 32 textos e os resultados quantitativos foram comparados para verificar a variação do uso da metáfora gramatical quanto à parte do texto, ou seja, introdução e conclusão, e à versão das mesmas, ou seja, inicial, intermediária e final.

Figura 1 – Janela de concordância para um texto, com sufixo de busca *ção

	Concordance	Set	Tag	Word #	t.
1	este contexto deve ser entendido como qualquer outro ambiente de interação social – e encontro face a face – com suas características particulare			83	:
2	isa que a sala de aula de qualquer natureza caracteriza-se pela manifestação de uma variedade de discursos que convivem em harmonia, fazendo d			121	!
3	deste contexto uma unidade sistêmica. Tal unidade é composta pela relação de interdependência entre as dimensões social, afetiva e cognitiva, co			142	!
4	avés da prática discursiva pedagógica, especificamente a partir da investigação das conversas periféricas, manifestação discursiva do tripé estrutural			193	!
5	eficacemente a partir da investigação das conversas periféricas, manifestação discursiva do tripé estrutural da sala de aula. As análises realizadas a			197	!
6	s a respeito das conversas periféricas revelaram que os agentes da interação escolar levam para sala de aula o que vivem fora da mesma, como po			219	!
7	ntes sociais atuando em um contexto social específico. Quanto à construção do conhecimento, as conversas periféricas se revelaram importante in			284	!
8	tivando, assim, construir significativamente seu processo de sócio-construção do conhecimento. Deste modo, as conversas periféricas podem ser e			342	!
9	as como uma forma de aprendizagem criada por alunos durante a construção de seu saber. Independentemente de métodos ou teorias de aprendiz			363	!
10	s espontaneamente pelos alunos com o intuito de, a partir de uma correlação com seu mundo particular, apropriar-se, bem como acomodar-se, ao			405	!
11	ntretanto, esta pesquisa apresenta algumas limitações, como a investigação da conversa periférica em um número restrito de grupo de alunos, co			428	!
12	a periférica em um número restrito de grupo de alunos, com pequena variação de idade, onde apenas um grupo de cada faixa etária foi selecionado.			442	!
13	línguas, sugiro como desdobramentos desta pesquisa que uma investigação das conversas periféricas seja realizada em um maior número de cont			521	!
14	m salas de aulas de diferentes disciplinas, outra faixa etária e maior variação no número de meninos e meninas participantes. Esta extensão do es			562	!
15	nversal em sala de aula de língua estrangeira, acredito que esta investigação pode contribuir para a prática de pedagógica, já que identificar e ente			619	!
16	tender as conversas periféricas pode gerar um processo de sócio-construção de conhecimento mais significativo a todos os envolvidos nesta prác			640	!
17	locais, ou seja, em contextos pedagógicos específicos, onde cada interação é única e particular a um determinado grupo de alunos com seu profe			677	!
18	m procurando, a partir de então, entender as diferentes formas de construção do conhecimento adotadas por alunos em sala de aula. Finalmente,			758	!

Resultados

(a) Ampliação do corpus:

Através da seleção e da compilação de textos predominantemente do discurso acadêmico, o CORPOBRAS PUC-Rio, no último ano, ultrapassou a meta de 1.000.000 (hum milhão) de palavras, equiparando-se a corpora internacionais considerados como médio-grandes, em relação ao seu tamanho (Valério et al, 2007). Atualmente, o corpus é composto por 26 (vinte e seis) gêneros discursivos, assim distribuídos: 19 (dezenove) gêneros do discurso escrito, 5 (cinco) gêneros do discurso oral, e 2 (dois) gêneros do discurso escrito para ser falado. O corpus totaliza 1.361 textos e 1.149.600 palavras (ver Anexo 1, Distribuição dos gêneros e textos do CORPOBRAS).

(b) Análise do corpus:

Um dos objetivos centrais desta fase do projeto foi aprofundar a análise de dissertações de mestrado e teses de doutorado produzidas por pós-graduandos da área de Letras, durante o seu desenvolvimento. Resultados da análise de 32 amostras, mostraram que a metáfora gramatical, investigada através de nominalizações, foi utilizada nos textos selecionados, nas introduções e conclusões, e que seu uso tende a aumentar durante o processo de escrita dos mesmos, principalmente das primeiras para as segundas versões. Percebeu-se também que nas amostras examinadas houve uma maior tendência para utilizar as nominalizações na forma singular do que no plural. Verificou-se ainda que, como em pesquisas anteriores sobre outros gêneros discursivos (Oliveira, 2006), os sufixos formadores de nominalizações mais frequentes nestes textos foram *-ção*, *-mento*, e *-cia* (ver Anexo 2, Tabela 1: Uso de sufixos em amostras de textos). A discussão destes resultados mostra que o uso das nominalizações parece estar relacionado com a densidade lexical das amostras, que aumenta durante o processo de produção da escrita das dissertações e teses, o que torna estes textos nominalmente mais complexos e adequados ao discurso acadêmico da área de conhecimento à qual pertencem.

(c) Participação da bolsista em atividades de pesquisa:

1) Grupo de pesquisa:

Grupo de Pesquisa do CNPq (Gr Pesp): ‘Linguística Sistêmico-Funcional, Linguística de Corpus e Análise do Discurso’, coordenado pela orientadora, e composto por alunos de iniciação científica, mestrado e doutorado do programa de pós-graduação da Pontifícia Universidade Católica do Rio de Janeiro.

2) Elaboração de projeto:

Elaboração de um projeto de pesquisa para aprimoramento e ensino de produção escrita em língua portuguesa em contextos escolares, a ser submetido, em 2008, ao Edital no.26 da Fundação de Apoio à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

3) Cursos:

2007.2 – Curso de Pós-Graduação ministrado pela orientadora: ‘Teoria Lingüística e Descrição do Português: Análise Sistêmico-Funcional aplicada ao Português’. PUC-Rio.

2008.1 – Curso de Graduação ministrado pela orientadora: ‘Análise Textual da língua Inglesa’. PUC-Rio.

(d) Produção científica:

1) Trabalhos publicados:

OLIVEIRA, L. P. (2007). Writing in the academic context: A corpus-based contrastive view. In: S. Zyngier, V. Viana & J. Jandre (Orgs.), *Textos e Leituras: Estudos Empíricos de Língua e Literatura*, pp. 53-64. Rio de Janeiro: Publit.

OLIVEIRA, L. P. (2007) . Letras. In: V. Daher (Org.), *Guia Megazine de Profissões*, pp.94-95. Rio de Janeiro: Ediouro. (Reportagem e pesquisa de Ediane Merola e Josy Fischberg).

VALÉRIO, R.G., BRITO, M.G. & OLIVEIRA, L.P. (2007). CORPOBRAS PUC-Rio: Um corpus do Português do Brasil e análise do discurso acadêmico. *Caderno de Resumos do VII Encontro da Ciência Empírica de Letras (VII ECEL)*, p.85. Rio de Janeiro: UFRJ,

2) Trabalho no prelo:

OLIVEIRA, L. P. (no prelo). Business Discourse in Brazil: Context, genres and practices. In F. Bargiela-Chiappini (Ed.), *The Handbook of Business Discourse*. Edinburgh: Edinburgh University Press.

3) Trabalhos apresentados:

VALÉRIO, R.G. & BRITO, M.G. (2007). Compilação e análise de um corpus representativo do português do Brasil. Trabalho apresentado na VI Semana de Letras do Departamento de Letras da PUC-Rio, Rio de Janeiro.

VALÉRIO, R.G., BRITO, M.G. & OLIVEIRA, L.P.(2007). CORPOBRAS PUC-Rio: Um corpus do Português do Brasil e análise do discurso acadêmico. Trabalho apresentado no VII Encontro da Ciência Empírica de Letras (VII ECEL). UFRJ: Rio de Janeiro.

4) Trabalhos aceitos para publicação em periódicos:

OLIVEIRA, L. P.. Involvement variation in the writing of academics: A cross-cultural analysis of three genres. *International Journal of Corpus Linguistics*. Amsterdam: John Benjamins.

OLIVEIRA, L. P.. Cross-cultural contrasts in discourse styles: Complexity-level variation in writing. *Language in Contrast*. Amsterdam: John Benjamins.

5) Trabalho aceito para apresentação em evento:

OLIVEIRA, L. P.. Construing experience in academic writing: A cross-disciplinary corpus-based study. 4º Congresso da Associação de Lingüística Sistêmico-Funcional da América Latina (ALSFAI), Outubro, 2008.

Referências

BIBER, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press

HALLIDAY, M.A.K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

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. Amsterdam: John Benjamins.

MARTIN, J. R. (1993). Genre and literacy: Modeling context in educational linguistics. *Annual Review of Applied Linguistics*, 13, 141-172.

OLIVEIRA, L.P. (2006). Grammatical metaphor in research articles: Linguistic and disciplinary contrasts. Trabalho apresentado na American Association for Applied Linguistics Annual Conference (AAAL), Montreal, Canada.

SCOTT, M. (2004). *WordSmith Tools version 4*. Oxford: Oxford University Press.

SIMON-VANDENBERGEN, A., Taverniers, M. & Ravelli, L. (Eds.) (2003). *Grammatical Metaphor: Views from Systemic Functional Linguistics*. Amsterdam: John Benjamins.

VALÉRIO, R.G., Brito, M.G. & Oliveira, L.P. (2007). CORPOBRAS PUC-Rio: Um corpus do Português do Brasil e análise do discurso acadêmico. *Caderno de Resumos do VII Encontro da Ciência Empírica de Letras*. Rio de Janeiro: UFRJ, p.85.

Anexo I
Distribuição dos gêneros, textos e números de palavras no CORPOBRAS PUC-Rio

Discurso Escrito		
Gêneros	Número de Textos	Número de palavras
Artigos científicos	12	63.818
Cartas ao editor	18	1.054
Cartas de reclamação	136	21.417
Cartas de recomendação	31	6.012
Cartas pessoais	16	7.829
Cartas profissionais	16	3.166
Cartas profissionais acadêmico	15	3.529
Circulares	16	2.608
Contos	14	15.253
Crônicas	26	17.434
Dissertações e Teses (Introduções e conclusões)	32	48.143
Editoriais	16	7.931
E-mails acadêmicos	15	1.816
E-mails pessoais	16	1.858
Notícias de jornal	99	40.409
Redações de alunos	16	3.416
Redações de alunos universitários	91	25.065
Redações de vestibular	139	28.646
Romances	28	27.061
Total: 19	Total: 752	Total: 326.465
Discurso Oral		
Conversas cariocas	53	353.678
Conversas de crianças	94	84.573
Entrevistas (acadêmicas)	17	88.769
Grupos de enfoque	7	40.513
Atendimento ao cliente	393	215.671
Total: 5	Total: 564	Total: 783.204
Discurso Escrito para ser Falado		
Discursos políticos	27	22.751
Roteiros cinematográficos	18	17.180
Total: 2	Total: 45	Total: 39.931
Total de Gêneros: 26	Total de Textos: 1.361	Total de Palavras no Corpus em 2008: 1.149.600

Tabela 1 – Uso de sufixos em amostras de textos.

Aluno	Língua	Parte	Versão	ção	ções	mento	mentos	cia	cias
Maria	1	1	1	32,1	5,6	11,2	2,8	12,6	0,0
Maria	1	1	2	42,8	9,3	13,9	6,5	8,4	0,0
Maria	1	1	3	44,0	9,4	14,1	6,6	8,4	0,0
Maria	1	2	1	25,3	3,5	16,1	2,3	6,9	0,0
Maria	1	2	2	27,1	0,6	20,4	0,6	1,2	1,2
Maria	1	2	3	27,0	5,1	22,5	0,0	7,7	1,3
Ana	1	1	1	27,5	7,5	12,0	3,1	6,2	1,3
Ana	1	1	2	27,0	8,4	11,5	3,1	3,5	1,3
Ana	1	1	3	24,7	6,0	12,4	2,8	6,4	1,1
Ana	1	2	1	21,9	6,1	6,1	4,9	0,0	1,2
Ana	1	2	2	22,0	6,1	6,1	4,9	0,0	1,2
Ana	1	2	3	22,9	6,5	6,5	4,4	0,0	1,1