

## 5

### Metodologia

Este capítulo será dividido em duas partes. Na primeira, descreveremos como foi feita a coleta de dados e a composição do corpus, apresentando os critérios adotados para a seleção dos artigos de pesquisa. E na segunda, apresentaremos as etapas do processo de análise do uso do metadiscorso no corpus selecionado, que se deu com o auxílio de cálculos estatísticos.

#### 5.1

##### A coleta de dados e a composição do corpus

Ao decidirmos utilizar artigos acadêmicos como dados para nossa pesquisa, nos deparamos com a seguinte questão: há uma imensa gama de periódicos publicados tanto em inglês quanto em português. Precisaríamos decidir, então, quais escolher. Optamos por trabalhar com periódicos selecionados pela CAPES (Coordenação de Aperfeiçoamento Pessoal de Ensino Superior), encontrados no *site* [www.periodicos.capes.gov.br](http://www.periodicos.capes.gov.br), os quais são indicados por professores de diferentes instituições de ensino superior como sendo importantes para suas respectivas áreas e de grande circulação entre seus alunos.

Nesse *site*, as diversas disciplinas encontram-se divididas em cinco áreas: (1) *Ciências Biológicas e Ciências da Saúde*; (2) *Ciências Biológicas e Ciências Agrárias*; (3) *Ciências Exatas e da Terra e Engenharias*; (4) *Ciências Sociais Aplicadas e Ciências Humanas*; (5) e *Linguística, Letras e Artes*, como pode ser visto no Anexo 1. Neste trabalho, investigamos o uso do metadiscorso nas áreas (1) e (5).

Embora saibamos que cada disciplina, dentro de uma mesma área, pode apresentar características retóricas próprias (Bazerman, 1988/2000), selecionamos a disciplina de Nutrição como representativa da área de *Ciências Biológicas e Ciências da Saúde*<sup>1</sup> e Linguística como representativa da área de *Linguística, Letras e Artes*. Alguns fatores contribuíram para a escolha dessas duas disciplinas para esta pesquisa, incluindo o interesse pessoal que temos por cada uma delas. Sempre nos interessamos por assuntos ligados à Nutrição e Linguística é nossa área de atuação profissional.

Além disso, uma pesquisa piloto que empreendemos sobre o metadiscorso nessas duas disciplinas revelou resultados interessantes (Moraes, no prelo), que nos levaram a querer investigá-las mais profundamente. Outro fator que favoreceu a inclusão de Nutrição e Lingüística nesta pesquisa foi o fato de que a fonte de coleta de dados utilizada para esta investigação, o *site* de periódicos da CAPES, possui uma grande quantidade de periódicos em língua inglesa e um número bem mais reduzido em língua portuguesa. Muitas disciplinas, tanto da área de *Ciências Biomédicas* quanto de *Lingüística, Letras e Artes*, nem sequer possuíam periódicos brasileiros cadastrados na época em que os dados foram coletados<sup>2</sup>. Nutrição foi uma das poucas disciplinas em que encontramos periódicos nas duas línguas. Lingüística, por sua vez, foi a disciplina que apresentou menor número de periódicos no *site*, na ocasião da coleta, despertando, assim, grande interesse pelo seu estudo. Em inglês, na área de Lingüística, havia um número pequeno de periódicos se comparado às outras áreas que o *site* abrange. Em português, só havia uma revista registrada, a *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, que não pôde ser utilizada neste estudo por não conter artigos no formato e período com os quais escolhemos trabalhar: artigos divididos em diferentes seções (ver Capítulo 4, item 4.1.3), publicados entre 2001 e 2003. Procuramos, ainda, outras bases de dados *online* que pudessem nos fornecer periódicos em português, tais como *EBSCOhost (Electronic Journals Service)*, *SciELO (Scientific Electronic Library Online)*, *ProQuest Direct (Academic Research Library)* e *Science Direct (Digital Library of the Future)*<sup>3</sup>, bem como a lista de periódicos da biblioteca da PUC<sup>4</sup>, mas não encontramos nenhum artigo de pesquisa com texto completo *online*, nesta língua. Os dados de Lingüística, em português, então, tiveram que ser digitados de textos publicados em versão impressa.

O fato de *Lingüística, Letras e Artes* se apresentar no site como uma área que possui número reduzido de periódicos disponíveis *online* pode indicar que esta área

<sup>1</sup> No decorrer do trabalho, chamaremos essa área de Ciências Biomédicas.

<sup>2</sup> Os dados foram coletados de outubro de 2003 a fevereiro de 2004.

<sup>3</sup> endereços eletrônicos: <http://ejournals.ebsco.com>; <http://www.scielo.org>; <http://www.proquest.umi.com>; e <http://www.sciencedirect.com>.

<sup>4</sup> endereço eletrônico: [www.dbd.puc-rio.br/periodicos](http://www.dbd.puc-rio.br/periodicos)

parece possuir menor prestígio, acadêmico ou financeiro, do que as demais, o que pode ser corroborado pelo número inferior de trabalhos publicados nesta área, bem como em *Ciências Sociais e Humanas*, em versão digital e impressa, em oposição à grande quantidade de trabalhos publicados em *Ciências Biomédicas* e *Ciências Exatas* (Atkinson, 1999a). O fato, ainda, de haver no site número bem superior de periódicos em língua inglesa em todas as áreas demonstra que tais periódicos parecem possuir maior circulação e credibilidade do que aqueles escritos em língua portuguesa. A discussão dessas questões, nesta pesquisa, poderão contribuir para uma maior valorização e reconhecimento de trabalhos em *Linguística, Letras e Artes*, bem como de trabalhos em língua portuguesa (ver Capítulo 7, item 7.4).

Foram coletados para este estudo 24 artigos acadêmicos, subdivididos da seguinte forma: 12 artigos de *Nutrição* e 12 de *Linguística*, sendo, em cada disciplina, 6 em português e 6 em inglês, como pode ser visto na Figura 5.1, a seguir:

Figura 5.1: Total de artigos coletados por área

ÁREA	PERIÓDICO	NÚMERO DE ARTIGOS	LÍNGUA
Ciências Biomédicas	The Journal of Nutritional Biochemistry	2	Inglês
	Nutrition	2	Inglês
	Nutrition Research	2	Inglês
	Revista Brasileira de Saúde Materno-Infantil <sup>5</sup>	2	Português
	Revista de Nutrição	2	Português
	Revista de Saúde Pública <sup>6</sup>	2	Português
Linguística, Letras e Artes	Brain and Language	2	Inglês
	Linguistics and Education	2	Inglês
	Journal of Second Language Writing	2	Inglês
	Linguagem e Ensino	2	Português
	Pesquisa em Discurso Pedagógico	2	Português
	The ESPecialist	2	Português

### 5.1.1

#### Critérios de seleção dos artigos

Com relação aos critérios de escolha dos artigos a serem utilizados na pesquisa, nos deparamos com algumas questões importantes.

<sup>5</sup> Apesar de esta revista conter artigos de diversas áreas, selecionamos, para esta pesquisa, apenas artigos de Nutrição.

<sup>6</sup> Ver nota 17

Em primeiro lugar, decidimos que seriam incluídos no corpus artigos escritos por autores nativos de português do Brasil e de inglês dos Estados Unidos. No que diz respeito ao português, todas as revistas selecionadas são brasileiras e, portanto, escritas por autores brasileiros, não havendo nenhum problema quanto à sua seleção. Raras vezes encontramos autores portugueses escrevendo para tais revistas, o que foi facilmente identificado pela universidade a que estavam vinculados e pelas características retóricas e lexicais dos textos. Os textos em inglês, contudo, precisaram sofrer uma observação mais detalhada, já que as revistas selecionadas são internacionais, ou seja, escritas por autores de diferentes nacionalidades. Como os periódicos não nos fornecem informações sobre o país de origem dos escritores, mas apenas sobre as instituições a que estão vinculados, escolhemos trabalhar com artigos escritos por autores de instituições norte-americanas, mesmo sabendo que isso não garante, necessariamente, que sejam nativos dos Estados Unidos. Muitos artigos escritos nestes periódicos foram elaborados por vários autores, às vezes chegando a um total de cinco, de diferentes nacionalidades, o que tornou impossível sabermos quem realmente os havia escrito, e por isso não foram incluídos no corpus. Desse modo, decidimos que artigos escritos por autores de universidades norte-americanas representariam os textos escritos em inglês.

Foram selecionados, então, artigos escritos por diferentes autores, de diferentes instituições do Brasil e dos Estados Unidos, a fim de conferirmos ao corpus maior representatividade e evitarmos tendências regionais ou estaduais.

Em segundo lugar, observamos que existem diferentes tipos de artigo dentro dos periódicos. Em nosso trabalho, consideramos apenas os que são relatos de pesquisa (*research articles*), mais especificamente, os artigos com foco mais aplicado e experimental, dispensando, portanto, os que tratam teoricamente de uma determinada área (*review articles*), editoriais, introduções, biografias de autores, resenhas de livros e resumos de teses, depoimentos, críticas, ou qualquer outro gênero que aparecesse nos periódicos.

Com relação ao tamanho do corpus a ser analisado em nosso trabalho, consideramos as seguintes questões: número de textos escolhidos, número de seções de cada texto, e número de palavras de cada seção. Por considerarmos que o número

de textos a serem analisados em uma pesquisa não deve ser muito pequeno, uma vez que "um único texto pode ter uma influência indevida nos resultados da análise" (Biber, Conrad e Rappen, 1994: 249), e baseados em estudos que afirmam que 10 textos são necessários para se obter resultados representativos quanto à ocorrência de traços lingüísticos em um determinado gênero (Biber, Conrad e Rappen, 1998: 249), optamos por trabalhar com 24 textos, de duas áreas, sendo 12 textos de cada língua.

Quanto às seções selecionadas de cada texto, considerando que características lingüísticas variam de acordo com diferentes micropropósitos ao longo das seções internas dos artigos científicos (Conrad, 1996; Biber e Finegan, 1989; Swales, 1990, capítulo 7), escolhemos para análise apenas artigos que se subdividissem em quatro seções, seguindo o formato IMRD de Swales (1990): Introdução, Metodologia, Resultados e Discussão. Observamos, no entanto, que nem sempre essas quatro seções recebiam exatamente estes rótulos nos artigos coletados, mas mantivemos esta nomenclatura na identificação das partes selecionadas por não considerarmos difícil identificar a que seção um determinado título pertencia. Incluímos a Conclusão na seção da Discussão, por ser esta uma prática adotada por vários autores (Swales, 1990; Salanger-Meyer, 1997).

Cabe aqui esclarecer que decidimos considerar cada seção do artigo como um texto diferente, já que apresentam propósitos e convenções retóricas diferentes. Assim, em vez de considerarmos 24 textos para análise, tivemos 24 x 4 (número de seções), ou seja, 96 textos analisados. Esta decisão de pesquisa também facilitou a aplicação de testes estatísticos e a obtenção de resultados específicos para a variação entre as partes do artigo científico.

Foram selecionadas, então, cerca de 250 palavras de cada seção (ver Anexo 6), somando, assim, um total de aproximadamente 24.000 palavras no corpus de 96 textos. Como não tínhamos um número exato e coincidente de palavras por texto, tornou-se necessária para a análise a normatização das frequências, isto é, fazer uma equivalência entre as frequências básicas encontradas para os elementos metadiscursivos e um determinado número de palavras (Biber, 1988: 75-76, Oliveira, 1997: 102-103), de forma a podermos comparar textos maiores e menores, sem que seu tamanho influenciasse nos resultados. Todas as frequências básicas foram assim

normalizadas para 250 palavras (ver Anexo 4: Frequências normalizadas em 96 textos). Desse modo, pudemos comparar o número de itens metadiscursivos por categoria. A fórmula utilizada para esse cálculo é a que se segue:

$\text{Frequência normalizada} = \frac{\text{fator de correção n palavras}}{\text{número de palavras no texto}} \times \text{frequência básica da variável no texto}$
---

Para melhor compreendermos a fórmula utilizada para a normalização de frequências, tomemos o texto 5 como exemplo, onde foram identificadas 6 ocorrências de conectivos lógicos. Dividimos 250 (fator de correção<sup>7</sup>) por 272 (número de palavras no texto) e multiplicamos esse resultado por 6 (frequência básica de conectivos lógicos). A frequência básica inicial (6 ocorrências) foi normalizada para 5,51, e este foi o valor considerado para a categoria dos conectivos lógicos nesse texto. O mesmo foi feito nos demais textos, para cada categoria de itens metadiscursivos.

## 5.2

### Etapas do processo de análise

A análise dos dados selecionados para esta pesquisa se deu em diferentes etapas. Primeiramente, selecionamos os artigos de periódicos brasileiros e internacionais via internet e digitamos os não encontrados no site da CAPES ou em outras bases de dados. Em seguida, separamos cerca de 250 palavras de cada seção dos textos (Introdução, Metodologia, Resultados e Discussão) para serem analisadas.

Para identificarmos os itens metadiscursivos que representam as categorias analisadas, pretendíamos utilizar uma ferramenta automática de busca de palavras em contexto, chamada *Concordancer*, que identifica e contabiliza o número de ocorrências de itens lexicais. Esta ferramenta nos forneceria não só o número de vezes em que cada palavra aparecesse no corpus, mas os contextos em que ocorria, ou seja, todas as sentenças em que estivesse presente. O *Concordancer*, no entanto, não pôde ser utilizado para a análise de nossos dados, uma vez que não havia uma lista

fechada de elementos metadiscursivos, sendo o esquema classificatório proposto para a pesquisa um modelo aberto. Desse modo, novos itens foram sendo acrescentados à classificação de acordo com o que era identificado nos textos. Se um novo item, por exemplo, fosse identificado no texto 10, uma volta aos 9 textos anteriores era feita a fim de verificar se este item aparecia ou não em tais textos. Todas as palavras ou expressões dos textos foram, então, identificadas e analisadas caso a caso, manualmente, com base nas funções que desempenhavam no discurso. É o caso do conectivo lógico *e*, que, conforme mencionado anteriormente (Capítulo 4, item 4.2.2.1), só foi considerado quando tinha a função de ligar orações ou períodos. Foi necessária, então, uma observação detalhada de cada texto para podermos identificar os itens relevantes para o nosso trabalho. Assim, nossa análise não foi automática, mas totalmente manual.

Após a identificação e classificação dos itens metadiscursivos, foi feita a tabulação dos resultados, que será melhor explicitada a seguir.

### **5.2.1**

#### **Identificação dos itens metadiscursivos**

Primeiramente, analisamos separadamente cada texto. Para isso, listamos o número de itens metadiscursivos encontrados em cada um dos 96 textos. Por exemplo, no texto 3, foram encontradas, dentro da categoria dos conectivos lógicos, **3** ocorrências de *e*, **1** de *embora* (ver Anexo 2). Somando todas as ocorrências, chegamos ao total de conectivos lógicos encontrados neste texto. O mesmo foi feito para todas as outras categorias em cada texto. Através dessa tabulação, pudemos contar as ocorrências de itens metadiscursivos em cada texto, agrupando-os em categorias e calculando o total por categoria, por texto.

### **5.2.2**

#### **Tabulação dos itens metadiscursivos**

Na segunda etapa da tabulação dos resultados, elaboramos uma tabela em planilha Excel, com a frequências básicas dos itens metadiscursivos por categoria, em

---

<sup>7</sup> O fator de correção escolhido foi equivalente a 250 palavras, que corresponde, aproximadamente, à

cada texto (ver Anexo 3). Nesta tabela, são considerados os 96 textos; a *língua* em que foram escritos, português e inglês; a *área* em que estão inseridos, Linguística e Nutrição; a *parte* (ou seção) do texto à qual cada texto corresponde, Introdução, Metodologia, Resultados e Discussão; o número de itens encontrados em cada categoria (por exemplo, no texto 1, **4** conectivos lógicos, **0** enumeradores, etc.), o total de itens metadiscursivos encontrados em cada texto (**20** itens no texto 1); e o número de palavras de cada texto (**248** palavras no texto 1).

Como os textos analisados são de tamanhos diversos (ver Capítulo 5, item 5.1.1), elaboramos, numa terceira etapa, uma tabela semelhante à anterior, porém com as frequências normalizadas para cada categoria (ver Anexo 4).

A partir desta tabela normalizada, foi elaborada, com o auxílio do software Excel, uma tabela dinâmica (ver anexo 5), que nos dá a frequência de traços metadiscursivos em cada categoria, de diferentes formas: através da soma a) de cada categoria por parte do texto, em linguística, em português; (b) de cada categoria em todos os textos de linguística, em português; (c) de cada categoria por parte do texto, em nutrição, em português; (d) de cada categoria em todos os textos de nutrição, em português; (e) de todos os textos de linguística e nutrição em português. A mesma sequência foi seguida para os textos em inglês. A tabela dinâmica nos forneceu, ainda, a soma de cada categoria em todos os textos, nas duas línguas. Através desta tabela, então, pudemos visualizar melhor as variações de frequências quanto à língua, à área e à parte do texto.

### 5.3

#### **Testes estatísticos: descrição, aplicação e interpretação**

Para caracterizar artigos de pesquisa escritos por brasileiros e norte-americanos em relação ao uso do metadiscorso, com base em resultados estatísticos, é essencial que se estabeleça se os valores com os quais se está trabalhando são estatisticamente significativos, ou seja, se as diferenças encontradas para os grupos de textos não ocorreram por acaso (Oliveira, 1997: 115). A MANOVA, que permite

---

média do número de palavras nos 96 textos (ver Anexo 3).



verificar se as diferenças entre as médias de frequência são significativas, confere, assim, ao estudo mais confiabilidade aos resultados estatísticos.

Testes univariados (ANOVA) foram também aplicados ao corpus para verificar o valor significativo de cada categoria analisada. Foram aplicados testes univariados para cada um dos efeitos (ou variáveis independentes) – língua, área e parte do texto – que poderiam influenciar a variação das categorias metadiscursivas (ou variáveis dependentes)<sup>8</sup>.

Após identificadas as categorias que apresentavam valores significativos, estas foram comparadas de diferentes perspectivas. Percentuais de frequência também foram calculados para estas categorias identificadas como significativas.

A seguir, os resultados quantitativos serão interpretados e discutidos, visando a caracterizar, no corpus, a variação metadiscursiva intercultural, interdisciplinar e retórica.

---

<sup>8</sup> Neste trabalho, *língua, área e parte do texto* correspondem às variáveis independentes, ou seja, variáveis que atuam sobre outras variáveis, chamadas de dependentes. As 18 categorias metadiscursivas representam estas variáveis dependentes, isto é, variáveis sobre as quais atuam as variáveis independentes (ver Brown, 1988: 10; Oliveira, 1997: 90). Assim, *língua, área e parte do texto* atuam sobre as 18 categorias analisadas.