

**CARLOS H. KAUFFMANN**

**O CORPUS DO JORNAL:  
VARIAÇÃO LINGÜÍSTICA, GÊNEROS E DIMENSÕES DA IMPRENSA  
DIÁRIA ESCRITA**

.

**LAEL - PUC/SP**

**SÃO PAULO**

**2005**

## ERRATA

PÁGINA	LINHA	ONDE SE LÊ	LEIA-SE
xii	3 <sup>a</sup> -4 <sup>a</sup>	teóricos-metodológicos	teórico-metodológicos
3	2 <sup>a</sup>	deles	de textos
49	Tabela 2	17.jan.2003	17.fev.2003
64	Nota 85	acgrupados	agrupados
69	23 <sup>a</sup>	teste de geral	teste geral
75	18 <sup>a</sup>	Estudos de Comunicação e a manuais	Estudos de Comunicação, isto é, obras acadêmicas e manuais
91	Nota 110	essa escolha	a escolha do número de fatores
105	Tabela 10	Outros	Outros (inclui 7 textos classificados como “nota de correção”)
130	22 <sup>a</sup>		<b>BICK, E. Gramática constritiva na análise automática da sintaxe portuguesa.</b> p. 91-112. In: BERBER SARDINHA, A. P. (org.). <b>A língua portuguesa no computador.</b> Campinas: Mercado de Letras, 2005.

**CARLOS H. KAUFFMANN**

**O CORPUS DO JORNAL: VARIAÇÃO LINGÜÍSTICA, GÊNEROS E  
DIMENSÕES DA IMPRENSA DIÁRIA ESCRITA**

**Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de MESTRE em Lingüística Aplicada e Estudos da Linguagem, sob a orientação do Prof. Dr. Antonio Paulo Berber Sardinha.**

**LAEL - PUC/SP**

**SÃO PAULO**

**2005**

Banca Examinadora:

---

---

---

Para Martina, Piero e Daniel Henrique

## AGRADECIMENTOS

Esta empresa não seria realizável sem a ajuda de algumas pessoas e instituições, mencionadas abaixo (não é uma lista exaustiva):

- Professor Dr. Tony Berber Sardinha, meu orientador e treinador – jamais sem perder a ternura –, a quem sou enormemente grato por me conduzir até onde cheguei por uma autopista do conhecimento (olha a metáfora!);
- Grupo Folha, que apoiou financeiramente este trabalho com uma bolsa parcial e gentilmente cedeu sua base de textos para compor o corpus da pesquisa;
- Um agradecimento especial ao Professor Dr. Douglas Biber, que teve a gentileza de analisar meus dados iniciais, comentá-los e sugerir caminhos;
- Professor Dr. David Lee, cujas páginas sólidas e palavras encorajadoras me deram segurança e entusiasmo para mergulhar na AMD;
- Professora Dra. Heloisa Collins, que vislumbrou soluções na interpretação dos resultados, e Izabella Martins, pela ajuda providencial;
- Patrícia Trudes da Veiga, Ana Estela de Souza Pinto e Shely Cayetano, que ajudaram diretamente a resolver uma dúvida que não me abandonava;
- José Arthur Fajardo, *designer* oficial da pesquisa e amigo de todas as horas;
- Banco de Dados e equipe – Marina Mariko Nihei, Florência do Céu Pereira e, em especial, Antonio Paulo Carretta –, impossível mencionar todos;
- Orientandos do Tony e afins, a quem agradeço pela infinita paciência em me ouvir: Elias, Giseli, Lindinalva, Adriana, Renata, Maurício, Glau, Denise, Roberto, Claudia, Daniela, Mauro, Dulce, Christina, Helena;
- Maria Eduarda, Andréa, Jacqueline, Fernanda, Totó: vocês me incentivaram;
- Todos do LAEL, pelo apoio: Maria Lúcia, Márcia, Ricardo, Paulo e Zan;
- Wayne Fernandes (Mineiro), revisor e primeiro leitor;
- Mary, Samuel, Gui, Bela, Bia, Roberta, João, sobrinhos: obrigado pela força;
- *Last, but not least*: Silvia, que carregou o piano enquanto a pesquisa era tocada.

Posso escrever “flor lindamente amarela e branca”, embora não deva. Mas, se em vez disso eu escrever **margarida**, ganho uma exatidão de cinema. Mostro em vez de comentar. Elimino os adjetivos e advérbios que se intrometem entre o leitor e a coisa. Se a língua fosse mais rica em substantivos e verbos, não precisaríamos usar tantos adjetivos e advérbios, que embaçam a exatidão e fazem o texto parecer chumbo em lugar de cristal.

Otávio Frias Filho

Antimanual de jornalismo. **Folha de S.Paulo**, 18.nov.1984. Caderno Folhetim, p. 7.

## SUMÁRIO

LISTA DE TABELAS	xi
LISTA DE TEXTOS	xii
LISTA DE ILUSTRAÇÕES	xiii
<b>RESUMO</b>	<b>xiv</b>
<b>ABSTRACT</b>	<b>xv</b>
<b>1 INTRODUÇÃO</b>	<b>1</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>8</b>
2.1 A Abordagem da Lingüística de Corpus	9
2.1.1 A Escola Neofirthiana	14
2.2 A Análise Multidimensional	21
2.2.1 Análise Multidimensional: Conceito de dimensão	25
2.2.2 Multidimensionalidade	27
2.3 Visões sobre o Conceito de Gênero	29
2.3.1 Gênero: origens	30
2.3.2 Visão sistêmico-funcional	30
2.3.3 Análise de gêneros: Swales e Bhatia	33
2.3.4 Gêneros discursivos: Bakhtin	34
2.3.5 Gênero/Registro: Biber	36
2.3.6 A perspectiva dos Estudos de Comunicação	38
2.3.7 Discussão	40
<b>3 METODOLOGIA DE PESQUISA</b>	<b>45</b>
3.1 Objetivo e Questões de Pesquisa	45
3.2 Desenho do Corpus	47

3.2.1	Semana Construída	49
3.2.2	Etiquetagem Morfossintática	51
3.3	Análise Multidimensional: Uma Abordagem	53
3.3.1	Introdução - Análise Multidimensional	54
3.3.2	Etapas da Análise Multidimensional	54
3.3.3	Análise Multidimensional da Língua Inglesa: Um Exemplo	56
3.4	Análise Multidimensional: Uma Prática	61
3.4.1	Organização Preliminar dos Dados	62
3.4.2	Análise Fatorial	65
3.4.2.1	Parâmetros	66
3.4.2.2	Informações Obtidas	68
3.5	Classificação Manual de Gêneros	74
3.5.1	Tipologia dos Gêneros do Discurso Jornalístico: Uma Proposta	75

## **4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS 82**

4.1	Seleção das Variáveis	83
4.2	Definição do Número de Fatores	90
4.3	Matrizes Padrão e Estrutural da Análise Fatorial	91
4.4	Cálculo de Escores de Textos Por Fator	95
4.5	Descrição dos Fatores	97
4.5.1	Textos nos extremos e na média do Fator 1	97
4.5.2	Textos nos extremos e na média do Fator 2	100
4.6	Classificação dos Textos em Gêneros	104
4.7	Cálculo dos Escores Médios dos Gêneros	106
4.7.1	Testes estatísticos <i>post-hoc</i>	109
4.7.2	Focos dimensionais nos gêneros	110
4.8	Dimensões do Texto Jornalístico	112
4.9	Apresentação Combinada das Dimensões: Uma Síntese	115
4.10	Discussão	116

<b>5 CONCLUSÃO</b>	<b>122</b>
<b>REFERÊNCIAS</b>	<b>129</b>
<b>LISTA DE ANEXOS</b>	<b>138</b>
ANEXO A - Lista de Etiquetas do VISL para o Português	139
ANEXO B - Exemplo de Texto Etiquetado pelo VISL	145
ANEXO C - Estatística Descritiva dos Fatores, por Gênero	148
ANEXO D - Textos “Típicos” dos Gêneros	149
ANEXO E - Estatística Descritiva de Frequência das Variáveis, por Gênero	167
ANEXO F - Versão Integral do Texto 6	181

## LISTA DE TABELAS

TABELA 1	- COMPOSIÇÃO E CARACTERÍSTICAS DO CORPUS DE ESTUDO .....	48
TABELA 2	- EDIÇÕES DIÁRIAS DA “FOLHA DE S.PAULO” UTILIZADAS NO CORPUS DE ESTUDO .....	49
TABELA 3	- VARIÁVEIS QUE COMPÕEM A DIMENSÃO 1 (BIBER, 1988).....	59
TABELA 4	- ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS NO CORPUS DE ESTUDO.....	87
TABELA 5	- COMUNALIDADES FINAIS DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES.....	89
TABELA 6	- MATRIZ PADRÃO DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES .....	92
TABELA 7	- MATRIZ ESTRUTURAL DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES.....	94
TABELA 8	- MATRIZ DE CORRELAÇÃO ENTRE OS FATORES DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES .....	94
TABELA 9	- ESCORES PADRONIZADOS DAS VARIÁVEIS DO FATOR 2 EM TEXTO SELECIONADO DO CORPUS DE ESTUDO.....	95
TABELA 10	- NÚMERO DE TEXTOS DO CORPUS DE ESTUDO, CLASSIFICADO POR GÊNERO .....	105
TABELA 11	- TESTES ESTATÍSTICOS DA VARIAÇÃO ENTRE GÊNEROS, POR FATOR .....	109
TABELA 12	- ETIQUETAS DO VISL PESQUISADAS ORIGINALMENTE NO TRABALHO .....	139
TABELA 13	- ESTATÍSTICA DESCRITIVA DO FATOR 1, POR GÊNERO.....	148
TABELA 14	- ESTATÍSTICA DESCRITIVA DO FATOR 2, POR GÊNERO.....	148
TABELA 15	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO ARTIGO .....	167
TABELA 16	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CARTA.....	168
TABELA 17	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CHAMADA .....	169
TABELA 18	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO COMENTÁRIO.....	170
TABELA 19	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CRÍTICA .....	171
TABELA 18	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CRÔNICA .....	172
TABELA 19	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO EDITORIAL .....	173
TABELA 20	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO ENTREVISTA.....	174
TABELA 21	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO NOTA DE CORREÇÃO.....	175
TABELA 22	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO COLUNA DE NOTAS .....	176
TABELA 23	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO NOTÍCIA .....	177
TABELA 24	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO OUTROS .....	178
TABELA 25	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO REPORTAGEM.....	179
TABELA 26	- ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO RESENHA.....	180

## LISTA DE TEXTOS

TEXTO 1	- MÁXIMO ESCORE NEGATIVO - FATOR 1 .....	98
TEXTO 2	- ESCORE MÉDIO - FATOR 1.....	98
TEXTO 3	- MÁXIMO ESCORE POSITIVO - FATOR 1.....	99
TEXTO 4	- MÁXIMO ESCORE NEGATIVO - FATOR 2 .....	100
TEXTO 5	- ESCORE MÉDIO - FATOR 2.....	101
TEXTO 6	- MÁXIMO ESCORE POSITIVO - FATOR 2.....	102
TEXTO 7	- AMOSTRA DE TEXTO ETIQUETADO PELO VISL .....	143
TEXTO 8	- EXEMPLO “TÍPICO” DE ARTIGO .....	150
TEXTO 9	- EXEMPLO “TÍPICO” DE CARTA .....	151
TEXTO 10	- EXEMPLO “TÍPICO” DE CHAMADA .....	152
TEXTO 11	- EXEMPLO “TÍPICO” DE COLUNA DE NOTAS.....	153
TEXTO 12	- EXEMPLO “TÍPICO” DE COMENTÁRIO .....	154
TEXTO 13	- EXEMPLO “TÍPICO” DE CRÍTICA .....	156
TEXTO 14	- EXEMPLO “TÍPICO” DE CRÔNICA .....	157
TEXTO 15	- EXEMPLO “TÍPICO” DE EDITORIAL.....	158
TEXTO 16	- EXEMPLO “TÍPICO” DE ENTREVISTA .....	159
TEXTO 17	- EXEMPLO “TÍPICO” DE NOTA DE CORREÇÃO .....	161
TEXTO 18	- EXEMPLO “TÍPICO” DE NOTÍCIA.....	162
TEXTO 19	- EXEMPLO “TÍPICO” DE “OUTROS GÊNEROS” .....	163
TEXTO 20	- EXEMPLO “TÍPICO” DE REPORTAGEM.....	164
TEXTO 21	- EXEMPLO “TÍPICO” DE RESENHA.....	166

## LISTA DE ILUSTRAÇÕES

FIGURA 1	- ANÁLISE FATORIAL: TELA INICIAL DO SPSS .....	66
FIGURA 2	- INFORMAÇÕES SOBRE AS VARIÁVEIS UTILIZADAS NA EXTRAÇÃO INICIAL DA ANÁLISE FATORIAL .....	84
GRÁFICO 1	- GRÁFICO “SCREE” DA EXTRAÇÃO COM 14 FATORES .....	90
GRÁFICO 2	- ESCORES MÉDIOS DO FATOR 1, POR GÊNERO .....	107
GRÁFICO 3	- ESCORES MÉDIOS DO FATOR 2, POR GÊNERO .....	108
GRÁFICO 4	- DISTRIBUIÇÃO DOS ESCORES DO FATOR 1, POR GÊNERO .....	110
GRÁFICO 5	- DISTRIBUIÇÃO DOS ESCORES DO FATOR 2, POR GÊNERO .....	111
GRÁFICO 6	- DISTRIBUIÇÃO BIDIMENSIONAL DOS ESCORES MÉDIOS, POR GÊNERO .....	115

## RESUMO

O objetivo desta pesquisa é identificar empiricamente semelhanças e diferenças de natureza lingüística nos textos e entre os gêneros de um jornal de expressão do Brasil. Para a sua consecução, foram empregados os recursos teóricos-metodológicos proporcionados pela Lingüística de Corpus (BIBER, 1988; SINCLAIR, 1991; KENNEDY, 1998; BERBER SARDINHA, 2004a).

Coletou-se uma amostra que representa a língua portuguesa do Brasil tal como é utilizada contemporaneamente na imprensa diária escrita de prestígio nacional. Ela se compõe de uma semana construída com sete edições aleatoriamente sorteadas, entre as publicadas em 2003 pela “Folha de S.Paulo”, totalizando um corpus de estudo de 1.431 textos (493.780 palavras).

O corpus jornalístico foi etiquetado automaticamente com o etiquetador morfossintático VISL para a língua portuguesa (BICK, 2005), de alta precisão. A metodologia escolhida para estudar o corpus foi a Análise Multidimensional (BIBER, 1988; LEE, 2000). Seu procedimento estatístico chave, a Análise Fatorial, busca encontrar grupos co-ocorrentes de características e categorias lingüísticas (ou “variáveis”). Das 19 variáveis selecionadas para a extração fatorial inicial, 14 delas compuseram a solução fatorial final. Foram extraídos dois fatores, interpretados em termos de suas funções comunicativas subjacentes – os eixos dimensionais “Narrativo versus Expositivo” e “Argumentativo versus Informativo”.

O trabalho propõe uma tipologia dos gêneros presentes no jornal, baseada em uma revisão da literatura da área de Estudos de Comunicação (MARQUES DE MELO, 1994 e outras fontes). Ela permitiu a classificação do corpus de estudo e a atribuição de escores, com os quais foi possível mapear os gêneros ao longo das dimensões. Essa perspectiva refinou tipologias anteriores, na medida em que ela não é apresentada de forma dicotômica, mas de acordo com as posições que os gêneros ocupam em relação às dimensões identificadas.

## ABSTRACT

The main objective of this research is to empirically identify linguistic similarities and differences found in texts and across genres in a major of daily newspaper. We employed theoretical and methodological resources from Corpus Linguistics (BIBER, 1988; SINCLAIR, 1991; KENNEDY, 1998; BERBER SARDINHA, 2004a).

A corpus representing Brazilian Portuguese language, as used nowadays in up-market national newspapers pages, was collected. It is a result of a constructed week of seven daily editions, randomly chosen from all of the issues of 2003 published by “Folha de S.Paulo”, the largest country’s newspaper. The research corpus has a total of 1,431 texts (493,780 words).

With the help of the high-precision VISL tagger for Portuguese (BICK, 2005), the press corpus was tagged. Our research uses as main methodology Multidimensional Analysis (BIBER, 1988; LEE, 2000). Its key statistical procedure, Factorial Analysis, seeks to find cocurrent groups of linguistic features and categories (the ‘variables’). Nineteen variables were initially selected and fourteen of them remained at the final factor solution. Two factors were extracted, which were then interpreted in terms of their underlying communicative functions: the dimensional axes “Narrative versus Expositive” and “Argumentative versus Informative”.

Based on Communication Studies literature (MARQUES DE MELO, 1994 and other sources), this research also proposed a typology of newspapers genres. This typology supported the genre classification of the research corpus, which formed the basis for the creation of scores and the subsequent genre mapping along the dimensions. Our typology refined previous ones in the sense that it was not based on dichotomies, but rather according to the position the genres belong to in relation to the dimensions.

## 1 INTRODUÇÃO

**Pieguice** – Jornal não é telenovela. O leitor compra jornal para se informar, não para se comover. Os textos jornalísticos devem ser secos, precisos, econômicos e, acima de tudo, exatos. (Folha de S.Paulo, **Manual Geral da Redação**, 2. ed., 1987, p. 93)

O jornal diário impresso brasileiro está há décadas incorporado à cultura de massa e, como ramo fundador da imprensa, é reconhecido como um dos atores sociais com incontestável influência cultural e política. Como veículo de comunicação, o jornal busca estar sempre em consonância com as necessidades de um conjunto mais ou menos amplo de leitores de se informarem, objetiva e/ou subjetivamente. Para atender a essa demanda, o jornal é portador regular de textos que possuem determinados parâmetros de forma e de conteúdo, produtos de um constante processo de consolidação e transformação de natureza histórica, social e cultural.

Os textos do jornal são amplamente utilizados como material didático para ensino de língua portuguesa (estimulados, inclusive, por formuladores de diretrizes pedagógicas, como os Parâmetros Curriculares Nacionais de Ensino Fundamental<sup>1</sup>), para esboçar apenas um dos interesses da área da Linguística Aplicada, uma entre as várias disciplinas que se relacionam com o assunto. Apesar disso, poucas<sup>2</sup> são as abordagens extensivas sobre a linguagem da imprensa escrita – especificamente, a imprensa brasileira atual – que levam em conta as numerosas *variedades* de textos encontradas em um jornal. Esta pesquisa pretende preencher essa lacuna, pois tem o objetivo básico de analisar a linguagem que circula nos jornais brasileiros de hoje. A maneira escolhida de estudar esse exemplo significativo da língua portuguesa será a abordagem teórico-metodológica proporcionada pela *Linguística de Corpus*, que

---

<sup>1</sup> Cf. Brasil/SEF/MEC (1997 e 1998).

<sup>2</sup> Entre as iniciativas que esta pesquisa localizou na literatura estão Vélez (1985), Marques de Melo (1987; 1994; 1998a; 1998b), Chaparro (1997) e Bonini (2004).

propugna observar *padrões* de frequência a partir de um *corpus* que representa uma amostra real da língua (BIBER, 1988; SINCLAIR, 1991; STUBBS, 1996; KENNEDY, 1998; BIBER et al., 1998; BERBER SARDINHA, 2004a).

Essa abordagem difere de pesquisas anteriores que estudaram a linguagem dos jornais na área de Lingüística Aplicada, como a Análise Crítica do Discurso. Van Dijk (1989) e Fowler (1991), trabalhos relevantes na área, extraem do texto jornalístico os componentes ideológicos constituídos na linguagem. De modo similar, Fairclough (1995) adota a teoria sistêmico-funcional (HALLIDAY, 1994) para revelar as práticas discursivas ideológicas subjacentes envolvidas na produção de um texto. Já Bell (1991) preocupa-se com etapas retóricas comuns que emergem da análise morfológica da notícia. No caso deste trabalho, a abordagem enfocará a contagem de frequência de palavras e classes morfossintáticas dos textos do jornal, no intuito de detectar na variação lingüística do corpus os padrões funcionais atuantes, como será explicado em detalhes a seguir.

Existem no país iniciativas recentes que esquadriham a linguagem do jornal em diferentes aspectos, entre elas o projeto Lácio-Web (ALUÍSIO et al., 2003), que considera o jornal e seus gêneros como uma de suas áreas de interesse, e o Projor – Projeto Gêneros do Jornal (BONINI, 2004), que realiza uma análise de gêneros na linha britânica (SWALES, BHATIA) e observa detidamente aspectos como a morfologia da página do jornal e os movimentos retóricos intratextuais. Na área de Estudos de Comunicação, Chaparro (1997) analisa comparativamente, em uma extensa pesquisa de campo, jornais brasileiros e portugueses; Marques de Melo (1998b), por sua vez, realiza um comparativo de gêneros entre jornais diários “up-market”, ou de prestígio nacional, e cotidianos provenientes de cidades do interior do Estado de São Paulo<sup>3</sup>. A análise realizada por esses trabalhos, no entanto, não põe em primeiro plano a descrição lingüística, como é o foco proposto por esta pesquisa.

---

<sup>3</sup> Utilizando, no entanto, como unidade de análise não o texto ou a quantidade de palavras, mas o espaço ocupado (centímetro por coluna).

De um ponto de vista lingüístico, o objeto de estudo de um jornal pode tanto ser qualquer fragmento de um texto, um texto inteiro, uma reunião deles ou mesmo corresponder à soma de todos os textos nele contidos. O conjunto de textos do jornal – que abarcaria *todos* os textos, incluindo publicidade, classificados, listas, apêndices não-periódicos, etc. – não se confunde, porém, com o conjunto formado pelos *textos jornalísticos*, interesse da presente pesquisa. O texto jornalístico corresponde ao material não publicitário, resultante do esforço editorial produzido por um corpo relativamente estabilizado de jornalistas e de colaboradores fixos e ocasionais. Nesse sentido, doravante o termo *jornal* designará somente o conjunto de textos jornalísticos.

Os textos encontrados no jornal, a despeito de sua padronização gráfica (têm harmonia visual), ideológica (obedecem a uma linha editorial) e pragmática (atingem um público amplo), estão longe de representar um conjunto homogêneo. Admite-se, ao contrário, a existência de categorias mais gerais, que reúnem grupos de textos que possuem semelhanças de várias ordens – formal, cognitiva, lingüística, etc. Ou seja, existe no jornal uma variedade de diferentes *gêneros*, convencionados socialmente e praticados em um contexto cultural (SWALES, 1990; BHATIA, 1993).

Se é ponto pacífico que se encontram vários gêneros no jornal, não é tão claro assim se eles são diferenciáveis sob o prisma da linguagem. A linha de investigação a ser conduzida neste trabalho procurará evidenciar diferenças e semelhanças lingüísticas entre os gêneros presentes no jornal e, assim, contribuir para essa discussão.

Para observar a variação lingüística existente no jornal foi coletado um corpus composto por sete edições de um veículo de prestígio nacional (“Folha de S.Paulo”). Esse *corpus de estudo*, que representa uma semana de jornal, soma 1.431 textos. A presunção assumida aqui é a de que essa amostra é representativa da imprensa escrita diária em português do Brasil tal como praticada contemporaneamente.

A base metodológica escolhida aqui para a análise do jornal e seus gêneros é a Análise Multidimensional de variação de registro (sigla AMD) (BIBER, 1988; LEE,

2000). É uma metodologia de observação da língua que permite captar a *variação*, existente na língua, de *características e categorias lingüísticas* em uso. Características e categorias lingüísticas são qualquer elemento *quantificável* encontrado na língua; as *características*<sup>4</sup> podem ser itens lexicais ou gramaticais, simples ou complexos, como os *lemas* (por exemplo, o lema do verbo *ser* corresponde à soma de todas as ocorrências das suas formas verbais flexionadas – *sou, é, és, etc.*). As *categorias*, por seu turno, reúnem classes gramaticais, como substantivos, adjetivos ou tempos verbais, entre outros. Dada a dimensão do corpus de estudo (493.780 palavras), a obtenção das características e categorias lingüísticas exige que os textos sejam submetidos a um processo de *etiquetagem* automática, por meio de programa de computador. A etiquetagem realiza no corpus uma classificação *morfo-sintática*<sup>5</sup>. O programa *on-line* VISL para a língua portuguesa, pelo alto índice de acuidade, foi o utilizado neste trabalho (cf. 3.2.2).

A Análise Multidimensional envolve um processo estatístico – a Análise Fatorial – que lida com múltiplas *variáveis*. Cada variável utilizada na Análise Fatorial corresponde a uma determinada característica ou categoria lingüística selecionada do corpus – as que potencialmente podem ter importância lingüística<sup>6</sup>. A Análise Fatorial considera a frequência das variáveis utilizadas em todos os textos do corpus para buscar *correlações* entre as variáveis e daí extrair um determinado número de *fatores*. Um fator, em outras palavras, enfeixa variáveis (características e categorias lingüísticas) co-ocorrentes. Cada fator, assim, é responsável por uma parcela da variação lingüística observada no corpus de estudo. Segundo Biber (1988), a partir da interpretação dos fatores é possível identificar *funções comunicativas* subjacentes à linguagem, chamadas pelo autor de *dimensões*. As dimensões são produto da análise e

---

<sup>4</sup> Biber (1988) e Berber Sardinha (2000b; 2004a) utilizam também nessa acepção o termo “traço”, ou “feature”.

<sup>5</sup> Em inglês, “POS tagging”; cf. Berber Sardinha, 2004a, p. 113-142.

<sup>6</sup> Cf. Biber, 1988, p.72.

não postas a priori (BIBER 1988, p.13): a co-ocorrência das características/categorias lingüísticas é que faz emergir o fator e, por conseguinte, a dimensão.

Alguns estudos especificamente utilizaram a Análise Multidimensional no Brasil, como Pacheco de Oliveira (1997), Conde (2002), Shergue (2003) e Santos (2003). Esta pesquisa pretende esclarecer algumas questões relativas aos procedimentos metodológicos da AMD em língua portuguesa, não abordadas anteriormente, de forma a facilitar a outros pesquisadores o manejo das ferramentas empregadas na AMD.

Entre os pressupostos teóricos da pesquisa a ser adotados no trabalho estão as idéias pioneiras do lingüista britânico Firth (MONAGHAN, 1979; DE BEAUGRANDE, 1991, p. 187-222) – entre elas, a da distribuição do significado em vários níveis na linguagem. As obras de Halliday (1994) e de Sinclair (1991) são vistas como um desenvolvimento dos conceitos fundadores de Firth e responsáveis por modelar a forma com que a Lingüística de Corpus realiza descrições na linguagem. Um aspecto a ser destacado é o princípio de que o sistema da língua possui uma natureza probabilística (STUBBS, 2001, p. 230; BERBER SARDINHA, 2004a, p. 30). Ou seja, a ocorrência de determinada palavra influencia e limita as escolhas subseqüentes do falante de uma forma estatisticamente mensurável (SINCLAIR, 1991).

Outro ponto teórico importante afirma a natureza *empírica* desta pesquisa (BIBER et al., 1998): busca-se extrair dimensões, que são constructos teóricos, a partir da exploração dos dados de um corpus jornalístico – e não o caminho inverso, que privilegiaria um modelo mental preexistente e seria confirmado (ou não) pelo exame do corpus.

Subsidiando uma reflexão específica sobre a questão do gênero, foram examinadas diversas linhas de análise na área de Lingüística Aplicada. Entre elas, os trabalhos de Swales (1990) e Bhatia (1993), a teoria sistêmico-funcional de gênero e registro (HALLIDAY; HASAN, 1989; EGGINS; MARTIN, 1997) e a teoria de gênero segundo Bakhtin (BAKHTIN, 2000). Algumas idéias utilizadas por Biber (1988)

também são examinadas, como a noção de tipo de texto.

Em complemento, foi feita uma revisão da literatura da área de Estudos de Comunicação<sup>7</sup> sobre os gêneros encontrados no jornalismo brasileiro. A pesquisa constata que não há uma tipologia consensual empregada para denominar os gêneros do jornal. Foram utilizados principalmente os trabalhos de Marques de Melo (1992, 1994), Chaparro (1997, 1998) e Rabaça e Barbosa (1998), além de manuais de jornalismo, para realizar uma compilação de definições sobre cada gênero.

Além do propósito de identificar semelhanças e diferenças lingüísticas no corpus, podem ser divisadas três outras finalidades para os resultados obtidos por esta pesquisa: subsidiar a construção de um modelo para a classificação automática de gêneros, capaz de ser aplicada em outros textos armazenados eletronicamente; apoiar o desenvolvimento de materiais didáticos sobre gêneros, visando atender demandas como os Parâmetros de Currículos Nacionais ou a formação contínua de professores; e auxiliar pesquisadores, professores e estudantes da área de jornalismo na identificação de características e categorias lingüísticas contrastivas ou semelhantes entre os principais gêneros presentes na imprensa escrita do Brasil.

As perguntas de pesquisa que se colocam diante da proposta de estudo estão enumeradas abaixo:

1. Utilizando a metodologia empregada em Biber (1988), que características e categorias do português sobressaem na variação lingüística de um corpus da imprensa escrita diária?
2. Quais são os principais *fatores* responsáveis pela variação da linguagem presente no corpus jornalístico e como eles se comportam em relação aos subcorpora de gêneros?

---

<sup>7</sup> Ou, mais especificamente, Ciências da Comunicação. Utilizaremos “Estudos de Comunicação” daqui em diante, na perspectiva de que o termo possa abarcar não somente estudos teóricos mas também compêndios utilitários como os manuais de redação.

3. Que funções comunicativas (isto é, *dimensões*) expressam os fatores obtidos e como os gêneros se posicionam ao longo das dimensões?

As três perguntas guardam, em primeiro lugar, um vínculo fundamental com o *corpus jornalístico* tomado como conjunto: o objetivo principal é analisar por meio da AMD a variação lingüística encontrada no corpus como um todo. A seguir, verifica-se em que medida a variação captada pela AMD pode ser responsável pela diferenciação entre os gêneros identificados no corpus. Para a identificação dos fatores e sua subsequente interpretação em dimensões, é preciso recorrer aos gêneros e ao seu comportamento de distribuição estatística para inferir quais funções comunicativas estão em jogo na variação lingüística observada.

Para responder a essas perguntas serão examinados inicialmente os pressupostos teóricos que embasam a pesquisa no terreno da Lingüística de Corpus (cf. 2.1), questões teóricas relativas à metodologia empregada, a Análise Multidimensional (cf. 2.2) e, particularmente, a questão dos gêneros (cf. 2.3).

A seguir, no Capítulo 3, abordar-se-á a metodologia empregada na pesquisa, como o desenho do corpus de estudo e procedimentos de etiquetagem morfossintática do corpus (cf. 3.2), mas, sobretudo, será examinada a Análise Multidimensional e seus parâmetros de pesquisa (cf. 3.3 e 3.4). Por fim, será apresentada uma proposta de tipologia dos gêneros do jornal (cf. 3.5.1).

O Capítulo 4 apresenta os resultados da Análise Fatorial e tece considerações a respeito dos fatores obtidos, de modo a propiciarem terreno para o estabelecimento de uma nomenclatura funcional aos fatores, que se tornam assim as dimensões do texto jornalístico em português (cf. 4.7 e 4.8).

## 2 FUNDAMENTAÇÃO TEÓRICA

Certos princípios e procedimentos tomados *partis pris*, que fundamentam o estudo da língua com a metodologia e o rigor requeridos pela Lingüística Aplicada, são analisados neste capítulo. Em primeiro lugar será feita uma reflexão a respeito da abordagem da Lingüística de Corpus (KENNEDY, 1998; BIBER et al. 1998; BERBER SARDINHA, 2004a), que norteia esta pesquisa teoricamente. Será examinada em detalhe a constituição do *corpus*, em termos de sua representatividade e de seu equilíbrio, propriedades necessárias para credenciá-lo como amostra *relativamente* confiável de uma dada variedade da linguagem – em especial, da linguagem existente no jornal. Subsidiariamente, será exposto como a Lingüística de Corpus articula-se com o arcabouço teórico formado pelas idéias introduzidas por Firth (MONAGHAN, 1979; DE BEAUGRANDE, 1991, p. 187-222) e desenvolvidas por Halliday (HALLIDAY, 1994) e Sinclair (SINCLAIR, 1991).

Em 2.2, será feita uma revisão dos fundamentos da Análise Multidimensional (BIBER, 1988), que será a metodologia utilizada nesta pesquisa. Nesse momento também será visto como a pesquisa de Biber (1988) se posiciona em relação à abordagem da Lingüística de Corpus.

A seção seguinte é dedicada à discussão do conceito de gênero, à luz de diversas perspectivas teóricas. A maioria delas encontra-se no âmbito da Lingüística Aplicada: a corrente sistêmico-funcional (HALLIDAY; HASAN, 1989; EGGINS; MARTIN, 1997); a análise de gêneros da escola britânica, cujos expoentes são Swales (1990) e Bhatia (1993); a conceituação de gênero/registro, em Biber (1988; 1995); e a análise de gêneros textuais/discursivos desenvolvida a partir das idéias de Bakhtin (2000). Outra perspectiva teórica examinada é proveniente da área de Estudos de Comunicação, em que se conceitua o gênero e, especificamente, são examinadas as fontes para a obtenção de suas variedades empregadas no jornal. Como conclusão, será declarado o posicionamento deste trabalho em relação à discussão tratada sobre os

gêneros, para explicitar as escolhas feitas pelo pesquisador.

## 2.1 A ABORDAGEM DA LINGÜÍSTICA DE CORPUS

Há na literatura considerável convergência sobre os objetivos e finalidades da Lingüística de Corpus (SINCLAIR, 1991; STUBBS, 1993; McENERY; WILSON, 1996; BIBER et al. 1998; KENNEDY, 1998; HUNSTON, 2002; BERBER SARDINHA, 2004a), que destacam seu papel fundamental como meio privilegiado de investigação da estrutura da linguagem. É adequada, desse ponto de vista, a acepção de que

“A Lingüística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador” (BERBER SARDINHA, 2000a, p. 325).

O termo “abordagem”, empregado para nomear esta seção, é utilizado como forma de definir o status da Lingüística de Corpus, que não deve ser considerada como uma disciplina – já que seu objeto de estudo não está delineado, como ocorre, por exemplo, com a Sociolingüística –, mas tampouco pode ser definida *apenas* como uma metodologia<sup>8</sup> (cf. BERBER SARDINHA, 2000a, p. 355-357, TOGNINI-BONELLI, 2001, p. 1). Existe uma afinidade neste trabalho com a definição de Leech, na qual “A Lingüística de Corpus define não somente uma nova metodologia emergente para o estudo da linguagem mas uma nova empreitada de pesquisa e, na verdade, uma nova abordagem filosófica.” (LEECH, 1992, p. 106, apud BERBER SARDINHA, 2000a, p. 357)

Essa posição, porém, não é unânime – Biber et al. (1998, p. 9-10), por exemplo, considera que a abordagem baseada em corpus pode complementar outras abordagens tradicionais, “mas não [deve ser vista] como a única abordagem correta”;

---

<sup>8</sup> Esta posição é sustentada por McEnery e Wilson (1996, p. 1).

Kennedy (1998, p. 8) igualmente não vê a Lingüística de Corpus como uma teoria de linguagem, mas como fonte de evidência da linguagem em uso, à disposição de pesquisadores de áreas diversas, na forma de dados empíricos.

Um importante componente associado aos modernos estudos baseados em corpus é a tecnologia empregada para a exploração da linguagem. Diferentemente dos primeiros estudos lingüísticos que fizeram recurso de um corpus (Käding, em 1897, coletou manualmente um corpus de 11 milhões de palavras do alemão; Thorndike elaborou em 1921 uma lista das palavras mais freqüentes em língua inglesa, aperfeiçoada por West, em 1953<sup>9</sup>), o estudo baseado em corpus contemporaneamente deixou de estar sujeito à crítica da imprecisão porque adquiriu notável segurança no processamento extensivo e organizado de dados com o advento da informática<sup>10</sup>.

Entre os corpora pioneiros da era do armazenamento digital das décadas de 1960-80, estão o Brown (1964, de 1 milhão de palavras do inglês americano escrito), o LOB (Lancaster-Oslo-Bergen, 1978, de 1 milhão de palavras do inglês britânico escrito) e o LLC (London-Lund Corpus, 1980, de 500 mil palavras do inglês britânico falado)<sup>11</sup>. Biber (1988) utilizou partes dos dois últimos – foram selecionados de ambos alguns subcorpora de gêneros textuais – para seu conhecido estudo sobre variação lingüística entre o inglês falado e escrito, que trataremos em detalhe em 3.3.3.

O computador, capaz de processar quantidade sobre-humana de informação a partir de um dado corpus, expandiu enormemente a capacidade de análise da linguagem por meio da contagem de freqüência de suas palavras e outras características/traços presentes no texto. Desse modo, uma variedade de ferramentas computacionais, como a suíte WordSmith Tools, etiquetadores morfossintáticos (CLAWS/JAWS, VISL, etc.) e testes estatísticos (como palavras-chave<sup>12</sup> e a Análise

---

<sup>9</sup> Cf. outros estudos pioneiros em McEnery e Wilson (1996, p. 2-3) e Kennedy (1998, p. 13-19).

<sup>10</sup> Cf. McEnery e Wilson (1996).

<sup>11</sup> Cf. mais detalhes em Biber et al. (1998, p. 281-287) e Berber Sardinha (2000a, p. 330).

<sup>12</sup> Cf. Scott (1997) e Berber Sardinha (1999).

Fatorial), pode ser combinada de modo a permitir o estudo de diferentes gêneros de textos para a verificação de semelhanças e diferenças entre seus componentes lingüísticos<sup>13</sup>.

Inicialmente é preciso fundamentar a noção de *corpus*, destacando em seguida sua tipologia e características, nos trabalhos que detalham o assunto em exame (SINCLAIR, 1991, p. 13-26, 171; BIBER et al., 1998, p. 246-253; KENNEDY, 1998, p. 60-75; BERBER SARDINHA, 2000a, p. 335-349; HUNSTON, 2002, p. 14-16, 25-37). Kennedy (1998, p. 1) o define como “um corpo de texto escrito ou fala transcrita que pode servir de base para a análise e descrição lingüística”. O autor tem a preocupação de distingui-lo de um *arquivo*, que significaria um “repositório de textos, em geral extenso e coletado de maneira oportunística, e normalmente não-estruturado” (KENNEDY, 1998, p. 4<sup>14</sup>).

Para Hunston (2002, p. 2), corpus pode ser definido “em termos de sua forma e seu propósito”. Ela agrega ao conceito tradicional de corpus – “coleção de exemplos que ocorrem naturalmente na língua”<sup>15</sup> – uma dimensão que envolve a questão tecnológica, propiciada pelas análises com o auxílio de computador: “coleção de textos (ou partes de texto) armazenados e acessados eletronicamente”. Em vista dos fins deste trabalho – o que demonstra a preocupação com as questões ligadas à representatividade do corpus e à sua consistência como fonte de dados empíricos, além de absorver as características modernas do conceito –, adotar-se-á a seguinte definição de corpus:

“Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em

---

<sup>13</sup> Cf. Capítulo 3.

<sup>14</sup> No original, “... an archive is a text repository, often huge and opportunistically collected, and normally not structured.”

<sup>15</sup> Cf. a definição de Sinclair (1991, p. 171, tradução de Berber Sardinha, 2000a, p. 336): “Uma coletânea de textos naturais, escolhidos para caracterizar um estado ou variedade de linguagem.”

amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise” (SANCHEZ, 1995, p. 8-9, apud BERBER SARDINHA, 2000a, p. 338).

Um traço funcional é observado na definição: o corpus deve servir a um propósito específico ao pesquisador, sob pena de seu conteúdo comprometer resultados e, portanto, sua interpretação. Tognini-Bonelli chama a atenção para esse fato, notando que na determinação da função de um corpus subjaz a relação entre a metodologia escolhida e sua amostragem. Um corpus, segundo a autora, “precisa ser justificado em termos lingüísticos” e seus textos devem ser “selecionados segundo um critério explícito de modo a capturar as regularidades de uma língua, variedade lingüística ou sublíngua” (TOGNINI-BONELLI, 2001, p. 55<sup>16</sup>). É preciso, pois, proporcionar ao corpus uma configuração de homogeneidade, sem que seus textos percam sua identidade, para uma pesquisa sobre sua tipologia.

A conformação do corpus é variável de acordo com seu “design, tamanho e natureza” (KENNEDY, 1998, p. 3-4), para descrição geral ou específica, de forma estática ou dinâmica<sup>17</sup>, mas sempre deve ser representativo. O autor reconhece que sempre haverá certa margem de dúvida se “uma amostra de textos pode ser efetivamente representativa em todos os possíveis gêneros ou mesmo de um tipo particular de gênero, assunto ou tópico” (KENNEDY, 1998, p. 62<sup>18</sup>). Tognini-Bonelli sintetiza da seguinte maneira a questão da representatividade do corpus:

---

<sup>16</sup> No original, “a corpus ... needs to be justified in linguistic terms. (...) The texts are selected according to explicit criteria in order to capture regularities of a language, a language variety or a sub-language”.

<sup>17</sup> Cf. também o conceito de corpus monitor, em Sinclair (1991, p. 25). Ele daria conta das transformações por que passa a língua em seu estado atual, sendo constantemente alimentado de dados novos e, assim, capaz de registrar, por exemplo, as novas palavras e as que caem em desuso, novos sentidos dados a uma palavra existente, etc.

<sup>18</sup> No original, “... it is not easy to be confident that a sample of texts can be thoroughly representative of all possible genres or even of a particular genre or subject field or topic”.

“Assim parece haver um acordo geral entre estudiosos que optaram por trabalhar com um corpus que este deveria ser representativo de certa população e que as afirmações derivadas da análise do corpus serão amplamente aplicáveis a uma amostra maior ou à linguagem como um todo” (TOGNINI-BONELLI, 2001, p. 57<sup>19</sup>).

Há que considerar ainda a diferença entre a presença ou a ausência de alguma característica da língua no corpus e o perfil da distribuição de frequência dessa característica. Uma característica linguística pode não ser representativa para o corpus como um todo, mas apresentar uma tipicidade em dada parcela do corpus, como, por exemplo, em um gênero.

A investigação lingüística que utiliza um corpus como objeto de estudo trabalha dentro de um quadro que comporta duas vertentes teóricas básicas: uma, de natureza confirmatória, em que a evidência dos dados pode assumir um papel de *exemplo* para uma teoria anteriormente formulada pelo pesquisador, e outra, de natureza exploratória e empírica, na qual o pesquisador busca no corpus padrões e distinções entre os dados para embasar a formulação de generalizações lingüísticas e, assim, chegar a uma teoria. Para Tognini-Bonelli (2001), a primeira vertente de estudo do corpus é uma abordagem “*corpus-based*” (baseada em corpus), enquanto a segunda é uma abordagem “*corpus-driven*” (conduzida pelo corpus; cf. TOGNINI-BONELLI, 2001; HUNSTON, 2002). Ambas as abordagens não têm um caráter excludente, apesar da existência de estudos radicados nos dois pólos teóricos – no primeiro grupo poderiam ser citados Aarts (1991) e Leech (1991); no segundo grupo, Hunston e Francis (1999), Sinclair (1991), Tognini-Bonelli (2001) e o projeto COBUILD.

A posição de Biber (1988, 1995), que coincide com esta pesquisa, é intermediária, seguindo um percurso metodológico iniciado numa abordagem “*corpus-based*”, na medida em que utiliza conceitos e categorias preexistentes na escolha das variáveis de estudo e funções da linguagem (CHAFE, 1982, QUIRK et al., 1985) para

---

<sup>19</sup> No original, “Thus there seems to be general agreement among scholars who choose to work on a corpus that this should be representative of a certain population and that the statements derived from the analysis of the corpus will be largely applicable to a larger sample or to the language as a whole”.

chegar a dimensões de linguagem e tipos de texto produzidos pelos dados disponíveis no corpus. De maneira similar, este trabalho percorre tal caminho, iniciado na classificação de gênero segundo critérios externos, passando pela seleção das características mais relevantes para uma Análise Multidimensional, e tendo como ponto de chegada resultados derivados da análise estatística processada e uma série de dimensões que expressam as funções comunicativas pelas quais se distribuem os textos coletados do corpus de estudo.

Para complementar os principais conceitos teóricos expostos acima, serão apresentadas as formas de descrição lingüística propostas pela escola neofirthiana e suas premissas principais, que formam uma espécie de moldura teórica à abordagem da Lingüística de Corpus.

### 2.1.1 A Escola Neofirthiana

Inicialmente será explanado o que Monaghan (1979) intitula a “tradição neofirthiana”. O termo faz referência direta a J. R. Firth – e, indiretamente pelo prefixo, a seus discípulos, cujos maiores representantes são M. A. K. Halliday e John Sinclair. Monaghan ressalva que a cunhagem desse termo não significa reconhecer o estabelecimento de uma “‘nova’ teoria de linguagem ou lingüística”, mas de uma “escola” (no sentido “descritivo, não prescritivo”) onde “estudiosos compartilham áreas de interesse comuns, uma mesma terminologia e maneiras comuns de enxergar os problemas da língua” (MONAGHAN, 1979, p. 6<sup>20</sup>).

Num breve plano geral, em relação às correntes principais da Lingüística, os

---

<sup>20</sup> No original, “I am not proposing a ‘new’ theory of language or linguistics but am only trying to make more available a part of modern linguistics which has received less attention than it deserves outside of Britain and the Commonwealth. The concept of linguistic ‘schools’ is a descriptive, not a prescriptive, notion and is meant to indicate groups of scholars who share common ideas of interest, common terminology and certain common ways of looking at the problems of language”.

neofirthianos encontram-se posicionados em terreno praticamente oposto à corrente chomskyana<sup>21</sup>, devido principalmente às fontes utilizadas para atingir o conhecimento da linguagem: enquanto os primeiros se valem do método empirista de construção do conhecimento, através da experiência e do exame da língua em situação real de uso, a segunda estuda a linguagem por meio da introspecção do pesquisador, incumbido de “verificar modelos de funcionamento estrutural e processamento cognitivo da linguagem” (BERBER SARDINHA, 2000a, p. 350). A linha estruturalista derivada de Saussure, por seu turno, tem várias de suas dicotomias basilares (língua X fala, sincronia X diacronia, eixos paradigmático X sintagmático, etc.) contestadas ou “reescritas” (STUBBS, 1993, p. 14-18, 22) pela escola firthiana. O próprio Firth (1935, p. 53; 1957, p. 2), citado por Stubbs (STUBBS, 1993, p. 22), explicitamente autodenomina sua “abordagem” de “monista”, visão distinta, portanto, dos dualismos saussurianos. Já a corrente funcionalista do Círculo de Praga, outra influente escola européia, apresenta alguma afinidade com as idéias de Firth e seus discípulos (principalmente Halliday, que posteriormente irá utilizar conceitos funcionalistas, como o de marca)<sup>22</sup>.

Monaghan (1979) mostra que uma das principais fontes de idéias formuladas por Firth é o trabalho pioneiro do etnólogo inglês Malinowski (1923), que considerou cientificamente a hipótese de a linguagem ser um fato social e um modo de ação<sup>23</sup>, “não um instrumento de reflexão” (p. 184)<sup>24</sup>. O conceito de “contexto de situação”, cunhado originalmente por Malinowski, é aplicado por Firth e assim *aggiornato*:

---

<sup>21</sup> McEnery e Wilson (1996, p. 4-10) enumeram a argumentação racionalista chomskiana contra abordagens baseadas em corpus. Cf. também De Beaugrande (1996).

<sup>22</sup> Cf. Monaghan (1979, p. 9).

<sup>23</sup> Cf. também De Beaugrande (1991, p. 195) e Ducrot e Todorov (1974, p. 88).

<sup>24</sup> De Beaugrande (1991, p. 192) menciona a afinidade de Firth com filósofos empiristas, como Wittgenstein, Locke, Hume e Moore, que têm em comum com a disciplina lingüística o interesse pelo estudo do significado.

“Contexto de situação é um grupo de categorias abstratas relacionadas a traços relevantes da interação” (MONAGHAN, 1979, p. 185<sup>25</sup>). O contexto passa a ser visto como parte indissociável da linguagem: “O foco de atenção é o significado do ato de fala no contexto” (MONAGHAN, 1979, p. 184<sup>26</sup>). De acordo com Malinowski, em um plano mais amplo, haveria também o *contexto de cultura*, que englobaria os “contextos de situação”.

Firth expande as fronteiras do significado na Lingüística, ultrapassando a sua circunscrição tradicional, no terreno da Semântica. Ele postula sua permeabilidade em todos os níveis lingüísticos: “O significado [...] deve ser considerado como um complexo de relações contextuais, e fonética, gramática, lexicografia e semântica, cada qual carrega seu próprio componente do complexo no seu contexto apropriado” (FIRTH, 1957, p. 19, apud MONAGHAN, 1979, p.31<sup>27</sup>).

A dispersão do significado em vários níveis na linguagem, como formula o autor, implica a possibilidade de explorar as relações existentes entre determinadas características ou categorias lingüísticas e as manifestações do significado na língua. Tal visão reforça e inspira este trabalho, na medida em que será proposto um meio de observar um fenômeno sociocultural, no caso a classificação dos textos do jornal em agrupamentos de gêneros, ante a sua manifestação como linguagem.

O contexto de situação revela, com o auxílio da linguagem, o significado. O mecanismo da “renovação da conexão”<sup>28</sup>, proposto por Firth, processaria vários contextos de situação semelhantes e assim qualquer teoria estaria sendo

---

<sup>25</sup> No original, “The context of situation is a group of abstract categories related to relevant features of the interaction”.

<sup>26</sup> No original, “The focus of attention is the meaning of the speech act in context”.

<sup>27</sup> No original, “Meaning ... is to be regarded as a complex of contextual relations, and phonetics, grammar, lexicography and semantics each handles its own component of the complex in its appropriate context”.

<sup>28</sup> Tradução minha para “renewal of connection”.

constantemente testada pela experiência da enunciação<sup>29</sup>. Firth explicita a necessidade de os fatos serem extraídos das “seqüências de fala” “operando em contextos de situação que são típicos, recorrentes e repetidamente observáveis” (FIRTH, 1957, p. 35 e 144, apud DE BEAUGRANDE, 1991, p. 200<sup>30</sup>). Ou seja, é a regularidade (ou freqüência) dos enunciados que irá revelar *padrões* que levam à teoria. E os enunciados só podem ser estudados a partir da coleta de exemplos reais na língua. Stubbs sistematiza esse princípio: “A língua deveria ser estudada em exemplos de uso concretos, atestados e autênticos, não com períodos intuitivos, inventados ou isolados” (STUBBS, 1993, p. 2<sup>31</sup>).

Uma das principais implicações do contexto de situação é a elevação de estatuto do texto como unidade de estudo do significado<sup>32</sup>, em detrimento do período: “O significado completo de uma palavra é sempre contextual, e nenhum estudo de significado que não leve em consideração seu *contexto completo* pode ser levado a sério” [grifo nosso] (FIRTH, 1935, p. 37, apud STUBBS, 1996, p. 53<sup>33</sup>).

Para examinar o significado de um texto integral, Firth – e, posteriormente, Halliday – propõe a divisão do significado em várias funções componentes. Sem deixar de levar em consideração a integralidade da linguagem, caberia ao lingüista aplicar esquemas teóricos e fazer afirmações em termos de estruturas e sistemas, em determinados níveis de análise. O significado total do texto estaria disperso em uma

---

<sup>29</sup> Cf. Monaghan (1979, p. 37 e 185).

<sup>30</sup> No original, “operating in contexts of situation which are typical, recurrent, and repeatedly observable”.

<sup>31</sup> No original, “Language should be studied in actual, attested, authentic instances of use, not as intuitive, invented, isolated sentences”.

<sup>32</sup> Cf. Sinclair (1991, p. 19), Stubbs (1993, p. 2 e 11) e De Beaugrande (1991, p. 201).

<sup>33</sup> No original, “the complete meaning of a word is always contextual, and no study of meaning apart from a complete context can be taken seriously”.

série de níveis (DE BEAUGRANDE, 1991, p. 203)<sup>34</sup>.

Outra idéia instigante introduzida por Firth é a da colocação, atualmente definida como a “associação entre itens lexicais, ou entre léxico e campos semânticos” (BERBER SARDINHA, 2005, p. 40). Para Firth, formas de expressão lingüística delimitadas na experiência e na ação, no original “*restricted languages*”, são um manancial de descobertas de colocações, para “o estudo de palavras-chave, palavras pivotais, palavras principais, mostrando-as em companhia de quem as acompanham” (FIRTH, 1968, p. 106, 113 e 182, apud DE BEAUGRANDE, 1991, p. 212)<sup>35</sup>.

A abordagem sistêmico-funcional da linguagem adotada por Halliday (também chamada de “sociosemiótica”) entende, como Firth, que o texto é essencialmente uma unidade semântica. O texto é definido como “linguagem” que assume uma forma “funcional”, ou, ainda, utilizando a terminologia de Malinowski, “qualquer exemplo de língua viva que toma parte em um contexto de situação” (HALLIDAY; HASAN, 1989, p.10). Na relação entre texto e contexto de situação, Halliday identifica três tipos principais de funções da linguagem (ou *metafunções*) – ideacional (subdividida, por sua vez, em lógica e experiencial), interpessoal e textual – , respectivamente ligadas aos componentes de situação “campo do discurso”<sup>36</sup>, “relações<sup>37</sup> do discurso” e “modo do discurso” (HALLIDAY; HASAN, 1989, p. 24-29)<sup>38</sup>.

Podemos chegar a um conceito central da escola neofirthiana, que diz

---

<sup>34</sup> Entre as funções componentes, estariam a fonética, lexical, morfológica, sintática e situacional.

<sup>35</sup> No original, “The ‘restricted language’ is a prime domain for discovering ‘**collocations**’: for ‘studying key words, pivotal words, leading words, by presenting them in the company they usually keep’”.

<sup>36</sup> Ligado mais especificamente ao componente funcional experiencial.

<sup>37</sup> Em inglês, “tenor”.

<sup>38</sup> Ao relacionar funções de traços lingüísticos particulares à variação de situações comunicativas, Biber (1988, p. 33-36) identifica sete categorias funcionais: ideacional, textual, pessoal, interpessoal, contextual, processual e estética.

respeito à *visão* de linguagem adotada: a linguagem deve ser considerada como um sistema de natureza *probabilística* (BERBER SARDINHA, 2000a, p. 350; STUBBS, 2001, p. 230). “As palavras não ocorrem por acaso no texto”, resume Sinclair (1991, p. 110). As restrições impostas pelo contexto influenciam diretamente o repertório de opções lingüísticas à disposição do falante, em vários níveis.

A conseqüência dessa visão é a necessidade de descrever a língua *em uso* para que seja possível observar o perfil de padronização utilizado por um determinado registro, de modo *empírico*, como conclui Berber Sardinha:

“Para se saber qual a probabilidade de um traço ou estrutura é necessária, portanto, a observação empírica da freqüência do emprego, realizado por diversos usuários, em contextos definidos. Destas considerações, tiram-se duas conclusões. A primeira é a importância primordial de um corpus como fonte de informação, pois ele registra a linguagem natural realmente utilizada por falantes e escritores da língua em situações reais. A segunda é a não-trivialidade da investigação da freqüência de ocorrência de traços lingüísticos de várias ordens (lexicais, sintáticos, semânticos, discursivos, etc.), pois é através do conhecimento da freqüência atestada que se pode estimar a probabilidade teórica” (BERBER SARDINHA, 2000a, p. 352).

Os estudos conduzidos por Sinclair (1972, 1987, 1991), entre outros autores<sup>39</sup>, examinaram exemplos de língua em uso, reunidos em um *corpus*<sup>40</sup>, para extrair noções gramaticais que dizem respeito à natureza de seu funcionamento. As áreas lexicográfica e pedagógica foram grandemente beneficiadas com a aplicação desse conhecimento, no desenvolvimento de dicionários, gramáticas, métodos de ensino e novos materiais didáticos<sup>41</sup>. Sua pesquisa reforçou, no âmbito da teoria, o caráter descompartmentalizado das instâncias lexical e gramatical, tradicionalmente estudadas de forma atomizada. A noção de léxico-gramática ganhou uma conformação mais definida com a pesquisa de Sinclair (STUBBS, 1993, p. 14-15). Ele mostrou a interdependência entre léxico e gramática examinando o fenômeno da co-ocorrência

---

<sup>39</sup> Quirk et al. (1985), Biber et al. (1999) são trabalhos representativos da área.

<sup>40</sup> Cf. 1.3.

<sup>41</sup> Como o *Longman Contemporary English Dictionary*, a série *Collins Cobuild English Guides* e o

lexical e os usos de formas lingüísticas em padrões de significado (SINCLAIR, 1991). O autor também identifica dois princípios que interpretam a maneira pela qual o significado surge no texto: os princípios idiomático e de escolha aberta. Enquanto este, na sucessão de escolhas que o texto coloca palavra a palavra, admite todas as possíveis combinações durante a sua criação (respeitada a gramaticalidade), aquele se pauta pela rigidez de modelos de natureza contextual e circunscreve a escolha aos seus *padrões* lingüísticos correspondentes (SINCLAIR, 1991, p. 109-115). O princípio de escolha aberta estaria manifestado no texto poético, extremamente imprevisível, ou na fala de um aluno intermediário de segunda língua que, apesar de articular corretamente seu repertório lexical, utiliza indistintamente vocabulário e construções sintagmáticas de registros<sup>42</sup> seletivos (formal/informal, por exemplo); o princípio idiomático, por seu turno, predomina em textos especializados, como uma ata societária ou um boletim meteorológico, mas pode ser responsável, em um nível microlingüístico, pelas construções preexistentes de uma locução ou por colocações em um dado texto.

Ante o exposto nesta seção, é importante relacionar as possíveis interfaces entre os conceitos examinados e a pesquisa aqui descrita. Firth, em retrospecto, desenvolveu vários conceitos que valorizam a importância do contexto na linguagem, tido como fundamental para a exploração do significado. Propostas como a utilização do texto autêntico e real como unidade central da análise lingüística, a distribuição do significado em vários níveis e o exame de colocações e palavras-chave no texto foram aproveitadas posteriormente e fortalecidas teoricamente por uma série de autores, entre os quais se destacam Halliday e Sinclair. A exploração da característica funcional da linguagem feita por Halliday levou-o à elaboração de uma teoria que pudesse interpretar o texto por meio de *metafunções*. Sinclair mostrou que não há fronteira clara entre léxico e gramática, além de revelar forças que atuam na produção da linguagem, com a proposição dos princípios idiomático e o de escolha aberta.

Tais idéias têm correspondência direta com este trabalho. Trata-se de uma

---

<sup>42</sup> No sentido dado pela Sociolingüística.

pesquisa de natureza empírica, na medida em que se busca observar no corpus *padrões* de frequência que possam permitir tirar conclusões sobre o funcionamento da língua. A organização do corpus obedece ao critério de autenticidade do texto, pois todos os textos coletados para o corpus foram efetivamente veiculados para leitores reais e sujeitos à interação social. A posição adotada na pesquisa, semelhante à de Sinclair, é a de que o gênero textual influencia as escolhas léxico-gramaticais do autor/falante e que as técnicas de análise propiciadas pela Linguística de Corpus, como a contagem de frequência vocabular, podem revelar certos padrões linguísticos existentes.

A seguir será discutida uma metodologia específica da Linguística de Corpus, a Análise Multidimensional, que será utilizada neste trabalho como ferramenta principal de análise de variação linguística.

## 2.2 A ANÁLISE MULTIDIMENSIONAL

A metodologia da Análise Multidimensional (AMD) será definida teoricamente nesta seção e detalhada tecnicamente no próximo capítulo. A AMD envolve o processo estatístico multivariado – ou seja, capaz de lidar com múltiplas variáveis – da Análise Fatorial, que agrupa e reduz os dados de forma a criar feixes de variáveis que ocorrem (ou ficam ausentes) em grupo com regularidade. Tais feixes de variáveis são chamados de *fatores*. As variáveis podem ser uma característica ou categoria linguística ou ainda ter um caráter situacional, desde que possam ser quantificáveis e ter o perfil adequado para serem utilizadas numa Análise Multidimensional (LEE, 2000, p. 175, 219-225; cf. 3.4.2). Extraídos os fatores, o pesquisador efetua uma seleção dos fatores mais relevantes por meio de uma análise do gráfico de seus *eigenvalues*<sup>43</sup>. Os fatores selecionados podem ser, a partir daí, comparados entre si, aplicados a subcorpora de gêneros, por exemplo, e interpretados

---

<sup>43</sup> *Eigenvalue* é o índice que representa a quantidade de variância compartilhada de cada fator (LEE, 2000, p.37).

de acordo com as suas características funcionais, tornando-se nesse momento *dimensões*. As dimensões, portanto, são determinadas *a posteriori*, como resultado de todo o processo da Análise Multidimensional.

Vê-se pela descrição acima que a Análise Multidimensional é uma metodologia eminentemente quantitativa. Ela depende da contagem de frequências de um grande número de características e categorias lingüísticas e situacionais para estabelecer padrões de ocorrência e assim produzir generalizações acerca da língua e seus gêneros.

Outra característica da AMD é sua dependência dos recursos proporcionados pela informática para o processamento de largas porções de textos e de diversos testes e procedimentos estatísticos. Utiliza-se uma série de programas combinados, em várias plataformas. O processamento manual, além de longo, está sujeito a erro humano e está praticamente descartado como alternativa às soluções automatizadas ou semi-automatizadas à disposição do pesquisador. No entanto, é necessário frisar que um controle manual para a verificação da qualidade dos resultados dados pelas ferramentas computacionais é imperioso, principalmente durante a etapa de contagem de frequências de características e categorias lingüísticas por meio de algoritmos.

Entre os trabalhos anteriores que buscaram identificar diferenças entre fala e escrita, está Chafe (1982), um estudo “exemplar”, segundo Biber (1988, p. 21), porque mostrou a necessidade de interpretar a diferença entre gêneros escritos e falados de acordo com a co-ocorrência de características e categorias lingüísticas e sugeriu dimensões para explicar essa variação (apesar de a metodologia utilizada por Chafe ter seguido caminho inverso, isto é, as dimensões terem sido enunciadas *a priori*, como base para detecção de características e categorias lingüísticas). Chafe (1982) identificou duas dimensões ao estudar dois gêneros típicos do modo falado (conversação) e escrito (artigo acadêmico): integração/fragmentação e afastamento/envolvimento. A integração estaria relacionada a características e categorias lingüísticas como nominalizações, adjetivos atributivos e locuções preposicionais, enquanto a fragmentação, à ausência de conectivos e à predominância

de conjunções coordenativas; a dimensão de afastamento estaria ligada à utilização de passivas e nominalizações; já o envolvimento seria marcado por pronomes de primeira pessoa, partículas enfatizadoras e salvaguardas (*hedges*).

Porém, Biber critica Chafe (1982) e outros estudos anteriores que buscaram contrastar os modos falado e escrito (tais como OCHS, 1979; AKINNASO, 1982; e CHAFE; DANIELEWICZ, 1987) por questões de ordem metodológica: tais trabalhos estudaram corpora pequenos demais, pouco representativos para uma pesquisa dessa extensão; selecionaram poucos gêneros textuais para realizar a análise (geralmente dois, um do modo falado e um do modo escrito); verificaram a existência de um número insuficiente de características/categorias lingüísticas – e os consideraram responsáveis únicos pelas diferenças de linguagem (BIBER, 1988, p. 52-53). A partir da constatação desses problemas, o estudo de Biber (1988) preocupou-se em escolher um corpus amplo e balanceado, que somou 960 mil palavras, distribuído em 481 textos de 23 gêneros falados e escritos dos corpora gerais LOB e London-Lund, além de selecionar o número expressivo de 67 características e categorias lingüísticas para realizar uma contagem de sua distribuição no corpus e em seus subcorpora de gêneros. Desse modo, o autor buscou representar com mais acuidade o universo da fala e da escrita na língua inglesa e analisou esse corpus de acordo com um número significativo de variáveis, para atingir um grau inédito de alcance no tipo de generalização advinda da pesquisa.

Outra crítica indireta é feita quando Biber (1988) explicita outra propriedade da dimensão: diferentemente dos trabalhos precursores, que viam as dimensões de uma forma polarizada, à maneira das dicotomias, Biber enxerga as dimensões como “parâmetros de variação quantificáveis em um contínuo” (BIBER, 1988, p. 22<sup>44</sup>). Ou seja, a dimensão representa um eixo que possui dois pólos opostos, mas que admite gradações nesse eixo e sobre o qual podem ser distribuídos os textos do corpus,

---

<sup>44</sup> No original, “... dimensions are identified as *continuous* quantifiable parameters of variation”

individualmente ou em grupo.

Para a seleção de variáveis que irão compor a Análise Multidimensional, o pesquisador deve avaliar a relevância e representatividade de sua distribuição nos textos do corpus (ou subcorpora), como já examinado acima. Essas variáveis, compostas de elementos de natureza lingüística e situacional, têm em comum o fato de serem passíveis de quantificação. Diversas, porém, são suas naturezas e formas de serem extraídas. O foco adotado neste trabalho é propositalmente amplo e considerará as seguintes possibilidades durante o processo de seleção:

a) características e categorias lingüísticas, isto é, qualquer elemento quantificável de natureza lingüística encontrado na língua. As características podem ser compostas por itens lexicais simples (por exemplo, “não”, para designar negação) e compostos, como os *lemas*, que são conjuntos de formas que compõem outro item lexical (quando, por exemplo, a freqüência das formas verbais flexionadas “sou”, “és”, “é”, etc. é atribuída à freqüência total do verbo *ser*). Já as categorias são classes morfossintáticas, como preposição, substantivo, pronomes demonstrativos, tempos verbais, etc. São grupos que podem também ser fundidos em uma só variável (como verbos e pronomes da primeira pessoa do singular, para detectar a presença dessa pessoa verbal no texto);

b) características não-lingüísticas, como a classificação de um texto pelo seu gênero, julgamento que pertence a uma esfera sociocultural;

O item a) é parte da terminologia empregada por Biber (1988) para definir as características e categorias lingüísticas que se quantificam, podendo ser considerada uma característica, por exemplo, tanto uma classe gramatical ou um tempo verbal, como um índice que indique a razão entre o número total de formas lexicais (“*types*”) e as suas ocorrências (“*tokens*”), ou mesmo a proporção de caracteres por palavra (“*word length*”). A contagem de características lingüísticas implica neste estudo a etiquetagem morfossintática do corpus, para que seja possível contabilizar algumas de

suas freqüências. O conceito de *lema* pode ser definido como um “conjunto composto de formas lexicais” (SINCLAIR, 1991, p. 173), isto é, um repositório em que se agrupam várias formas (“*types*”), cuidadosamente monitorado pelo pesquisador, de modo a evitar ambigüidades e generalizações indevidas. Sobre esse tema, Sinclair (1991, p. 8) adverte que “cada forma distinta é potencialmente uma unidade lexical única, e formas deveriam ser agrupadas em lemas somente quando suas características apresentassem certa similaridade”<sup>45</sup>.

O item b) refere-se à atribuição situacional, segundo um critério externo à língua, de características como gênero, tópico do assunto, participantes, etc. Esta pesquisa lida com esse tipo de informação devido ao processo de atribuição de um gênero a cada um dos textos do corpus de estudo, fato que torna o corpus passível de ser agrupado em um conjunto de subcorpora distintos.

Selecionar as características e categorias lingüísticas no corpus de estudo é uma das tarefas críticas para a consecução da Análise Multidimensional. Os critérios utilizados para a inclusão das variáveis que irão compor a Análise Multidimensional devem ser explicitados porque essa decisão afeta o resultado da análise (LEE, 2000; cf. detalhes dos critérios utilizados aqui em 3.2 e 4.1).

Nas seções seguintes, serão discutidos os conceitos de dimensão e de multidimensionalidade, dois pressupostos teóricos da AMD.

### 2.2.1 Análise Multidimensional: Conceito de dimensão

O lingüista norte-americano Douglas Biber introduziu a metodologia

---

<sup>45</sup> No original, “each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity”.

conhecida como Análise Multidimensional<sup>46</sup> (AMD) na Lingüística de Corpus ao publicar *Variation Across Speech and Writing*, em 1988<sup>47</sup>. O ambicioso estudo pretendeu descrever as diferenças entre os modos falado e escrito da língua inglesa em termos de dimensões – seis ao todo<sup>48</sup> –, que seriam cada qual produto de um processo estatístico de agrupamento de características e categorias lingüísticas co-ocorrentes. Cada dimensão, por sua vez, descreve funções atuantes na linguagem, como justifica o autor:

“Traços não co-ocorrem aleatoriamente no texto. Se certos traços co-ocorrem consistentemente, logo é razoável procurar uma influência funcional subjacente que estimule o seu uso. Desse modo, as funções não são postuladas numa base a priori, mas são requeridas para explicar padrões de co-ocorrências entre os traços lingüísticos” (BIBER, 1988, p. 13<sup>49</sup>).

O conceito de dimensão para Biber está fortemente apoiado à sua correspondência com uma função de caráter genérico da linguagem, que o embasa. As dimensões representam, assim, funções comunicativas que influenciam, em maior ou em menor grau, todos os textos e gêneros do corpus.

A citação acima também especifica que qualquer proposição que identifique uma dimensão só pode ser válida se seus traços constitutivos “co-ocorrerem

---

<sup>46</sup> Biber (1988) denomina a técnica de “Multi-feature Multi-dimensional Analysis of Register Variation” (MF / MD). Este trabalho utilizará simplificadaamente o termo “Análise Multidimensional”, e a abreviatura “AMD”, para descrevê-la.

<sup>47</sup> Antes, porém, Biber já havia publicado artigos sobre o tema (BIBER, 1985; 1986), um desenvolvimento de sua tese de doutorado (não publicada).

<sup>48</sup> Originalmente, Biber (1988) detectou sete dimensões, mas a menos relevante delas foi descartada durante a pesquisa. Biber et al. (1998) reconhece no mesmo trabalho de 1988 cinco dimensões principais (cf. 2.3).

<sup>49</sup> No original, “Features do not randomly co-occur in texts. If certain features consistently co-occur, then it is reasonable to look for an underlying functional influence that encourages their use. In this way, the functions are not posited on an a priori basis; rather they are required to account for the observed co-occurrence patterns among linguistic features”.

*consistentemente*”, o que significa agregar à noção da dimensão um vínculo direto com a representatividade e precisão, só proporcionada pela utilização de um extenso corpus de estudo de língua autêntica, fundamental para a consecução da pesquisa.

### 2.2.2 Multidimensionalidade

Um pressuposto teórico importante em Biber (1988) é a questão da complexidade que envolve a descrição da linguagem, assim colocada pelo autor:

“... a variação lingüística em qualquer língua é muito complexa para ser analisada em termos de uma dimensão única. (...) a descrição da variação lingüística de uma dada língua deverá ser multidimensional” (BIBER, 1988, p. 22<sup>50</sup>).

Biber refere-se a uma descrição de uma língua como um todo (no caso, a inglesa) e, portanto, não apenas a um gênero ou a um modo específico. Provavelmente o autor esteja certo em afirmar que deve haver mais de uma dimensão para dar conta da descrição da língua, e tomaremos também como nosso esse pressuposto. Este estudo, porém, realiza um corte diferente na seleção do corpus, preocupando-se apenas em registrar com rigor textos da imprensa diária escrita em português, de vários gêneros. É possível supor, destarte, que o procedimento da Análise Multidimensional produza aqui um número menor de dimensões significativas, em relação ao alcançado pelo estudo de Biber (1988).

Será feita a seguir uma breve reflexão sobre alguns dos vários pontos de contato existentes entre a AMD e as idéias de Firth, que nortearam o desenvolvimento da Lingüística de Corpus. A metodologia encerrada na Análise Multidimensional converge em larga medida com o arcabouço teórico proposto pela corrente neofirthiana, exposto nas seções iniciais deste capítulo. A AMD utiliza um corpus de língua autêntica, como defendido por Firth, Halliday e Sinclair, sem recorrer a modelos apriorísticos oriundos da intuição do pesquisador. Está, portanto,

---

<sup>50</sup> No original, “linguistic variation in any language is too complex to be analyzed in terms of any single dimension. (...) linguistic variation in a given language will be multi-dimensional”.

fundamentada em uma sólida base empírica: os dados da AMD são provenientes da frequência de características e categorias lingüísticas de um corpus representativo da língua, gênero ou situação que se quer estudar, permitindo sua reprodutibilidade em ocasiões futuras por outros lingüistas.

A visão de linguagem subjacente no trabalho de Biber imbrica-se com a da escola neofirthiana, que a toma como tendo um caráter probabilístico. A técnica estatística envolvida apóia-se nesse fundamento e é o que sustenta questões como a da representatividade do corpus.

A Análise Multidimensional realiza uma pesquisa ao mesmo tempo macrolingüística (entre gêneros textuais, por exemplo) e microlingüística (como a descrição de características e categorias lingüísticas em textos individuais), aproximando-a das proposições de Sinclair (1991) que apóiam de um modo não-excludente tanto aspectos quantitativos e de representatividade<sup>51</sup>, como também a instância de comportamento individual de *cada* palavra em relação ao seu entorno contextual, ao discutir um conceito como o da colocação (SINCLAIR, 1991, p. 115-121).

A perspectiva aberta com o trabalho de Biber (1988), mostrando ser possível reunir características e categorias lingüísticas e situacionais comuns por meio de uma análise estatística de suas frequências para tirar conclusões a respeito de uma língua, gênero ou variedade lingüística, assemelha-se à proposição de Firth de que o significado encontra-se expresso e distribuído em vários níveis lingüísticos. Para Biber, os traços a serem pesquisados não obedecem a um critério rígido de escolha, não se limitando às características e categorias lingüísticas, sendo possível admitir vários níveis de análise – de ordem estrutural, discursiva, etc. Isso indica que novos estudos<sup>52</sup> que promovam outras linhas de investigação na seleção de traços a analisar

---

<sup>51</sup> Cf., por exemplo, a afirmação “... o corpus deve ser tão grande quanto possível” (SINCLAIR, 1991, p. 18). Ou, no original, “... a corpus should be as large as possible”.

<sup>52</sup> Exemplos pioneiros em português são Pacheco de Oliveira (1997), Shimazumi (1998), Conde (2002) e Shergue (2003).

em um corpus terão a chance de refinar os resultados encontrados por Biber (1988) e, provavelmente, lhes dar novas feições<sup>53</sup>.

Nesta seção procurou-se evidenciar que a abordagem metodológica da Análise Multidimensional possui um instrumental analítico que permite nuançar pólos conceituais antes dicotômicos. A AMD cria uma perspectiva capaz de descrever padrões textuais com maior grau de sutileza, dispondo subgrupos semelhantes de textos, como os gêneros, ao longo de dimensões que atuam na língua como funções comunicativas mais gerais. Essa gradação, em suma, é enriquecedora e visa proporcionar uma nova base para os estudos de gênero, assunto que será discutido na próxima seção.

### 2.3 VISÕES SOBRE O CONCEITO DE GÊNERO

A abordagem da Lingüística de Corpus e, em específico, da Análise Multidimensional não tem uma teoria de gêneros consolidada (EAGLES, 1996a; 1996b; LEE, 2001). Como visto acima, a metodologia da AMD utiliza largamente a classificação baseada em gêneros para estudar a variação na linguagem (cf. 3.3.3) e suas dimensões. No entanto, Biber (1988) não discute especificamente a questão relativa à classificação dos textos em gêneros, uma vez que preferiu empregar uma terminologia de classificação de gêneros previamente disponibilizada pelos corpora que utilizou.

A Lingüística de Corpus vale-se, portanto, da adoção de perspectivas teóricas a respeito do gênero existentes na Lingüística Aplicada e em outras áreas do conhecimento para desenvolver suas pesquisas.

Serão examinadas brevemente a seguir diversas perspectivas sobre o conceito de gênero. Inicialmente são examinadas as origens do gênero e como o conceito foi difundido pelo senso comum. Em seguida, o gênero é observado por cinco

---

<sup>53</sup> Cf. Lee (2000).

perspectivas diversas: a abordagem sistêmico-funcional, derivada das idéias de Halliday; a linha de análise de gêneros britânica, de Swales e Bhatia; a teoria de Bakhtin sobre os gêneros do discurso; o conceito de gênero e registro, proposto por Biber; e uma exposição sobre como os gêneros jornalísticos são tratados na área de Estudos de Comunicação. Uma discussão sobre essas perspectivas teóricas completa esta seção, de modo a explicitar a linha teórica a respeito de gênero utilizada neste trabalho.

### 2.3.1 Gênero: origens

Em uma perspectiva histórica, foram depositados sobre o termo *gênero* vários sentidos e interpretações. Sua origem remonta à obra de Aristóteles, que o via como um fenômeno da Retórica. Diferentes gêneros, para o filósofo grego, designavam tipos retóricos diversos – deliberativo, judiciário e epidítico (ou demonstrativo), por exemplo –, de emprego específico. Todos eles seriam variações de um tipo de comunicação oralizada, destinada a uma audiência própria e historicamente situada: a peça de oratória chamada de “discurso” (cf. ARISTÓTELES, 1991).

Uma leitura influenciada pela tradição literária, e adotada pelo senso comum, considera o gênero um abrigo para os diversos tipos de produção literária. Os agentes que levam a efeito essa categorização compõem-se de critérios de várias naturezas em coexistência, não raro conflitantes, razão pela qual seu conceito é em certa medida considerado vago e impreciso. Tais critérios poderiam ser relacionados à composição, forma e conteúdo (o soneto, por exemplo); àqueles que influenciam os modos de representar o real (o “gênero romântico”); e aos que são baseados na organização enunciativa (o “gênero policial”) (CHARAUDEAU; MAINGUENEAU, 2004, p. 249; cf. TODOROV, 1980).

### 2.3.2 Visão sistêmico-funcional

Os autores de linha sistêmico-funcional – do qual fazem parte Hasan, Eggins

e Martin – descrevem a natureza do gênero a partir das funções sociais da linguagem que o texto assume em sua materialidade. O papel crucial do contexto na teoria de linguagem proposta por Firth (1957), como visto em 2.1.1, criou condições para que o conceito de texto obtivesse um estatuto de unidade semântica. Halliday e Hasan (1989) avançam nessa abordagem e explicitam como são conferidos ao texto funções de cunho sociocultural, enquanto Eggins e Martin (1997) aprofundam as diferenças conceituais entre gênero e registro e investigam a capacidade de ambos captarem a variação funcional do texto.

Para uma parcela de autores que se basearam nas idéias de Halliday (cf. 2.1) e estudaram a questão do gênero (HALLIDAY; HASAN, 1989; EGGINS; MARTIN, 1997), há uma instância que mediará o gênero no âmbito da linguagem: o *registro*, que seria “uma configuração de significados que são tipicamente associados com uma configuração particular de campo, modo e relações” (HALLIDAY; HASAN, 1989, p. 38-39). O registro se manifesta por meio de traços lingüísticos que se realizam no texto – traços fonológicos, locucionais, expressões, palavras particulares e sinais gramaticais particulares. Os parâmetros contextuais ou situacionais de variação determinam o registro, aos quais o texto está funcionalmente associado, e desse modo imprimem características lingüísticas particulares que definem uma variedade da língua. A noção sistêmico-funcional de registro assemelha-se, portanto, à de “*restrictive language*”, elaborada por Firth (cf. 2.1.1), ressalvando-se que subsiste uma clara diferença no grau de especialização e de quantidade nos traços lingüísticos à disposição de um e de outro conceito.

Para Halliday e Hasan,

“Um gênero é reconhecido pelos significados a ele associados; em realidade, o termo ‘gênero’ é uma forma abreviada de uma expressão mais elaborada, ‘potencial semântico específico de gênero’. ... Gêneros variam em delicadeza do mesmo modo que contextos o fazem. Mas, para que alguns textos pertençam a um gênero específico, é necessário que a sua estrutura seja uma realização de uma determinada EGP [estrutura genérica potencial]. ... Daí decorre o fato de que os textos que pertencem a um mesmo gênero possam variar em relação a suas estruturas; o único aspecto em relação ao qual eles não podem variar sem

conseqüência à sua alocação genérica são os elementos obrigatórios e a disposição da EGP” (HALLIDAY; HASAN, 1989, p. 108<sup>54</sup>; tradução de Tony Berber Sardinha).

Além das considerações sobre a ligação do texto (ou de grupos de textos) de um determinado gênero a um contexto social<sup>55</sup> – e aí haveria a ressonância do conceito malinowskiano de “contexto de cultura”<sup>56</sup> –, cabe destacar a respeito da citação acima o caráter de *predição* contido nos conceitos de “estrutura genérica potencial” e “potencial semântico de especificidade genérica”, sobretudo pela escolha do termo “potencial”. A “estrutura genérica potencial” seria o conjunto de traços funcionais em comum entre os textos, reconhecíveis por um dado conjunto de indivíduos em um contexto determinado. Reside nessa característica do reconhecimento de gêneros uma leitura do texto não apenas como produto mas como processo<sup>57</sup> que reflete a importância das escolhas<sup>58</sup> sucessivamente efetuadas pelo falante.

Em um desdobramento da abordagem sistêmico-funcional sobre gênero e registro, Eggins e Martin (1997) propõem uma teoria de gênero e registro capaz de captar a “variação funcional” (EGGINS; MARTIN, 1997, p. 236) do texto, que permitisse tanto uma “predição textual” como uma “dedução contextual”:

“Ou seja, dada uma descrição do contexto, seria possível prever os sentidos que estão em jogo e as características lingüísticas possivelmente utilizadas para codificá-los. Alternativamente, dado um texto, seria possível deduzir o contexto no qual foi produzido,

---

<sup>54</sup> No original, “A genre is known by the meanings associated with it; in fact the term ‘genre’ is a short form for the more elaborate phrase ‘genre-specific semantic potential’. ... Genres can vary in delicacy in the same way as contexts can. But for some given texts to belong to one specific genre, their structure should be some possible realisation of a given GSP [Generic Structure Potential]. ... It follows that texts belonging to the same genre can vary in their structure; the one respect in which they cannot vary without consequence to their genre-allocation is the obligatory elements and dispositions of the GSP”.

<sup>55</sup> A citação também mostra que um gênero pode admitir mais de um registro.

<sup>56</sup> Cf. 2.1.1.

<sup>57</sup> Cf. Halliday e Hasan (1989, p. 10).

<sup>58</sup> Cf. Stubbs, 1993, p. 14.

pois as características lingüísticas selecionadas em um texto irão codificar dimensões contextuais, tanto seu contexto imediato de produção como sua identidade genérica, qual papel é desempenhado pelo texto na cultura” (EGGINS; MARTIN, 1997, p. 236-237<sup>59</sup>).

Para Eggins e Martin (1997), a capacidade de transmissão de características do contexto para o texto opera em dois planos: o do gênero, em que há uma aproximação com o contexto da cultura e das convenções sociais, e o do registro, que, “inserido no contexto de situação, associa-se às opções léxico-gramaticais” (VIAN JR., 1997, p. 32).

### 2.3.3 Análise de gêneros: Swales e Bhatia

Para Swales e Bhatia, o gênero liga-se à organização da cultura e dos propósitos sociais em torno da linguagem (SWALES, 1990; BHATIA, 1993). Ambos os autores examinam um corpus e tentam identificar na linguagem utilizada a estrutura de gêneros subjacente, mas utilizando na metodologia passos socio-retóricos (cf., por exemplo, SWALES, 1990, p. 141). Há também um comprometimento dessa linha de pesquisa com as implicações pedagógicas que a identificação da estrutura genérica pode proporcionar.

Bhatia (1993) estudou a estrutura intratextual do gênero, na pesquisa com cartas de promoção de vendas (BHATIA, 1993). Já o conceito de *comunidade discursiva* foi desenvolvido por Swales (1990) quando analisou um corpus de artigos acadêmicos de pesquisa em inglês. Para o autor, a comunidade discursiva detém a “propriedade” do gênero. Ou seja, “gêneros pertencem a uma comunidade discursiva, não a indivíduos, a outros tipos de grupo ou a comunidades lingüísticas mais amplas”

---

<sup>59</sup> No original, “That is, given a description of the context, it should be possible to predict the meanings that will be at risk and the linguistic features likely to be used to encode them. Alternatively, given a text, it should be possible to deduce the context in which it was produced, as the linguistic features selected in a text will encode contextual dimensions, both of its immediate context of production and of its generic identity, what task the text is achieving in the culture”.

(Swales, 1990, p. 9<sup>60</sup>).

A relação entre gênero e comunidade se dá na medida em que o gênero sintetiza uma classe de propósitos comunicativos e a compartilha entre seus membros. A composição da comunidade discursiva envolve, em um paralelo com o exemplo do discurso jornalístico dos cotidianos, uma complexidade de papéis atribuídos à sociedade e a seus indivíduos no processo de comunicação. Uma ínfima parte dessa comunidade é produtora de textos do jornalismo diário impresso, ante a quantidade de leitores. O produtor do texto não é o responsável pela sua veiculação; o momento de publicação do texto depende de outros operadores, ligados a um processo industrial de distribuição de material editorial, de lógica econômica de atendimento a um mercado de consumo (cf. FOWLER, 1991, p. 19-24). O escritor não se encontra presente quando seu texto é visto pelos leitores, assincronicamente. A interação entre membros dessa comunidade discursiva é, em regra, indireta e rara – quando se realiza, ademais, possivelmente o é por meio de outros gêneros, como a carta. Esse exemplo sugere que uma comunidade discursiva inscreve-se num contínuo em cujos pólos estão indivíduos com uma maior ou uma menor intensidade de contato com o gênero, em gradações correspondentes ao seu conhecimento.

#### 2.3.4 Gêneros discursivos: Bakhtin

Para Eggins e Martin (1997, p. 236), do ponto de vista da Lingüística, deve-se ao filósofo russo Bakhtin a ampliação do conceito de gênero, para além da classificação herdada pela tradição literária. O gênero, para Bakhtin, inscreve-se como tipologia dos *enunciados*, que, por sua vez, formam a expressão e a unidade do *discurso*. De acordo com a formulação proposta, o gênero passa a abarcar situações vividas no dia-a-dia, nos modos oral e escrito, em circunstâncias cuja enunciação se

---

<sup>60</sup> No original, “... genres belong to discourse communities, not to individuals, other kinds of grouping or to wider speech communities”.

manifesta com certa homogeneidade. Os chamados “gêneros do discurso” são definidos pelo autor como “tipos *relativamente* estáveis de enunciados” [grifo nosso] (BAKHTIN, 2000, p. 279).

Segundo a visão bakhtiniana, o enunciado é definido como a “unidade real da comunicação verbal” (BAKHTIN, 2000, p. 287 e 293). O enunciado coloca-se, portanto, em contraposição a um estudo *descontextualizado* da língua. Bakhtin aponta nessa direção quando critica a utilização da oração e do período (BAKHTIN, 2000, p. 295-7) como recurso de análise e estratégia de ensino (BAKHTIN, 2000, p. 301). Outro elemento da teoria formulada por Bakhtin diz respeito ao caráter dialógico da linguagem (BRAIT, 1997): sendo os enunciados a materialização do discurso, é justificado inferir que às formas assumidas pelos enunciados corresponde um substrato original, formado por enunciados anteriormente efetuados. “A língua materna (...), não a aprendemos nos dicionários e nas gramáticas, nós a adquirimos mediante enunciados concretos que ouvimos e reproduzimos durante a comunicação verbal” (BAKHTIN, 2000, p. 301). Daí os conceitos fundamentais de *dialogismo* e *polifonia*:

“É uma experiência que se pode, em certa medida, definir como um processo de *assimilação*, mais ou menos criativo, das *palavras do outro* (e não das *palavras da língua*). ... nossos enunciados (que incluem obras literárias) estão repletos das palavras dos *outros*” [grifos do autor] (BAKHTIN, 2000, p. 314).

Quando os enunciados se tornam comparáveis, estabelece-se uma relação de imanência entre ambos, metaforicamente chamada por Bakhtin de dialogismo:

“O enunciado está voltado não só para o seu objeto mas também para o discurso do outro acerca desse objeto. A mais leve alusão ao enunciado do outro confere à fala um aspecto dialógico que nenhum tema constituído puramente pelo objeto poderia conferir-lhe” (BAKHTIN, 2000, p. 320).

A medida comum de comparação entre enunciados sugere a formação do gênero e de uma tipologia correspondente. Bakhtin propôs a divisão dos gêneros em dois ramos principais: gêneros primários e secundários. Enquanto a primeira categoria representa formas primeiras e espontâneas da expressão lingüística, como o diálogo e a carta, a segunda, por sua vez, designa formas elaboradas, mais complexas, de

circulação numa esfera pública de comunicação, entre as quais poderiam ser citados o romance e o artigo científico.

Completando o quadro teórico proposto por Bakhtin, deve-se mencionar que são, para o autor, as *condições de produção* do discurso que modelam a existência dos gêneros, de acordo com as funções que desejam expressar: “Uma dada função (científica, técnica, ideológica, oficial, cotidiana) e dadas condições, específicas para cada uma das esferas da comunicação verbal, geram um dado gênero, ou seja, um dado tipo de enunciado, relativamente estável do ponto de vista temático, composicional e estilístico” (BAKHTIN, 2000, p. 284).

### 2.3.5 Gênero/Registro: Biber

Como foi dito, nem a Lingüística de Corpus nem a AMD têm uma teoria consolidada sobre gênero. Apesar disso, muitos trabalhos da área, como Biber (1988; 1995), de maneira pioneira, investigaram a língua utilizando como suporte teórico o conceito de *gênero* e, posteriormente, *registro*. Biber utiliza indistintamente os termos gênero e registro (BIBER, 1988; 1995), preferindo a segunda forma em seus trabalhos mais recentes (BIBER et al., 1998; 1999), como se preponderasse o viés sociolingüístico que a palavra *registro* carrega em, por exemplo, Labov (1972).

Biber também considera o registro como uma categoria mais ampla e abstrata, que congregaria vários sub-registros<sup>61</sup>. Em Biber et al. (1999), quando analisa a gramática da língua inglesa, o autor opõe o *registro jornalístico* às formas da *conversa*ção, da *ficção* e do *registro acadêmico* (BIBER et al., 1999). Para o autor, o registro jornalístico (“news register”), nesta acepção, é descrito como detentor de *características funcionais* comuns e próprias, se comparado aos outros registros. O

---

<sup>61</sup> Como os mencionados em Biber (1988): cartas ao editor, editoriais, reportagem, etc.; esses, por sua vez, podem ser subdivididos por tópico (esporte, política, economia, etc.).

registro jornalístico (BIBER et al., 1999, p. 16) é assim definido:

- 1) Pertencente ao modo escrito;
- 2) Sem interatividade e produção imediata (“*no interactiveness and online production*”);
- 3) Em situação imediata não-compartilhada (“*no shared immediate situation*”);
- 4) Tem como principal propósito comunicativo/conteúdo a informação/julgamento (“*information/evaluation*”);
- 5) Atinge um grande público;
- 6) Possui domínio dialetal de alcance regional e nacional.

Uma importante distinção teórica proposta por Biber é a diferença entre gênero/registo e *tipo de texto* (“*text type*”; cf. BIBER, 1988, p. 70 e 170): enquanto os gêneros/registros são “categorias de texto situacionalmente definidas”, os tipos de texto são “categorias lingüisticamente definidas” (BIBER, 1994, p. 380, apud TOGNINI-BONELLI, 2001, p. 60). No contexto da AMD, os tipos de texto corresponderiam aos textos cuja distribuição se concentra ao longo das dimensões (em qualquer ponto delas) e que formam grupos lingüisticamente semelhantes. EAGLES (1996a; 199b), da mesma forma, sugere a existência de elementos inerentes à linguagem capazes de igualmente realizar uma classificação de gêneros.

Biber (1993) combina ambas as abordagens ao propor uma estratégia de classificação automática de gêneros, por meio de análise discriminante, uma metodologia estatística que lida com predição probabilística. No estudo, cada gênero ocupa uma posição determinada em relação a cada uma das cinco dimensões consideradas, das sete originais de Biber (1988). Cada gênero adquire, portanto, um perfil próprio em relação às dimensões, que o distingue dos perfis que outros gêneros assumem. Dessa forma é possível classificar textos individuais de acordo com a semelhança do seu perfil em relação aos perfis de gênero (BIBER, 1993).

Ainda na área de Lingüística de Corpus, o lingüista de corpus Michael Hoey considera o jornal uma "colônia discursiva". Ou, no original, "discourse colony" (HOEY, 2001, p. 76), congregando diferentes textos, mas de natureza semelhante em um nível geral. O conceito de colônia discursiva ocorre também em subníveis – em um jornal, por exemplo, os textos às vezes estão publicados sob uma hierarquia produzida pelas páginas de determinados tópicos ou seções. Em um nível ainda mais básico, um texto-colônia manifesta-se no texto composto pela soma de vários pequenos textos, como ocorre em uma coluna de notas.

### 2.3.6 A perspectiva dos Estudos de Comunicação

Subsidiariamente à explanação dos fundamentos teóricos sobre gênero em algumas linhas de pesquisa da Lingüística Aplicada, exposta acima, será abordado nesta seção como a questão é tratada na área de Comunicação Social – mais especificamente, nos Estudos de Comunicação. Convém analisar como se definem as expressões de gêneros textuais no universo da imprensa escrita, num contexto sociocultural determinado, com o intuito de subsidiar uma classificação empírica no corpus de estudo.

Marques de Melo (1994) é um trabalho fundamental na área de Estudos de Comunicação, porque faz um extenso retrospecto da literatura, descrevendo classificações de gêneros feitas por autores europeus, norte-americanos e hispano-americanos, além de brasileiros. O autor propõe ainda uma terminologia própria e razoavelmente sistematizada (v. abaixo). A percepção de Marques de Melo é a de que os *gêneros jornalísticos* sofrem influência direta da cultura em que se inscrevem, sendo necessário particularizar suas ocorrências e definições de acordo com variáveis de tempo e espaço – ou seja, deve-se circunscrever os gêneros jornalísticos à cultura de um determinado país ou região, em um determinado período histórico, para que a sua tipologia seja consistente:

“O jornalismo é concebido como um processo social que se articula a partir da relação

(periódica/oportuna) entre organizações formais (editoras/emissoras) e coletividades (públicos receptores), através de canais de difusão (jornal/revista/rádio/televisão/cinema) que asseguram a transmissão de informações (atuais) em função de interesses e expectativas (universos culturais ou ideológicos). ... Assim cada processo jornalístico tem suas próprias peculiaridades, variando de acordo com a estrutura sociocultural em que se localiza, com a disponibilidade de canais de difusão coletiva e com a natureza do ambiente político e econômico que rege a vida da coletividade” (MARQUES DE MELO, 1994, p. 15).

Outro autor de destaque na área é Chaparro (1997; 1998). Ele parte da noção de *superestruturas e macroestruturas*, de Van Dijk (1990, p. 54-59, 80-82) – termos que designam, respectivamente, a ordem externa e a ordem interna do texto. Van Dijk, segundo Chaparro (1997, p. 60), considera que o discurso jornalístico funciona em dois esquemas de superestrutura: o esquema argumentativo e o esquema narrativo. A partir desses dois perfis basilares, Chaparro (1997) defende a existência de dois gêneros básicos, o *comentário*, incumbido de expressar o esquema argumentativo, e o *relato*, designado para expressar o esquema narrativo. A partir dessa divisão principal, Chaparro (1997; 1998) estabelece uma tipologia de subespécies que, na prática, é bastante próxima da tipologia de gêneros proposta por Marques de Melo (1994).

É inegável que os jornais diários tenham um repertório estável de formatos de veiculação de informações, em grande parte homogêneo, a despeito dos eventuais desvios (estilísticos, discursivos, estruturais, etc.). Marques de Melo (1992, 1994, 1998) tipifica os gêneros em categorias mais amplas, historicamente situadas<sup>62</sup>, que chamou de *jornalismo informativo* (compreendendo os gêneros *nota*, *notícia*, *reportagem* e *entrevista*), *jornalismo opinativo* (do qual fazem parte os gêneros *editorial*, *comentário*, *artigo*, *resenha*, *coluna*, *carta* e *crônica*) e *jornalismo interpretativo* (que conteria os gêneros *perfil*, *análise*, *cronologia* e *enquete*); a essas poderia ser adicionada outra categoria, como o *jornalismo utilitário*, relativamente

---

<sup>62</sup> Atribui-se a Samuel Buckley, diretor do jornal inglês *The Daily Courant* (1702-1735), a divisão básica *news/comments* (“notícias/comentários”). Cf. Marques de Melo (1994, p. 21-23) e Chaparro (1997, p. 6-8).

recente (do qual fariam parte os gêneros *roteiros*, *indicadores* e *chamadas*).

Numa abordagem influenciada pela classificação semiótica de Jakobson, Vélez (1985) teoriza sobre os gêneros encontrados no jornal “Folha de S.Paulo” e conclui que há três tipos básicos de textos jornalísticos: notícia, reportagem e crônica, que adquirem conformações próprias de acordo com as funções lingüísticas que assumem – respectivamente, referencial, emotiva-referencial e emotiva. Apesar de o trabalho não ser utilizado aqui, Vélez (1985) merece menção por detectar uma aproximação entre a reportagem e a entrevista, porque ambas têm como referentes “personagens ou pessoas” (VÉLEZ, 1985, p. 232).

Em 3.5.1 serão apresentadas as definições encontradas a respeito dos gêneros jornalísticos, que fundamentarão a classificação do corpus de estudo (cf. 4.6). Há citações dos autores já mencionados acima, de outros trabalhos (como NABANTINO RAMOS, 1970; LAGE, 1998; RABAÇA; BARBOSA, 1998; NOBLAT, 2002) e de definições extraídas de manuais de redação (principalmente as edições publicadas pela “Folha de S.Paulo”, que abordam diretamente o assunto). Propositamente buscou-se colher um maior número de definições daqueles gêneros sobre os quais há maior dificuldade de reconhecimento – entre eles, *notícia*, *reportagem*, *artigo*, *crítica* e *crônica*. Gêneros como *editorial* e *carta* dificilmente serão problemáticos para serem reconhecidos, pois, na maioria das vezes, encontram-se identificados graficamente de forma inequívoca na página do jornal.

### 2.3.7 Discussão

Os pontos de vista apresentados nesta seção mostram a riqueza de análise proporcionada pelo estudo dos gêneros, dando margem para a exploração de vários aspectos de sua constituição, forma interna e tipologia.

Sob a égide da Lingüística Aplicada, é preponderante a visão geral, também compartilhada nesta pesquisa, de que o contexto sociocultural *arbitra* e *modela* o gênero (HALLIDAY; HASAN, 1989; SWALES, 1990; BHATIA, 1993; BAKHTIN,

2000), seja ele visto como fenômeno que ocorre no *texto* ou no *discurso*. O gênero deixa suas marcas na linguagem como *registro* (HALLIDAY; HASAN, 1989; EGGINS; MARTIN, 1997), e isso implica a possibilidade de verificar se existem em um corpus semelhanças e diferenças de natureza lingüística entre, por exemplo, grupos de gêneros.

Este trabalho decidiu adotar uma definição que pode indicar caminhos para a superação das diferenças entre os conceitos de gênero e registro, proposta por Lee (2001):

É proveitoso enxergar o gênero e o registro como de fato dois diferentes ângulos ou pontos de vista, o *registro* sendo usado quando falamos de padrões léxico-gramaticais e discursivo-semânticos associados a situações (por exemplo, padrões lingüísticos), e *gênero* sendo usado quando falamos sobre a *afiliação a categorias reconhecidas culturalmente*. [último grifo nosso] (LEE, 2001, p. 46<sup>63</sup>)

Biber (1995; 1998; BIBER et al., 1998) é uma referência igualmente relevante na composição e análise dos grupos de gêneros. Suas idéias são empregadas neste trabalho na medida em que criam um terreno propício para a comparação entre gêneros ao longo dos eixos das dimensões, nuançando oposições antes consideradas como dualistas.

Como dito em 2.3.5, uma importante distinção teórica proposta por Biber (1994, p. 380) é a diferença entre gênero/registo e tipo de texto. Ora, esta pesquisa lidará positivamente com essa distinção, de forma a aproveitar sua ambivalência intrínseca para aperfeiçoar um modelo de classificação em que haja certa correlação entre os critérios estabelecidos *externamente* (ou seja, por meio de uma classificação do corpus em gêneros) e *internamente* (por meio de uma classificação do corpus em tipos de texto) – portanto, convergindo com as recomendações feitas por EAGLES

---

<sup>63</sup> No original, “it is useful to see the two terms *genre* and *register* as really two different angles or points of view, with *register* being used when we are talking about lexico-grammatical and discourse-semantic patterns associated with situations (i.e., linguistic patterns), and *genre* being used when we are talking about memberships of culturally-recognisable categories”.

(1996a e 1996b) e possibilitando estudos como Biber (1993). Tognini-Bonelli salienta que os “critérios situacionais e lingüísticos não são obviamente independentes um do outro; um dos principais usos de um corpus é investigar as correlações entre os dois” (TOGNINI-BONELLI, 2001, p. 60<sup>64</sup>), porém lembra que se deve afastar o risco de ocorrer um círculo vicioso na investigação, decorrente do fato de não haver clara separação entre os dois tipos de critério. Iniciar a pesquisa de um corpus com critérios situacionais de seleção, caso deste trabalho, é um “primeiro passo seguro”, pois são reconhecíveis de imediato, diferentemente dos critérios lingüísticos (BIBER, 1994, p. 380-381, apud TOGNINI-BONELLI, 2001, p. 60).

Apesar de o conceito de Bakhtin (2000) sobre gênero não estar diretamente empregado neste trabalho, no exame de sua teoria foram encontradas semelhanças com alguns princípios da Lingüística de Corpus, que serão mostradas abaixo a título de reflexão.

A primeira semelhança é em relação à definição de gênero feita por Bakhtin (2000, p. 279) – “tipos *relativamente* estáveis de enunciados”. A utilização do termo “relativamente” é indício de que o gênero, para Bakhtin, admite *variação*. Aí reside uma afinidade teórica com a conceituação desenvolvida em 2.1, que destaca a importância da observação da língua em uso, em toda a sua variedade de formas, como meio de identificação de padrões na linguagem.

Do mesmo modo, também é possível considerar a definição bakhtiniana de *enunciado* em convergência com a idéia de *texto*, oriunda de Firth e Halliday, na qual o texto é indissociável do contexto (cf. 2.1.1). Por sua vez, o inventário composto pelas “palavras dos outros”, defendido por Bakhtin (2000), embute uma noção de *freqüência*, a qual foi também examinada por Firth e Sinclair (cf. 2.1). Na medida em que *co-ocorrem* os enunciados é que existe a possibilidade de emergirem *padrões* na língua (cf. 2.1), percebidos de maneira consciente ou não.

---

<sup>64</sup> No original, “Situational and linguistic criteria are obviously not independent of one another; one of the principal uses of a corpus is to investigate the correlations between them”.

Vian Jr. (2001) faz o mesmo exercício de comparação – embora seja ali comparada a linha sistêmico-funcional com a linha bakhtiniana – e observa que existe uma preocupação mútua com o “meio social (a condição concreta) em que a linguagem é utilizada/produzida” em ambas as premissas de análise, embora frise que há divergência quanto aos “pontos de chegada”.

O conceito de gênero utilizado nesta pesquisa, em síntese, coincide com Swales (1990) quando esse autor qualifica o gênero como produto de uma convenção social e de uma prática inscrita em um contexto cultural. Não nos debruçaremos, porém, sobre a questão da comunidade discursiva, proposta pelo autor – que recentemente, aliás, reviu esse conceito (cf. SWALES, 1992). Da corrente neofirthiana, será levado em conta um pressuposto teórico importante, que postula a existência de marcas do significado presentes em vários níveis da linguagem. As ferramentas de análise da gramática sistêmico-funcional, propostas por Halliday (1994), não serão utilizadas, bem como o instrumental de análise de gêneros proposto por Halliday e Hasan (1989) e Eggins e Martin (1997). O motivo é que a análise que os autores propõem, de descoberta das estruturas genéricas dos textos, impossibilita o exame de um corpus de grandes dimensões, como o empregado aqui.

Dado que o objetivo deste trabalho é estudar a variação lingüística de uma amostra da língua portuguesa produzida nos jornais e utilizar, subseqüentemente, grupos de gêneros definidos socioculturalmente, na etapa de *classificação* dos gêneros propriamente ditos, justifica-se o emprego da tipologia oferecida pela área de Estudos de Comunicação – ela será, portanto, o substrato teórico escolhido para a classificação de gêneros do corpus de estudo (cf. 3.5).

A terminologia proposta por Biber, no tocante à distinção entre gênero/registo e tipo de texto, é considerada útil para esta pesquisa, como mencionado anteriormente. Foi feita a opção, no entanto, de designar “gênero” o que o autor hoje em dia considera “registo”. A proximidade conceitual dos dois termos, como salienta Lee (2001), faz-nos pensar também que ambos são perspectivas diferentes de um mesmo fenômeno. Por fim, foi procurado mostrar que as abordagens da área de

Estudos de Comunicação a respeito dos gêneros que ocorrem no jornal são absolutamente necessárias para o estabelecimento de um rol de princípios que norteiem uma categorização – porque, afinal, o gênero tem como função principal atuar como

“... um meio de classificar textos. ... Proporcionando um meio de descrever sistematicamente os textos, habilita-se a discussão sobre a natureza das diferentes formas textuais e dá-se sustento à criação de uma metalinguagem” (STANTON, 1996, p. 7<sup>65</sup>).

Este capítulo, em conclusão, preocupou-se fundamentalmente em explicitar a base teórica para fundamentar a consecução desta pesquisa, cuja metodologia será explanada a seguir.

---

<sup>65</sup> No original, “One of the primary functions of the category of genre is to act as a means of classifying texts. (...) In providing a means of systematically describing texts, it enables the discussion of the nature of different text forms and also makes provision for a metalanguage to be created”.

### 3 METODOLOGIA DE PESQUISA

O intuito deste capítulo é expor os principais procedimentos metodológicos utilizados na coleta de dados (cf. 3.2) e na Análise Multidimensional (AMD), instrumento fundamental para a investigação proposta (cf. 3.3). Inicialmente, porém, serão reiterados o objetivo deste trabalho e as questões de pesquisa formuladas (cf. 3.1), bem como a forma pela qual ambos orientam os critérios escolhidos para o desenho do corpus e a metodologia aplicada na pesquisa.

#### 3.1 OBJETIVO E QUESTÕES DE PESQUISA

O objetivo desta pesquisa é identificar empiricamente semelhanças e diferenças entre gêneros jornalísticos, a partir de um corpus composto por textos em língua portuguesa da imprensa escrita diária veiculados no Brasil. Para a sua consecução, a pesquisa emprega os recursos teóricos e de análise proporcionados pela Lingüística de Corpus, como o estudo de palavras-chave, a obtenção da freqüência de características e categorias lingüísticas e a metodologia introduzida por Biber (1988), a Análise Multidimensional de variação de registro. Diante desse objetivo, formulamos três questões de pesquisa básicas, que expressam as linhas de investigação sobre as quais se dá este trabalho:

1. Utilizando a metodologia empregada em Biber (1988), que características e categorias do português sobressaem na variação lingüística de um corpus da imprensa escrita diária?
2. Quais são os principais fatores responsáveis pela variação da linguagem presente no corpus jornalístico e como eles se comportam em relação aos subcorpora de gêneros?
3. Que funções comunicativas (isto é, dimensões) expressam os fatores obtidos e como os gêneros se posicionam ao longo das dimensões?

As três questões apóiam-se no pressuposto teórico expresso por Firth (cf. 2.1) de que o significado está distribuído ao longo do texto, entre os seus componentes constituintes, devendo haver, portanto, uma instância de expressão do significado nas características e categorias lingüísticas que o formam – como nos de ordem gramatical, lexical e discursiva. A esse respeito, como ilustração de como se dá essa distribuição, Hoey (1991), ao estudar as relações entre léxico e texto, ressalta a intensa ligação entre esses dois campos na produção comunicativa, afirmando: “Cada seleção textual coage as escolhas lexicais possíveis, e é nessa combinação entre escolhas lexicais e textuais efetuadas por escritores ou falantes que a sua criatividade é expressa” (HOEY, 1991, p. 217<sup>66</sup>). A pesquisa, nesse sentido, irá utilizar as ferramentas de análise desenvolvidas pela Lingüística de Corpus para observar os padrões lingüísticos que emergem da comparação entre a variedade de textos presentes em um jornal brasileiro para extrair descrições lingüísticas caracterizadoras de gêneros classificados originalmente de acordo com valores provenientes do contexto sociocultural.

Assim, como visto no capítulo anterior, para obter as respostas necessárias a essas questões proceder-se-á empiricamente a uma observação do corpus em um nível macrolingüístico, combinada a uma análise microlingüística de identificação de traços relevantes, com auxílio de uma metodologia de pesquisa de natureza essencialmente quantitativa, para a avaliação da freqüência das características e categorias lingüísticas co-ocorrentes.

---

<sup>66</sup> No original, “each textual selection constrains the lexical choices possible, and it is in the combination of the lexical and textual choices that writers or speakers make that their creativity is expressed”.

### 3.2 DESENHO DO CORPUS

Nesta seção descreveremos os principais procedimentos utilizados na pesquisa para a coleta dos dados. Inicialmente, é preciso justificar a escolha do jornal “Foha de S.Paulo” como um representante válido do jornalismo dito “de prestígio” (CHAPARRO, 1997). A primeira razão para a escolha é que o jornal, fundado em 1921, atualmente é o título mais vendido no país<sup>67</sup>, além de possuir um alcance nacional reconhecido; em segundo lugar, porque é o jornal que foi objeto do maior número de estudos relacionados ao assunto aqui tratado (cf. VÉLEZ, 1985; MARQUES DE MELO, 1987, 1998a e 1998b; CHAPARRO, 1997; BONINI, 2003).

Para a montagem do corpus de estudo, foram coletados, de forma integral<sup>68</sup>, textos de sete edições de um jornal de grande circulação nacional (“Folha de S.Paulo”), de modo a ser possível extrair um ciclo reconhecidamente completo, no âmbito do contexto sociocultural, de *uma semana* (MARQUES DE MELO, 1987, 1998a). Esta semana, no entanto, não coincide com uma série seqüencial de edições diárias. A semana foi artificialmente construída (cf. 3.2.1), para que a cobertura diária dos acontecimentos do jornal não exerça uma influência que possa distorcer os dados de contagem de freqüência lexical.

Esquemáticamente o corpus de estudo é assim descrito, de acordo com o número de textos, palavras e formas:

---

<sup>67</sup> Entre todos os jornais brasileiros, a “Folha de S.Paulo” teve a maior circulação média diária no ano de 2004 (307.703 exemplares). Em segundo lugar, está “O Globo” (257.451 exemplares), seguido por “Extra” (243.357 exemplares) e “O Estado de S. Paulo” (233.471 exemplares). Fonte: Instituto Verificador de Circulação (IVC).

<sup>68</sup> Entende-se aqui por extração integral dos textos de uma edição diária a soma dos conteúdos publicados nas edições local e nacional do veículo, eliminados os textos repetidos ou que sofreram redução de uma edição para a outra.

TABELA 1 - COMPOSIÇÃO E CARACTERÍSTICAS DO CORPUS DE ESTUDO

	<b>Edições diárias da “Folha de S.Paulo”</b>	<b>Textos</b>	<b>Palavras (tokens)</b>	<b>Formas (types)</b>
<b>Corpus de estudo</b>	7 (semana construída)	1.431	493.780	35.236

Observa-se que o corpus de estudo pode se afigurar como um corpus de dimensão média, segundo terminologia proposta por Berber Sardinha (2000a, p. 346). Tendo como base a tipologia empregada pelo autor (BERBER SARDINHA, 2000a, p. 339-342) para a classificação de corpora, o corpus de estudo desta pesquisa possuiria o seguinte perfil:

- a) *modo* escrito (e, mais especificamente, impresso);
- b) de um *corte temporal* contemporâneo (tempo corrente);
- c) *selecionado* por um critério de amostragem (uma porção finita de textos representa a linguagem como um todo);
- d) de *conteúdo* especializado (representa gêneros razoavelmente delimitados) e relativamente equilibrado (reflete os pesos relativos de cada gênero em um ciclo semanal de edições diárias);
- e) de *autoria* de falantes nativos;
- f) com *fins* de estudo (é o corpus que se busca descrever);
- g) pluriautorais;
- h) formado por textos completos.

O último ponto listado difere da metodologia utilizada por Biber (1988) na montagem do seu corpus de estudo, na qual foram utilizados fragmentos de textos maiores para igualar os textos em 2.000 palavras cada. O procedimento adotado por

Biber (1988) foi criticado pela despreocupação com “certas características constitutivas da textualidade” (BERBER SARDINHA, 2000b, p. 123-124) e por privilegiar o nível da palavra, e não o do texto. A utilização de textos completos adotada aqui é, ademais, convergente com os argumentos que explicitam a necessidade de considerar o texto como uma unidade (FIRTH, HALLIDAY, SINCLAIR, STUBBS), como exposto no capítulo anterior.

### 3.2.1 Semana Construída

O conceito de “semana construída” foi extraído de Kennedy (1998), mas, segundo Westin (2002, p. 11), que utilizou o mesmo procedimento, foi originalmente desenvolvido por Jones e Carter Jr. (1959) e posto em uso em trabalhos como Bell (1991). A Tabela 2 mostra a composição do corpus de estudo, que aplica o conceito.

TABELA 2 - EDIÇÕES DIÁRIAS DA “FOLHA DE S.PAULO” UTILIZADAS NO CORPUS DE ESTUDO

<b>Edição da “Folha de S.Paulo”</b>	<b>Dia da semana</b>	<b>Número de textos selecionados</b>
17.jan.2003	Segunda-feira	192
03.jun.2003	Terça-feira	157
06.ago.2003	Quarta-feira	211
27.mar.2003	Quinta-feira	213
30.mai.2003	Sexta-feira	215
19.jul.2003	Sábado	177
19.out.2003	Domingo	266
<b>Total de textos:</b>		<b>1.431</b>

A seleção de sete edições diárias do jornal, reunidas aleatoriamente de um

universo determinado – as 52 semanas que, somadas, formam o ano de 2003 –, foi realizada com o sorteio de uma edição de segunda-feira de uma semana  $x$ , de uma edição de terça-feira da semana  $y$ , e assim por diante. O intuito do procedimento é diluir o impacto de coberturas jornalísticas longas, que se repetem em dias seguidos (eleições, eventos esportivos, guerras, etc.), capazes de provocar “distorção excessiva na amostragem” (KENNEDY, 1998, p. 75).

Foram *incluídos* todos os textos jornalísticos veiculados nas edições local e nacional do dia (há uma versão “Nacional” e uma versão “São Paulo”, finalizadas em horários distintos e destinadas a diferentes localidades), compilados manualmente de modo a evitar textos repetidos ou reduzidos de uma edição para a outra. Foram *excluídos* da composição do corpus os seguintes elementos: tabelas, gráficos, infográficos (*story-board*, ilustrações com textos inseridos, etc.), mapas, cotações, fragmentos de obras literárias, frases, legendas de fotos, textos-legendas (texto curto que acompanha uma imagem), publicidade de qualquer natureza, propaganda legal (atas, avisos, etc.), apostilas e testes de educação formal, roteiros (listas de estabelecimentos, atrações, etc.), horóscopo, obituário, expediente, classificados e previsão do tempo.

A escolha em favor de um período de uma semana de jornal para determinar o corpus de estudo também converge com as considerações de Hunston (2002) sobre a necessidade de o desenho do corpus relacionar as características de equilíbrio com as de representatividade. A obtenção de um corpus representativo de uma língua envolve “quebrar o todo em partes componentes e buscar incluir quantidades iguais de dados de cada uma das partes”. No desenho de corpora provenientes de textos jornalísticos, uma medida de equilíbrio é “incluir todos os números de uma seleção de publicações de uma dada semana, mês ou ano. Isso deixaria que as proporções fossem determinadas por si mesmas” (HUNSTON, 2002, p. 28<sup>69</sup>). Dada a dificuldade de

---

<sup>69</sup> No original, “in the case of newspapers, the solution may be to include all issues of a selection of publications from a give week, month or year. This will allow the proportions to determine themselves”.

estudar uma quantidade tão grande como a resultante de um ano de textos, ou mesmo um mês, a pesquisa estabeleceu o horizonte de uma semana de textos veiculados pelo jornal.

A pesquisa tem como objeto de análise, em resumo, um total de 1.431 textos, que são o resultado da coleta de sete edições diárias de um jornal de circulação nacional no Brasil, formando uma semana completa – *unidade cíclica cultural*, para a veiculação de conteúdo editorial estável na imprensa escrita diária, relativamente consensual. O equilíbrio buscado nesse procedimento residiria no fato de que o corpus coletado contém proporcionalmente a quantidade de textos de cada gênero de acordo com sua veiculação efetiva, distribuídos tematicamente em editorias pelos cadernos veiculados na semana, sejam eles de periodicidade diária (cobertura de assuntos ligados à cidade, à política, ao esporte, etc.) ou semanal (como os temas ligados a literatura, turismo, automóveis, entre outros).

### 3.2.2 Etiquetagem Morfossintática

O corpus de estudo possui duas versões utilizadas na prática de pesquisa: a primeira é a composta pelo grupo de textos descrito acima, de 1.431 textos provenientes de sete edições; a segunda, por sua vez, é o mesmo corpus transformado formalmente por sofrer uma etiquetagem de cunho morfossintático<sup>70</sup>. Essa etiquetagem pode ser definida como uma ferramenta capaz de “inserir automaticamente, no corpus, códigos que indicam a classe gramatical de cada palavra” (BERBER SARDINHA, 2004a, p. 113; cf. vários tipos de etiquetadores em BERBER SARDINHA, 2004a, p. 113-142).

A etiquetagem é, assim, o instrumento que possibilita a contagem extensa de frequência de características e categorias lingüísticas. Daí a importância de a etiquetagem, que é automática, necessitar ter um alto índice de confiabilidade. A

---

<sup>70</sup> Em inglês, utiliza-se o termo *part of speech* e a designação abreviada “POS tagging”.

ferramenta adotada na pesquisa para a montagem da versão etiquetada do corpus de estudo foi o VISL (“Visual Interactive Syntax Learning”; cf. BERBER SARDINHA, 2004a, p. 136-142), criado pela Universidade de Odense, na Dinamarca. Segundo Bick, a versão do VISL para a língua portuguesa possui um índice de confiabilidade de 99% em termos de classe de palavras e flexão e de 97% a 98% para sintaxe (BICK, 2005, p. 100). Em paralelo, o acesso via internet<sup>71</sup> e a interface amigável permitem uma utilização facilitada a pesquisadores, o que faz do VISL uma escolha natural<sup>72</sup> no âmbito da presente pesquisa. Acrescente-se o fato de haver poucas opções disponíveis de etiquetadores para a língua portuguesa (BERBER SARDINHA, 2004a, p. 153).

Em contrapartida, muito da definição de variáveis de estudo ficou condicionado ao sistema de classificação adotado pelo VISL, baseado na Gramática Constritiva<sup>73</sup> (KARLSSON et al., 1995).

A relação das etiquetas inicialmente estudadas na pesquisa está no Anexo A (cf. Anexo B com exemplo de texto etiquetado). Foi realizada uma pré-seleção das etiquetas obtidas no processo de etiquetagem, de acordo com os seguintes critérios:

- a) **frequência**: foram excluídas etiquetas com frequência muito baixa no corpus, pois são inadequadas para integrarem um análise multivariada como a Análise Fatorial (cf. LEE, 2000);
- b) **sobreposição**: foram excluídas etiquetas contadas coincidentemente sobre o mesmo tipo de palavra, ocasionando uma correlação direta entre duas variáveis; por exemplo, as variáveis participio passado e voz passiva;

---

<sup>71</sup> A página do VISL disponível para a etiquetagem de textos em português é <http://visl.sdu.dk/visl/pt/parsing/automatic/parse.php>. Consultado em 21/08/2005.

<sup>72</sup> Reforça a decisão de utilizá-lo a afirmação: “O etiquetador do projeto VISL é o único confiável dentre os disponíveis on-line para a nossa língua” (BERBER SARDINHA, 2004a, p. 153).

<sup>73</sup> Em inglês, “Constraint Grammar”.

- c) **interpretabilidade:** foram incluídas variáveis que, direta ou indiretamente, foram estudadas por outros autores;
- d) **acuidade:** foram incluídas somente variáveis cujas etiquetas morfossintáticas apresentaram um alto índice de confiabilidade;
- e) **contrastabilidade:** foram mantidas variáveis que, nas diversas extrações que precederam a solução de Análise Fatorial final, obtiveram seguidamente cargas contrastadas entre os grupos de gêneros.

Acreditamos ter fornecido as definições necessárias para configurar devidamente o desenho do corpus de estudo para esta pesquisa, bem como para esclarecer questões metodológicas concernentes à escolha das variáveis utilizadas. A seguir serão tratados os procedimentos de análise dos dados, vinculados à metodologia da Análise Multidimensional.

### 3.3 ANÁLISE MULTIDIMENSIONAL: UMA ABORDAGEM

Esta seção apresenta os procedimentos estatísticos utilizados sobre os dados provenientes da contagem de frequências do corpus de estudo, para a obtenção de resultados que satisfaçam o cumprimento do objetivo e que provejam respostas adequadas às questões de pesquisa, ambos explicitados no início deste capítulo. Primeiramente, na introdução (cf. 3.3.1), será apresentada uma visão panorâmica do método desenvolvido por Biber, a Análise Multidimensional, seguida de uma explicitação da terminologia utilizada e dos trabalhos que a utilizaram. A seguir serão discutidas em profundidade todas as etapas da Análise Multidimensional, bem como aspectos da técnica estatística original de Biber (1988) que serão adaptados para esta pesquisa, a partir de sugestões encontradas em Lee (2000).

### 3.3.1 Introdução - Análise Multidimensional

Conforme visto em 2.2, a Análise Multidimensional compõe-se de um processo estatístico complexo, em que é analisada por meio de uma Análise Fatorial a frequência de ocorrências de diversas variáveis em um grupo de textos, de modo a obter combinações de variáveis que se distribuem com certa uniformidade no corpus. Tais conjuntos são designados fatores, cabendo ao pesquisador escolher a melhor solução estatística que os agrupe. Depois de interpretados, os fatores passam a ser chamados de dimensões, pois a partir daí haveria uma correspondência entre as características lingüísticas agrupadas e uma função percebida de cunho comunicativo.

A Análise Fatorial foi desenvolvida originalmente há cerca de um século (1904), pelo matemático Spearman, que a aplicou na área da psicologia, onde até hoje é utilizada, tendo como fonte de dados usual indivíduos e testes a eles aplicados (LEE, 2000, p. 156-157). O uso pioneiro da técnica na área da linguagem é atribuído a Carroll (1960). Biber (1988), como dito anteriormente, utilizou o modelo para descrever as dimensões envolvidas na produção oral e escrita da língua inglesa, e observar suas semelhanças e diferenças. Outros autores (e o próprio Biber) expandiram o modelo utilizado em *Variation Across Speech and Writing* em várias perspectivas de análise – diacrônica, outras línguas, etc. (cf. 3.4). Antes de examinar essas questões, porém, será exemplificado o funcionamento da metodologia da Análise Multidimensional como um todo e esclarecidos alguns pontos ligados à terminologia.

### 3.3.2 Etapas da Análise Multidimensional

Esquemáticamente, utilizando o elenco feito por Berber Sardinha (2000b), essas são as etapas para a consecução de uma Análise Multidimensional:

- 1) Coleta ou adoção de um corpus de dados lingüísticos representativo e compatível com as metas da análise.
- 2) Levantamento das características lingüísticas relevantes para análise por meio de ampla consulta à literatura disponível.
- 3) Transformação das características lingüísticas em variáveis quantificáveis.
- 4) Codificação dos dados baseada nas variáveis selecionadas, usando-se ferramentas computacionais para análise automática, semi-automática (interativa) ou manual.
- 5) Conferência manual da codificação feita por computador para se assegurar de sua exatidão.
- 6) Computação de frequências médias de cada variável.
- 7) Padronização das frequências (em geral por 1.000 palavras), para permitir a comparação entre variedades (textos, registros ou corpora) de extensões diferentes (normalização).
- 8) Análise Fatorial inicial, a fim de se obterem os pesos (“*loadings*”) de cada variável em cada variedade.
- 9) Determinação do número de fatores, por meio da aplicação de técnicas como observação dos valores eigen (“*eigenvalues*”) em um gráfico “*scree*” (“*scree plot*”).
- 10) Análise Fatorial posterior, fazendo-se a rotação dos fatores para se evitar sobrecarga de um fator.
- 11) Interpretação de cada fator e rotulação das dimensões.
- 12) Cálculo de escores de cada texto por fator, através da padronização dos escores com base na média e desvio padrão.
- 13) Cálculo de escores médios de cada variedade por fator.

14) Testes de precisão e “*recall*”<sup>74</sup> entre gêneros e tipos de texto (opcional).

Cada item será explorado em maior ou menor grau de detalhe neste capítulo e no seguinte, de Apresentação e Discussão dos Resultados. Antes, porém, será descrito de maneira global o processo da Análise Multidimensional.

### 3.3.3 Análise Multidimensional da Língua Inglesa: Um Exemplo

Biber (1988) realizou uma inédita investigação exploratória na língua inglesa, tentando abarcar em seu corpus uma ampla variedade de gêneros escritos e falados (23 ao todo), como prosa acadêmica, editoriais ou ficção de mistério (no primeiro grupo), e conversas face a face públicas, discussão de pôster ou transmissão de rádio (no segundo), em conjuntos uniformes de 481 textos de 2.000 palavras cada, totalizando um corpus de cerca de 960 mil palavras.

O passo inicial do estudo foi identificar quais características e categorias lingüísticas poderiam ser pesquisadas no corpus, e para isso Biber valeu-se de estudos anteriores, como Chafe (1982) e Quirk et al. (1985), para determinar qual conjunto de variáveis com poder de “associação funcional” (BIBER et al., 1998, p. 145) seria incluído na Análise Fatorial posterior. Ele selecionou características e categorias lingüísticas pertencentes a 16 “categorias gramaticais”: marcadores de tempos e aspectos verbais (presente, pretérito perfeito), advérbios de lugar e de tempo, pronomes, interrogações, formas nominais, passivas, verbos de estado (como “*do*” no

---

<sup>74</sup> É possível realizar testes de precisão e “*recall*” (ou cobertura) sobre um universo amostral que possua um determinado subgrupo de indivíduos conformes (sob qualquer critério). Uma pesquisa que intentasse separar só os indivíduos do subgrupo teria resultado de grande precisão se trouxesse em conseqüência um número de indivíduos desconformes próximo de zero. Já um resultado que apresentasse grande “*recall*” proporcionaria um número de conformes muito próximo ao valor total do subgrupo observado.

papel de verbo principal), características de subordinação (complementos, formas de particípio, pronomes relativos, orações subordinadas adjetivas e adverbiais), adjetivos (atributivos e predicativos) e advérbios (total), especificidade lexical (razão forma / palavra, tamanho das palavras), classes lexicais (conjunções, salvaguardas, pronomes demonstrativos, etc.), modais (de possibilidade, de necessidade, de predição), classes especiais de verbo (públicos, pessoais, persuasivos), formas reduzidas (contrações, por exemplo), coordenação e negação. As características selecionadas, de ordem lingüística e situacional, foram formalmente descritas, na forma de listas de palavras, expressões algorítmicas e fórmulas de cálculo. A partir desse momento, as características lingüísticas são tratadas como variáveis que compõem a Análise Fatorial.

Em seguida o autor utilizou programas de computador para identificar e contar as freqüências de cada uma das variáveis, revisadas manualmente ao cabo do processo automatizado (ou semi-automatizado). A atenção agora é voltada para a distribuição das variáveis entre os textos: sua freqüência média, mínima e máxima, além do desvio padrão (que indica se há homogeneidade na distribuição da variável). Buscando encontrar uma medida comum de comparação entre textos de extensões diferentes, Biber normalizou os resultados obtidos – ou seja, padronizou as freqüências (por 1.000 palavras) a fim de igualar o peso das variáveis em textos de diversos tamanhos.

Com os dados de freqüência reunidos, Biber submeteu as variáveis à Análise Fatorial, que processa as informações e identifica os conjuntos de variáveis que se distribuem de forma semelhante. Para que os fatores tenham uma composição de variáveis otimizada e sejam mais facilmente interpretados, é feito um procedimento estatístico de *rotação oblíqua*, porque o método de Análise Fatorial produz originalmente uma matriz fatorial que é difícil de interpretar. É necessário, desse modo, utilizar um método de rotação, que faz com que os fatores resultantes da matriz fatorial se tornem mais inteligíveis. A opção adotada na pesquisa de Biber (1988) foi a de utilizar a rotação oblíqua Promax (que aumenta a intercorrelação entre os fatores),

supondo que as dimensões que expressam funções na linguagem também o sejam.

A seguir os fatores são dispostos em um gráfico “*scree*”, onde é feita a seleção definitiva dos fatores que comporão a Análise Fatorial. Esses fatores resultantes serão interpretados e convertidos em dimensões, segundo a função comunicativa que exercem no corpus.

As dimensões, produto da interpretação dos fatores resultantes da Análise Fatorial, são interpretadas com base nas “funções situacionais, sociais e cognitivas” (BIBER et al. 1998, p. 146-147) que as características e categorias lingüísticas presentes nos fatores representam e compartilham entre si. O estudo de Biber (1988) chegou originalmente a sete dimensões de variação da língua inglesa oral / escrita, mas no mesmo trabalho esse número baixou para seis; posteriormente (BIBER et al., 1998) esse número foi reduzido a cinco dimensões, abaixo denominadas:

- 1) Produção com interação versus produção informacional;
- 2) Discurso narrativo versus não-narrativo;
- 3) Referências explícitas versus referências dependentes do contexto;
- 4) Expressão explícita de argumentação;
- 5) Estilo impessoal versus não-impessoal<sup>75</sup>.

No intuito de esclarecer como se distribuem as variáveis nas dimensões, listaremos as variáveis que compõem a Dimensão 1 de Biber (1988), na Tabela 3. Como se perceberá abaixo, cada variável recebe um *peso* (chamado também de *carga*<sup>76</sup>), que corresponde à “correlação da variável à dimensão ou ao fator subjacente” (LEE, 2000, p.157<sup>77</sup>). O peso pode adquirir um valor entre -1 e 1 (quanto mais

---

<sup>75</sup> Tradução adaptada de Pacheco de Oliveira (1997), com modificações revistas por Biber et al. (1998, p. 148) incorporadas.

<sup>76</sup> Em inglês, “loading”.

<sup>77</sup> No original, “correlation of the variable to the underlying dimension or factor”.

próximo esse valor estiver de 1 ou -1, mais significativo será o peso da variável; alternativamente, se o peso estiver próximo de zero, seja positivo ou negativo, ele não terá uma carga significativa na dimensão). Enquanto valores positivos indicam que a variável co-ocorre acima da frequência média, o contrário ocorre com variáveis de valores negativos (a ausência da variável é que passa a ser significativa, encontrando-se em relação inversamente proporcional à presença de uma outra variável de valor positivo):

TABELA 3 - VARIÁVEIS QUE COMPÕEM A DIMENSÃO 1 (BIBER, 1988)

Variável	Peso
Verbos pessoais	0,96
Apagamento de “that”	0,91
Contrações	0,90
Verbos no tempo presente	0,86
Pronomes pessoais de 2ª pessoa	0,86
“Do” como verbo auxiliar	0,82
Negação analítica	0,78
Pronomes demonstrativos	0,76
Enfatizadores em geral	0,74
Pronomes de 1ª pessoa	0,74
Pronome “it”	0,71
“Be” como verbo principal	0,71
Subordinação causativa	0,66
Partículas discursivas	0,66
Pronomes indefinidos	0,62
Salvaguardas em geral	0,58
Amplificadores	0,56
Pronomes relativos	0,55
Perguntas WH	0,52
Modais de possibilidade	0,50
Coordenação não-frasal	0,48
Orações WH	0,47
Preposições finais	0,43
Substantivos	-0,80
Tamanho das palavras	-0,58
Reposições	-0,54
Razão forma / palavra	-0,54
Adjetivos atributivos	-0,47

(BIBER, 1998, p. 148; traduzido por BERBER SARDINHA, 2000b, p. 110)

Biber (1988) identifica na Dimensão 1 que os traços que obtiveram valores

positivos indicam

“um foco não-informacional e interativo, devido a um propósito interativo ou afetivo primordial e/ou a circunstâncias de produção altamente coercitivas. Esses traços podem ser caracterizados como verbais, interativos, afetivos, fragmentados, reduzidos na forma e generalizados em termos de conteúdo.” (BIBER, 1988, p. 105<sup>78</sup>)

Inversamente, os traços com valores negativos sugerem foco informacional e integração cuidadosa da informação no texto (representada pela presença de preposições), além de implicar um alto grau de densidade lexical (pelos índices de razão palavra / forma e extensão das palavras). Berber Sardinha assim resume os gêneros associados à Dimensão 1:

“Os registros que melhor representam o modo de produção com interação são as conversas, tanto ao telefone quanto cara a cara; os registros que melhor representam a produção informacional são documentos oficiais, reportagem jornalística e prosa acadêmica” (BERBER SARDINHA, 2000b, p. 112).

A partir dos valores atribuídos às variáveis, podem-se calcular os escores de cada texto, de modo a classificá-los, conforme sua relevância, ao longo das dimensões. Igualmente podem ser comparados gêneros como um todo (que representam um subgrupo de textos no corpus), tanto em relação à contabilização geral como entre os demais gêneros.

Esta exposição da Análise Multidimensional, paradigmática para a Lingüística de Corpus feita atualmente, teve o objetivo de ilustrar o *modus operandi* básico da principal metodologia utilizada neste trabalho. A seguir, serão discutidos tópicos relacionados à prática da Análise Multidimensional e algumas questões ligadas à terminologia utilizada na pesquisa.

---

<sup>78</sup> No original, “All of [the features] can be associated in one way or another with an involved, non-informational focus, due to a primarily interactive or affective purpose and/or to highly constrained production circumstances. These features can be characterized as verbal, interactional affective, fragmented, reduced in form, and generalized in content”.

### 3.4 ANÁLISE MULTIDIMENSIONAL: UMA PRÁTICA

Biber (1988) utilizou-se da técnica estatística multivariada da *Análise Fatorial*, responsável pela maior parte do arcabouço metodológico da Análise Multidimensional.

Como visto na seção precedente, a ambiciosa pesquisa realizada por Biber (1988) visou descrever a variação da língua inglesa nos modos oral e escrito por meio de *dimensões* – isto é, conjuntos de características lexicogramaticais co-ocorrentes que podem ser interpretadas em termos do papel que assumem *funcionalmente* no processo de comunicação. As dimensões formam *continua* distintos, ao longo dos quais é possível mapear a distribuição dos diversos gêneros textuais (ou registros, termo que posteriormente Biber preferiu utilizar). Assim, cada gênero é representado de forma mais ou menos marcada em cada uma das dimensões resultantes de uma Análise Multidimensional. Um dos pressupostos teóricos dessa análise é a existência de várias dimensões subjacentes à linguagem, que dão conta da variação lingüística existente – nesse caso, entre a fala e a escrita em inglês –, daí o fato de ser chamada de “*multidimensional*”.

O estudo de Biber (1988) alcançou notável repercussão entre os lingüistas de corpus (KENNEDY, 1998, p. 185-190; BERBER SARDINHA, 2000 e 2004a, p. 297-329; cf. LEE, 2000, p. 389-391) e influenciou a criação de pesquisas similares em outras línguas (BIBER, 1995), que valorizam o aspecto diacrônico (BIBER E FINEGAN, 1989; ATKINSON, 1992) ou, ainda, que estão voltadas para corpora mais especializados (BIBER, 1991; TRIBBLE, 1998). No Brasil, Pacheco de Oliveira (1997), Conde (2002), Santos (2002), Shergue (2003) e Berber Sardinha (2004b) são trabalhos relevantes que utilizaram a metodologia introduzida por Biber (1988). Tal interesse despertado justificou também a tentativa de replicar criticamente o experimento de Biber (LEE, 2000), de modo a testar o seu modelo com outros dados e aumentar a transparência metodológica do complexo procedimento estatístico em que se assenta a Análise Multidimensional, a Análise Fatorial.

A seguir serão descritos e detalhados em termos práticos os passos que compõem a metodologia desenvolvida por Biber (1988), incorporando a ela algumas sugestões dadas por Lee (2000) como medidas necessárias para a obtenção de soluções estatísticas mais rigorosas e de melhor qualidade.

### 3.4.1 Organização Preliminar dos Dados

A etapa preliminar da Análise Multidimensional está relacionada com a obtenção do corpus de estudo e com o levantamento de características lingüísticas que serão observadas no corpus. Nesse estágio, é necessário levar em conta alguns pré-requisitos: a) o corpus de estudo deve ser *representativo*, ou seja, ser composto por uma amostra equilibrada do que se queira representar, seja uma variedade (gênero, dialeto, nível de aprendizado, etc.), um modo (fala, escrita) ou até um idioma inteiro<sup>79</sup>; b) a seleção das características e categorias lingüísticas que irão compor o leque de *variáveis*<sup>80</sup> deve ser criteriosa – resultado de uma pesquisa extensa na literatura – e abranger um bom número de parâmetros que potencialmente sejam capazes de acusar a ocorrência de variação lingüística no corpus de estudo; c) os processos de *inserção* e

---

<sup>79</sup> Biber (1988) tentou extrair as dimensões do inglês utilizando um corpus de estudo que tinha textos orais / escritos na proporção de 30% / 70%, respectivamente: 290.000 palavras de gêneros do modo oral (conversas face a face e telefônicas, debates e entrevistas radiofônicas, discursos planejados e espontâneos) e 670.000 palavras de gêneros do modo escrito (gêneros jornalísticos, prosa acadêmica, ficção, cartas pessoais e profissionais, etc.). Como é impossível determinar qual é a proporção relativa de cada modo e gênero em toda uma língua e, mais importante, sabendo-se que a proporção dos diversos gêneros que compõem o corpus de estudo influencia diretamente os resultados da Análise Multidimensional, Lee (2000) entende que os resultados obtidos em Biber (1988) são heurísticos e preliminares, não devendo as dimensões encontradas serem tomadas categoricamente.

<sup>80</sup> O procedimento estatístico da Análise Fatorial, utilizado a seguir, trata toda característica / categoria quantificável como uma *variável*.

*contagem* de etiquetas que representam as características lingüísticas (variáveis) no corpus de estudo, em geral realizados com o auxílio de programas automáticos ou semi-automáticos (cf. BERBER SARDINHA, 2004a, p. 113-142), devem ser confiáveis (admite-se em geral uma margem de erro de até 5%) e checados manualmente.

A seguir, é necessário efetuar uma *normalização* dos dados obtidos na contagem das variáveis em cada texto do corpus de estudo. Ela visa estabelecer uma mesma medida de freqüência entre textos de tamanho diferentes. O padrão normalmente adotado (por 1.000 palavras) é obtido pela fórmula:

$$(\text{número de ocorrências da variável no texto} \div \text{número de palavras do texto}) \times 1.000$$

O uso de um programa de planilhas como o Excel é útil nessa fase da análise, porque armazena os dados brutos provenientes da contagem das características (variáveis) e os transforma facilmente no padrão “por 1.000 palavras”.

Na etapa seguinte, cada variável passa a ser estudada de acordo com sua *estatística descritiva*, em que se levantam informações sobre o comportamento da sua distribuição de freqüência no corpus. O objetivo, além de permitir uma comparação básica com dados de outros trabalhos, é reunir subsídios para avaliar se uma determinada variável possui uma freqüência rara<sup>81</sup> ou se possui uma distribuição muito irregular<sup>82</sup> – indícios de que poderiam comprometer a estabilidade e os resultados da Análise Fatorial (LEE, 2000)<sup>83</sup>.

A estatística descritiva das variáveis pode ser realizada no Excel ou no

---

<sup>81</sup> Menos de 3 ocorrências por 1.000 palavras, segundo Sigley, 1997.

<sup>82</sup> Desvio padrão > 1 ou curtose > 1 (valor positivo ou negativo) já indica que a distribuição da variável não é normal (Lee, 2000, p. 213).

<sup>83</sup> Porém, como será visto a seguir, Lee (2000) utiliza outros parâmetros para a exclusão de variáveis da Análise Fatorial.

programa estatístico em que será processada a Análise Fatorial. É preciso extrair os seguintes valores de cada variável (entre parênteses, o nome do rótulo que designa a função no Excel): Número mínimo de ocorrências da variável entre os textos do grupo (MÍNIMO); Número máximo de ocorrências da variável entre os textos do grupo (MÁXIMO); Média aritmética do total de ocorrências (MÉDIA); Amplitude, que é a diferença entre o número máximo e mínimo de ocorrências em um texto; Desvio padrão (DESVPAD), que designa a medida de dispersão ou variabilidade de uma distribuição de frequência; Assimetria<sup>84</sup> (DISTORÇÃO), que indica se há concentração de ocorrências nas extremidades da curva de distribuição; e Curtose (CURT), propriedade de uma distribuição de frequência que caracteriza a sua semelhança ou não com a curva de distribuição normal.

Com todos os dados preliminares obtidos, a pesquisa estará pronta para uma etapa crítica da Análise Multidimensional: a Análise Fatorial, procedimento estatístico que realiza a *redução* dos dados e propicia o agrupamento das variáveis co-ocorrentes em *fatores* (o mesmo que dimensões, exceto pelo fato de que os fatores ainda não possuem rótulos que posteriormente designarão a sua função). Todavia, é conveniente esclarecer previamente alguns pontos metodológicos relativos à Análise Fatorial e os meios para obtê-la.

A expressão “Análise Fatorial” pode ser entendida de duas formas: genericamente, denominando um conjunto de procedimentos estatísticos semelhantes<sup>85</sup>, ou, na acepção mais específica, com o método de extração chamado de “Principal Axis Factoring” (ou “Iterated Principal Factor Analysis”)<sup>86</sup>. Neste trabalho, “Análise Fatorial” equivale ao método “Principal Axis Factoring” (PAF), que foi o

---

<sup>84</sup> Em inglês, “skewness”.

<sup>85</sup> Sob o qual são agrupados vários métodos de extração, como o “Principal Components Analysis”, “Principal Axis Factoring”, “Maximum Likelihood” e “Image Factoring”, entre outros.

<sup>86</sup> Conforme a terminologia utilizada pelo tipo de software que o processa (SPSS e SAS, respectivamente). Biber (1988) utiliza a expressão “Principal Factor Analysis”.

utilizado na pesquisa de Biber (1988). É preciso ressaltar, no entanto, que vários autores também realizaram “Análise Fatorial” recorrendo ao método de análise de componentes principais (“Principal Components Analysis”)<sup>87</sup>. Segundo Lee (2000, p. 158-159, 162-167), apesar de o método de análise de componentes principais ser menos complicado de operar, não é o mais indicado para uma investigação exploratória como a Análise Multidimensional, porque leva em conta em seus resultados *toda* a variação encontrada nos dados (incluindo variância de erro e a variância específica), e não apenas a parcela de variação *compartilhada* entre as outras variáveis, como faz o “Principal Axis Factoring”.

No aspecto da prática de pesquisa, a Análise Fatorial é realizada com o auxílio de softwares estatísticos especializados, como o SPSS e o SAS. Recomendamos a utilização do SPSS, que tem uma interface mais amigável e recursos úteis, como a exportação de resultados em arquivos de várias extensões (.txt, .html, etc.). No SPSS, é possível importar diretamente as planilhas Excel com os dados da pesquisa (as variáveis devem estar dispostas nas colunas, enquanto as respectivas contagens de frequências encontradas nos textos, já normalizadas, ficam nas linhas da planilha), que depois são convertidos em arquivos com a extensão .sav (formato nativo do programa SPSS).

### 3.4.2 Análise Fatorial

Nesta seção, serão esclarecidos os parâmetros utilizados para a realização da Análise Fatorial, visando uma *extração inicial* que buscará reunir informações para as extrações seguintes. Além da apresentação de parâmetros utilizados para a extração e as caixas de diálogo do SPSS<sup>88</sup>, serão descritos as informações que resultam da

---

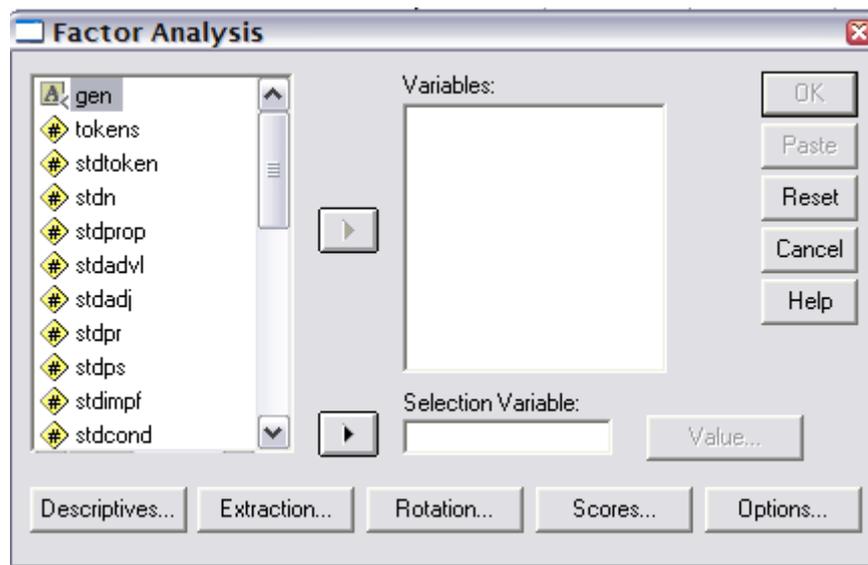
<sup>87</sup> No Brasil, Pacheco de Oliveira (1997), Shergue (2003) e Conde (2002), trabalhos baseados na Análise Multidimensional de Biber (1988), utilizaram todos a análise de componentes principais.

<sup>88</sup> Utiliza-se aqui a versão 11 do programa “SPSS for Windows”.

extração inicial e os elementos que servirão como base para a interpretação dos resultados, como as matrizes padrão e estrutural. A extração inicial difere das demais extrações apenas por não trazer definido o número de fatores a extrair. Não obstante, as extrações realizadas com um número determinado de fatores trazem as mesmas informações da extração inicial, que também devem ser analisadas pelo pesquisador, principalmente no tocante às comunalidades e às medidas de adequação de amostragem (MSA), ambas definidas a seguir.

### 3.4.2.1 Parâmetros

FIGURA 1 - ANÁLISE FATORIAL: TELA INICIAL DO SPSS



Uma extração inicial da PAF é feita com os dados normalizados<sup>89</sup> para gerar uma *matriz de correlação* (“correlation matrix”) entre as variáveis. A matriz de correlação produzida (não apresentada aqui) apresenta muitas redundâncias, e a Análise Fatorial nada mais é que um processo encarregado de reduzi-las a uma série

---

<sup>89</sup> Por razões algébricas, o número de textos deve ser maior que o número de variáveis. Alguns estatísticos propõem a proporção mínima de cinco textos por variável (cf. LEE, 2000, p. 166).

menor de variáveis, os fatores (LEE, 2000, p. 165). Para efetuar uma extração, é preciso em primeiro lugar abrir o arquivo com os dados da pesquisa, com a extensão .sav. Em seguida, selecionar no menu do SPSS a opção “Analyze” e os submenus “Data Reduction” e “Factor”.

A caixa de diálogo da Análise Fatorial que é aberta contém vários parâmetros a serem definidos (Figura 1), que serão explicitados a seguir.

a) Seleção das variáveis: a janela da esquerda lista as variáveis encontradas nos dados do arquivo. Selecionam-se as variáveis que irão compor a Análise Fatorial clicando com o mouse sobre a variável escolhida e apertando em seguida o botão central em que aparece uma seta para a direita (ou, diretamente, dando um clique duplo sobre a variável). As variáveis escolhidas aparecerão na janela da direita (“Variables”).

b) Botão “Descriptives”: há duas caixas de opções, “Statistics” e “Correlation Matrix”. Na primeira, é preciso deixar selecionada a opção “Initial solution”. Na segunda, “Anti-image” e “KMO e Bartlett’s test of sphericity”. O intuito, como será explicado a seguir, é verificar a adequação dos dados com o auxílio dos testes gerais preliminares, o de esfericidade de Bartlett e a medida de adequação de amostragem de Kaiser-Meyer-Olkin (cujas siglas em inglês são “KMO MSA” ou, simplesmente, “MSA”). Também faz parte da análise o exame da matriz antiimagem de correlação.

c) Botão “Extraction”: na caixa de seleção “Method”, selecionar “Principal Axis Factoring”. Devem também ser marcadas as opções “Correlation matrix” (em “Analyze”) e “Scree plot” (em “Display”). A área “Extract”, nessa etapa, deve ter selecionada apenas a alternativa “Eigenvalues over:”, com o valor default “1”. O valor “25”, default para o campo “Maximum iterations for Convergence”, fica mantido (nesta e na caixa de diálogo seguinte). Somente nas extrações seguintes, para definir o número de fatores que terá a solução da Análise Fatorial, deve-se marcar o número

desejado de fatores no campo “Number of factors”.

d) Botão “Rotation”: na área “Method”, selecionar “Promax”, com valor de “Kappa” igual a 3 (o default do SPSS é 4, mas deve ser alterado para alinhar-se com o mesmo grau de intercorrelação utilizado por Biber). Fica também marcado o campo “Rotated solution”, em “Display”, para produzir o resultado da rotação oblíqua.

e) Botão “Scores”: não utilizar. O cômputo de escores dos fatores originalmente utilizado por Biber (1988) é feito com o auxílio de *escalas somadas*, ou baseadas em fatores, que serão examinadas adiante.

f) Botão “Options”: o default da área “Missing Values” é a alternativa “Exclude cases listwise”. Para facilitar a leitura dos resultados da Análise Fatorial, selecionar a alternativa “Sorted by size”, que agrupa as variáveis de maior peso nos fatores, e marcar também a opção “Supress absolute values less than”, ajustando a janela adjacente com o valor de “0,30”.

Quando se aperta o botão “OK”, é feito o processamento da Análise Fatorial. Após alguns segundos, os resultados da extração serão visualizados em uma nova janela, no ambiente “SPSS Viewer”. A sua janela principal é subdividida em duas áreas: a da esquerda elenca os itens da extração pelo seu título, enquanto a da direita mostra os resultados correspondentes ao título em detalhe, como as matrizes de correlação e os testes. A cada nova extração, seus resultados virão visualizados abaixo dos anteriormente obtidos, caso o arquivo produzido no Viewer não seja fechado (é necessário salvar o arquivo antes de fechá-lo). Com o botão direito do mouse, é possível abrir um menu e selecionar a opção “Copy”, para a transferência do material selecionado para uma planilha ou editor de textos, ou “Export”, que produz um arquivo em formato .htm, .xml ou .txt da seleção (figuras são salvas em .jpg).

#### 3.4.2.2 Informações Obtidas

A análise dos resultados da extração inicial da Análise Fatorial é de

fundamental importância para os passos seguintes, que definirão novas extrações e análises subsequentes, até a extração “final”, definida pela estabilidade dos fatores e sua interpretabilidade. A seguir serão mostradas as informações obtidas na Análise Fatorial (pelo título que aparece na janela da esquerda do SPSS Viewer), assim como os critérios de avaliação a serem utilizados pelo pesquisador para considerar a qualidade da extração efetuada:

a) “Notes”<sup>90</sup> (notas): traz a documentação dos parâmetros utilizados na extração, como o nome do arquivo-base, as variáveis escolhidas, o método de extração e outros critérios selecionados. É útil para o pesquisador conservar esses dados, que servem para organizar futuras extrações.

b) “KMO and Bartlett’s Test” (KMO e teste de Bartlett): mostram o resultado de dois testes preliminares da extração, a medida geral de adequação de amostragem (MSA) e o teste de esfericidade de Bartlett. O primeiro quantifica globalmente (a partir da totalidade dos dados) em que medida as variáveis mantêm relação com as demais, justificando, portanto, uma Análise Fatorial, que se baseia na intercorrelação entre as variáveis. Valores acima de 0,70 são desejáveis, na faixa de 0,60 são medíocres, na dos 0,50, pobres e abaixo de 0,50 são inaceitáveis (NORUŠIS, 1988, p. 129, apud LEE, 2000, p. 175). O MSA também é obtido individualmente para cada variável, por meio da matriz antiimagem de correlação (v. abaixo). Já o teste de esfericidade de Bartlett verifica se a matriz resultante *não* é uma matriz de identidade (que ocorre quando não há correlação entre as variáveis, situação que invalidaria toda a extração). É, portanto, um teste de geral que apenas afere se a análise não é inválida e pode prosseguir (esteja ela certa ou não, como lembra Lee, 2000, p. 174). Para que a extração fatorial também não seja considerada inválida é preciso que o valor do teste atribuído à significância (“Sig.”) seja inferior a 0,05.

---

<sup>90</sup> Se as informações não estiverem visíveis na janela do lado direito, faça um clique duplo sobre o ícone, para mostrá-las.

c) “Anti-image Matrices” (matrizes antiimagem): traz uma tabela com duas matrizes. Só a referente à correlação (“anti-image Correlation”), na metade inferior da tabela, é analisada. Há dois aspectos importantes a serem verificados nessa matriz, a cada extração: os dados dispostos na *diagonal*, que traz os valores de MSA específicos de cada variável<sup>91</sup>, e os valores que se situam na parte inferior da diagonal, que revelam se há ocorrências de correlações extemporâneas à Análise Fatorial<sup>92</sup> – quanto mais próximos estiverem de zero, melhor. A informação mais relevante é a derivada da análise do MSA de cada variável. Da mesma forma que o mencionado acima sobre o valor do MSA geral da Análise Fatorial, o ideal é atingir valores maiores que 0,70. Aquelas variáveis com um índice MSA menor que 0,50 devem ser consideradas pelo pesquisador como passíveis de exclusão<sup>93</sup>. Todavia, como observa Lee (2000, p. 176-177), a exclusão de variáveis deve levar em conta não só a análise do MSA mas também a análise das comunalidades finais (v. item seguinte).

d) “Communalities” (comunalidades): a tabela do SPSS Viewer traz a comunalidade estimada de cada variável com duas configurações possíveis – apenas a comunalidade inicial (coluna “Initial”)<sup>94</sup>, definida como “a força da associação linear

---

<sup>91</sup> O programa SPSS marca o valor com um “a” sobrescrito (“<sup>a</sup>”).

<sup>92</sup> Segundo Lee (2000, p. 173-174), os valores que aparecem ao redor da diagonal (os valores abaixo e acima da diagonal são idênticos) na matriz antiimagem de correlação são os atribuídos a fatores únicos (“unique factors”), ou seja, “o que não é explicado pelos fatores comuns, que são o que se quer extrair” na Análise Fatorial.

<sup>93</sup> Lee (2000, p. 379) propõe a exclusão de variáveis com valor menor que 0,60. Porém, como ocorre em Biber (1988), algumas variáveis têm um MSA mais baixo, entretanto possuem valor teórico relevante e podem ocorrer com frequência apenas em textos de determinado gênero, não no corpus de estudo inteiro. Daí a proposição de um valor-limite de MSA menos rigoroso.

<sup>94</sup> Isso ocorre quando a extração não consegue encontrar uma solução satisfatória para a Análise Fatorial com os parâmetros originalmente fornecidos. Nesse caso, conforme será explicado adiante, será preciso fazer novas extrações com um número definido de fatores a extrair.

entre as variáveis” (NORUŠIS, 1988, p. 130, apud LEE, 2000), ou a informação da comunalidade inicial seguida da comunalidade final da extração, já levando em conta a Análise Fatorial e a rotação (coluna “Extraction”). A comunalidade indica quão bem uma variável pode estar correlacionada com as demais. A comunalidade final, por sua vez, “mede a proporção de variância de uma variável que é explicada pelos fatores comuns retidos (isto é, aqueles extraídos)” (LEE, 2000, p. 177). Uma variável com baixa comunalidade final (menor que 0,10<sup>95</sup>) indica ao pesquisador que é preciso considerar a possibilidade de sua exclusão da Análise Fatorial.

e) “Total Variance Explained” (variância total explicada): a Análise Fatorial, em seu processo de redução de dados, agrupa as variáveis que se correlacionam em fatores, os quais ficam responsáveis por explicar a variância geral dos dados. O primeiro fator explica a maior parcela de variância dos dados, enfeixando as variáveis mais fortes que se correlacionam. O segundo fator, por seu turno, vai tentar explicar a variância que restou e, conseqüentemente, sua solução vai responder por uma proporção menor de variância, agrupando variáveis mais fracas. O processo de criação de novos e mais fracos fatores continua, até que 100% da variância seja explicada pelos fatores. A tabela que acompanha este item, Total de Variância Explicada, traz relacionados, portanto, os fatores que irão explicar a variância dos dados do corpus inicialmente (na coluna “Initial Eigenvalues”) e, quando há a Análise Fatorial consumada, da solução da extração (na coluna “Extraction Sums of Squared Loadings”) e da solução rotacionada (na coluna “Rotation”). O índice que representa a quantidade de variância compartilhada por um dado fator antes da extração é chamado de valor *eigen* (ou “eigenvalue”). Uma das colunas apresentadas soma as variâncias parciais cumulativamente, até chegar a 100% de variância total. A informação da tabela do Total de Variância Explicada mais o gráfico “*scree*” (item seguinte) serão ferramentas importantes para a determinação do número de fatores que a Análise

---

<sup>95</sup> Lee (2000, p. 177) considera valores abaixo de 0,30 como indicativos de baixa comunalidade final. Utilizo um valor mais tolerante, de 0,10, proposto por Biber (comunicação pessoal).

Fatorial final terá.

f) “Scree plot” (gráfico “*scree*”): representa graficamente a coluna dos valores eigen da tabela comentada no item anterior. O gráfico “*scree*” serve como indicativo auxiliar para a determinação do número de fatores para a Análise Fatorial final<sup>96</sup>. Para tanto, deve-se observar a curva produzida pelos valores eigen em cada fator e verificar a partir de quais pontos a curva deixa de ficar pronunciada, na passagem para um padrão plano. O número ideal de fatores deve estar próximo a essa região do gráfico, pois a partir daí os valores eigen dos fatores ficam pouco diferenciados, deixando de haver, portanto, relevância.

Algumas análises fatoriais iniciais produzidas no SPSS terminam de fornecer informações a partir desse ponto, porque não conseguem encontrar uma solução adequada para a extração<sup>97</sup>. Outras, no entanto, vão além, consumando a Análise Fatorial com os fatores que obtiveram valor eigen acima de 1, e fornecendo os seguintes dados: matriz fatorial (“Factor Matrix”), matriz padrão (“Pattern Matrix”), matriz estrutural (“Structure Matrix”) e matriz de correlação entre os fatores (“Factor Correlation Matrix”). Essa não deve ser considerada, no entanto, a solução definitiva para a Análise Fatorial, mas, sim, a primeira de uma série de extrações que têm como meta uma solução estável para a Análise Fatorial.

As matrizes, descritas abaixo, completam o quadro de informações

---

<sup>96</sup> Lee (2000, p. 180-181) recomenda que o gráfico “*scree*” seja gerado com a utilização do método de componentes principais, mas aceita que seu uso combinado com outros métodos de Análise Fatorial pode gerar resultados semelhantes.

<sup>97</sup> É preciso lembrar que não foi determinado, entre os parâmetros do SPSS, o número de fatores para a realização da extração. Esse procedimento, como será visto mais tarde, será realizado após a análise dos dados resultantes da extração inicial. Pode também ocorrer, em algumas extrações com valor definido de fatores, de o programa não encontrar uma solução rotacionada satisfatória. Nesse caso, o Viewer não irá apresentar a Análise Fatorial e nova extração deverá ser feita, com outro número de fatores ou com outra composição de variáveis.

fornecidas pelo SPSS durante uma extração da Análise Fatorial. As matrizes mais importantes para a Análise Fatorial são as matrizes padrão, estrutural e de correlação entre os fatores<sup>98</sup> – obtidas com a rotação oblíqua, que torna a Análise Fatorial mais facilmente interpretável. A matriz fatorial (“Factor Matrix”), por sua vez, não é analisada porque traz a solução inicial da extração, antes da rotação oblíqua.

g) “Pattern Matrix” (matriz padrão): é considerada a matriz mais importante da Análise Fatorial. É a partir dela que são computados os escores dos fatores, que servirão para parametrizar individualmente os textos do corpus. A matriz padrão mostra os pesos (“loadings”) que cada variável possui nos fatores extraídos. A ênfase, porém, é dada aos fatores: a matriz padrão mostra como os fatores contribuem para a variância de uma variável (LEE, 2000, p. 249).

As variáveis vão estar agrupadas pelos maiores pesos (maiores ou iguais a 0,30, positivos ou negativos) nos fatores. Pesos menores que 0,30 não têm importância, apesar de estatisticamente significativos (BIBER, 1988, p. 87).

Com isso, aparecem na matriz apenas os pesos com variância relevante compartilhada, seja no fator no qual recebe o maior peso, seja em outros fatores que recebam pesos secundários, de menor importância. É preciso interpretar os resultados obtidos na matriz padrão para proceder à etapa de rotulagem dos fatores em dimensões, considerando também como os escores dos fatores incidem sobre os textos ou sobre determinados gêneros (v. discussão sobre os escores a seguir).

h) “Structural Matrix” (matriz estrutural): quando, como neste caso, uma extração da Análise Fatorial sofre uma rotação oblíqua (procedimento que facilita a interpretação dos dados, mas aumenta o grau de intercorrelação entre os fatores), a matriz estrutural deve ser interpretada (cf. LEE, 2000). A matriz estrutural reflete uma

---

<sup>98</sup> Biber (1988, 1995) não leva em conta as informações dadas pela matriz estrutural na realização da Análise Fatorial. No entanto, como demonstrado em Lee (2000), a matriz estrutural revela aspectos importantes para a interpretação da Análise Fatorial, como aqueles ligados à intercorrelação entre fatores e entre variáveis.

maior complexidade em relação à matriz padrão porque na primeira é mostrada em que medida uma variável está correlacionada a um ou a mais fatores. A ênfase, diferentemente da matriz padrão, está voltada para o comportamento das variáveis. Podem ocorrer deslocamentos de variáveis de um fator a outro, em maior ou menor grau, na comparação entre as duas matrizes. Isso pode indicar instabilidade na solução, atribuída à presença de alguma variável de baixa comunalidade ou MSA, sugerindo uma nova extração com outra composição de variáveis. Quando as duas matrizes têm fatores compostos por variáveis com perfis semelhantes, é sinal de que a solução é estável.

i) “Factor Correlation Matrix” (matriz de correlação entre os fatores): mostra em que grau existe correlação entre os fatores, ou, em outras palavras, qual é a proporção de variância que é conjuntamente explicada por dois fatores. A título de ilustração, como explicado em Lee (2000, p. 250), um índice de -0,49 (encontrado na solução de Biber, 1988) significa cerca de 25% ( $0,49^2$ ) de variância negativa compartilhada entre fatores, o que indica um valor expressivo. Para Lee (2000), os dados obtidos na matriz de correlação entre fatores devem ser levados em consideração na interpretação e no uso das dimensões resultantes da Análise Fatorial, devido ao fato de que a extração proposta em Biber (1988), com rotação oblíqua, previa a intercorrelação entre dimensões, sem um caráter de independência entre si.

Buscou-se nesta seção apresentar os parâmetros necessários para a realização de uma extração inicial da Análise Fatorial no SPSS, bem como as características principais das informações obtidas em cada extração. O procedimento final da Análise Fatorial será apresentado no Capítulo 4. A seguir serão discutidos os procedimentos adotados para classificar os textos do corpus de estudo em gêneros.

### 3.5 CLASSIFICAÇÃO MANUAL DE GÊNEROS

Com base nas definições dos gêneros jornalísticos descritas nesta seção, o corpus de estudo, composto por 1.431 textos, foi classificado manualmente por este

pesquisador.

Um estudo preliminar (KAUFFMANN, em fase de elaboração) constatou que, em uma amostra estatisticamente representativa do corpus de estudo, 85% dos textos classificados nos gêneros pelo autor coincidem com a opinião majoritária da classificação efetuada por mais três outros especialistas, das áreas de produção editorial e de documentação da informação. Isso apóia a idéia de que o gênero não é definido de maneira homogênea pelos integrantes de uma comunidade discursiva, porém há convergência de opiniões, na maioria dos casos. Vistos em maior detalhe, percebe-se que há gêneros “consensualmente convergentes”, em maior ou menor grau, e gêneros em que se observou divergência de opiniões – por exemplo, a classificação de um texto como reportagem é mais consensual que a classificação de um texto como notícia; os gêneros resenha e comentário são os mais discordantes (KAUFFMANN, em fase de elaboração).

Esta seção dedica-se a buscar um referencial sociocultural para embasar uma classificação de gêneros textuais, de acordo, portanto, com um critério externo ao nível lingüístico. Para assegurar que as definições a respeito de gêneros sejam suficientemente precisas para essa etapa da pesquisa, recorreu-se a trabalhos vinculados aos Estudos de Comunicação e a manuais de redação produzidos por empresas jornalísticas.

### 3.5.1 Tipologia dos Gêneros do Discurso Jornalístico: Uma Proposta

A existência do gênero pressupõe um contexto representado pelo universo sociocultural (SWALES, 1990), cujas fronteiras ultrapassam os limites da linguagem encontrada no texto. Sob esse ponto de vista é possível considerar o gênero um critério externo ao plano léxico-gramatical do estritamente lingüístico (BIBER, 1988, p. 70, 170; EAGLES, 1996a e 1996b; LEE, 2000, p. 51). O texto que pertence a um gênero não tem somente certa estabilidade formal e lexical, mas também um assentado julgamento cognitivo comum sobre a que gênero o texto pertence. Desse modo, as

referências a respeito de gêneros, colhidas para subsidiar o esforço de classificação dos textos do corpus, buscam traduzir o contexto sociocultural de que são expressão; suas fontes provêm da área de Comunicação Social, mais especificamente dos Estudos de Comunicação (NABANTINO RAMOS, 1970; VIVALDI, 1973; RABAÇA; BARBOSA, 1998; MARQUES DE MELO, 1992, 1994, 1998a; NOBLAT, 2002). Também foram consultados para realizar esta tipologia de gêneros os principais manuais de redação utilizados por jornais brasileiros ou editoras – “Folha de S.Paulo”, principal fonte<sup>99</sup>, “O Estado de S.Paulo”, “O Globo”, “Zero Hora” e Editora Abril<sup>100</sup>.

Buscaram-se nessas fontes basicamente citações que pudessem caracterizar gêneros textuais do jornalismo escrito diário, para que pudesse haver um mínimo substrato consensual entre tão movediço assunto. Como advertido por Kennedy (1998, p. 62), “mesmo dentro de um corpus escrito a questão sobre quais gêneros incluir não é clara. Não há, por exemplo, taxonomia abrangente de gêneros da qual se possa selecionar”. Na mesma direção, Vivaldi destaca os liames frágeis entre os gêneros presentes no jornalismo impresso:

“Metodologicamente, admitimos e reconhecemos a dificuldade de explicitar campos, de precisar claramente, de diferenciar um gênero jornalístico de outro. (...) há um cruzamento de traços: artigos que têm muito de crônicas; crônicas que são propriamente artigos, e reportagens especiais que, por seu tom e enfoque, beiram o campo da crônica ou do artigo.” (VIVALDI, 1973, p. 22<sup>101</sup>)

A tipologia proposta a seguir é, portanto, produto da pesquisa às fontes do

---

<sup>99</sup> É o único manual a tratar diretamente do assunto dos gêneros. Há um diálogo com a tipologia proposta por Marques de Melo (1994). Alguns gêneros foram redefinidos de acordo com cada nova edição do “Manual da Redação” (FOLHA DE S.PAULO, 1984, 1987, 1992 e 2001), como será visto nesta seção.

<sup>100</sup> Os dois últimos são manuais de estilo e não fazem referência a gêneros.

<sup>101</sup> No original, “Metodológicamente, admitimos y reconocemos la dificultad de deslindar campos, de precisar netamente, de diferenciar um gênero periodístico de outro. (...) hay un entrecruce de rasgos: artículos que tienen mucho de crónicas; crónicas que son propriamente artículos, y reportajes especiales que, por su tono y enfoque, rozan el campo de la crónica o del artículo.”

jornalismo e as definições para cada gênero têm origem nas principais citações encontradas na pesquisa:

**Artigo:** as menções encontradas têm em comum o fato de haver uma nomeação expressa de autoria e trazer uma marca de opinião: “Gênero jornalístico que traz interpretação ou opinião do autor. Sempre assinado” (FOLHA DE S.PAULO, 1992, p. 123); “Texto jornalístico interpretativo e opinativo, mais ou menos extenso, que desenvolve uma idéia ou comenta um assunto a partir de determinada fundamentação. Geralmente assinado” (RABAÇA; BARBOSA, 1998, p. 51); “Trata-se de uma matéria jornalística onde alguém (jornalista ou não) desenvolve uma idéia e apresenta sua opinião” (MARQUES DE MELO, 1994, p. 116); “Escrito, de conteúdo amplo e variado, de forma diversa, na qual se interpreta, julga ou explica um fato ou uma idéia atuais, de especial transcendência, segundo a conveniência do articulista” (VIVALDI, 1973, p. 176, trad. por MARQUES DE MELO, 1994).

**Carta:** é um gênero pouco lembrado nos compêndios de jornalismo, mas está presente no jornal, como veículo no qual “leitores expressam seus pontos de vista” (FOLHA DE S.PAULO, 1992, p.129). Pode comportar uma instância de interação, na forma de respostas, ou réplicas, dadas por jornalistas, individualmente, ou pela redação, institucionalmente.

**Chamada:** “Texto curto na Primeira Página, resume as informações publicadas pelo jornal a respeito de um assunto. Remete o leitor para as páginas que trazem a cobertura extensiva” (FOLHA DE S.PAULO, 2001, p. 58). Chaparro (1997) chama o gênero de “Resumo-Chamada”, numa definição que tenta dar conta da característica morfológica do texto encontrado na primeira página, um extrato de outros textos, localizados no interior do jornal: “Têm textos mais elaborados e que, sob o ponto de vista formal, apresentam estrutura de Notícia – um resumo descritivo-informativo com a essência de conteúdos mais desenvolvidos e densos, de grande

relevância, normalmente narrados em forma de Reportagem, localizados em páginas internas e para os quais esses resumos chamam a atenção” (CHAPARRO, 1997, p. 53).

**Coluna de notas:** configura-se como um “texto-colônia” (HOEY, 2001, p. 72-92), ou uma coleção de pequenos textos que podem ser lidos em qualquer ordem pelo leitor, sem prejuízo do seu sentido. Pode ser também chamada de “coluna”. Para Rabaça e Barbosa (1998), significa “Seção especializada de jornal ou revista, publicada com regularidade, geralmente assinada, e redigida em estilo mais livre e pessoal do que o noticiário comum. Compõe-se de notas, sueltos, crônicas, artigos ou textos-legendas, podendo adotar, lado a lado, várias dessas formas. As colunas mantêm um título ou cabeçalho constante e são diagramadas geralmente em uma posição fixa e sempre na mesma página” (RABAÇA; BARBOSA, 1998, p. 143).

**Comentário:** as definições a respeito de comentário são vagas: “Gênero jornalístico opinativo (sem o rigor de análise que caracteriza a crítica), sobre qualquer fato, evento ou assunto” (RABAÇA; BARBOSA, 1998, p. 144); “Pequeno artigo interpretativo” (FOLHA DE S.PAULO, 1992, p. 61); “Mantém vinculação estreita com a atualidade, sendo produzido em cima dos fatos que estão ocorrendo. Vem junto com a própria notícia” (MARQUES DE MELO, 1994, p. 109). No entanto, o gênero foi mantido na lista de gêneros para averiguar até que ponto haveria escolhas a seu favor<sup>102</sup>.

**Crítica:** “Discussão fundamentada e sistemática, a respeito de determinada manifestação artística” (RABAÇA; BARBOSA, 1998, p. 186); “Avalia trabalho

---

<sup>102</sup> O mesmo não ocorreu com o termo “análise”, mais indefinido que o comentário. Exemplo: “[...] Situa-se num campo intermediário entre a crônica, a crítica ou o comentário (mais subjetivos) e a notícia propriamente dita (mais objetiva)” (FOLHA DE S.PAULO, 2001).

artístico, acadêmico ou desempenho esportivo” (FOLHA DE S.PAULO, 2001, p. 71).

**Crônica:** é “um texto jornalístico desenvolvido de forma livre e pessoal, a partir de fatos e acontecimentos da atualidade, com teor literário, político, esportivo, artístico, etc.” (RABAÇA; BARBOSA, 1998, p. 187). Chaparro (1997) considera a crônica um gênero próprio da imprensa brasileira, com características peculiares. De uma perspectiva mais ampla, “designa composição breve, relacionada com a atualidade.” (RÓNAI, 1971, apud MARQUES DE MELO, 1994, p. 146)

**Editorial:** um exemplo de citação, coincidente com as demais encontradas, é “texto jornalístico opinativo, escrito de maneira impessoal e publicado sem assinatura [...]. Define e expressa o ponto de vista do veículo ou da empresa responsável pela publicação (jornal, revista, etc.)” (RABAÇA; BARBOSA, 1998, p. 227).

**Entrevista:** texto redigido “sob a forma de perguntas e respostas. Reproduz o diálogo mantido entre o repórter e o entrevistado” (RABAÇA; BARBOSA, 1998, p. 238). Para os mesmos autores, também significa o trabalho realizado na apuração jornalística pelo repórter.

**Nota de correção:** a “Folha de S.Paulo” adota a publicação sistemática de notas de correção para retificar as informações anteriormente veiculadas. Daí a proposição de integrar o gênero nesta tipologia, apesar de não ter encontrado na literatura citações a respeito.

**Notícia:** as diversas definições coletadas sobre a notícia são semelhantes: “Relata a informação de maneira mais objetiva possível” (FOLHA DE S.PAULO, 2001); “... o relato de uma série de fatos a partir do fato mais importante, e este, de seu aspecto mais importante” (LAGE, 1998, p. 25-26). Há um sentido complementar, de

extensão<sup>103</sup>: “é a informação concisa de fato jornalístico” (NABANTINO RAMOS, 1970, p. 171), “... um relato mais ou menos breve sobre um fato” (NOBLAT, 2002, p. 94). Chaparro (1998, p. 125), em uma definição mais completa para o gênero, define a notícia como “o resumo informativo para a descrição jornalística de um fato relevante que se esgota em si mesmo, e para cuja compreensão bastam as informações que o próprio fato contém”.

**Outros:** este campo foi proposto para que pudesse conter tudo o que não estivesse adequado a ser classificado nos demais gêneros e também incluir gêneros pouco expressivos em termos de representatividade, como, por exemplo, efemérides, resultado de loterias e análise.

**Reportagem:** apesar de haver um grande número de definições convergentes sobre a reportagem, existe certa tensão terminológica: Rabaça e Barbosa (1998, p. 509) a consideram designador de um processo, não de um produto: “Considera-se incorreto designar reportagem como um tipo de notícia descritiva, mais apurada e ampla, acompanhada com documentação e testemunhos. Na verdade esse tipo de notícia é resultado de uma reportagem, e não a reportagem em si”. Essa opinião não é compartilhada pelas demais referências encontradas, como a que define reportagem como “... um relato mais extenso, abrangente e contextualizado” (NOBLAT, 2002, p. 94), que “traz informações mais detalhadas sobre notícias, interpretando os fatos” (FOLHA DE S.PAULO, 2001, p.71). Para Chaparro (1998, p. 125), na definição que parece ser a mais completa, reportagem é “o relato jornalístico que expande a Notícia, para desvendamentos ou explicações que tornam mais ampla a atribuição de significados a acontecimentos ocorridos ou em processo de ocorrência”. Uma visão

---

<sup>103</sup> Nesse aspecto, aproximando-se de algumas definições esparsas encontradas sobre “nota”. O termo não foi incluído na tipologia por não ser consensual entre os autores compulsados e por possuir um grau de sobreposição em relação a “notícia” e “coluna de notas”.

ditada pela prática jornalística define o gênero como “o relato do acontecimento importante, feito pelo jornalista que tenha estado no local em que o fato ocorreu ou tenha apurado as informações relativas a ele. É o produto essencial da atividade jornalística” (FOLHA DE S.PAULO, 1987, p. 158; cf. FOLHA DE S.PAULO, 1992, p. 42).

**Resenha:** difere do gênero crítica apenas pelo objeto analisado pelo texto, os livros: “É uma espécie do gênero crítica. Resume e critica um ou vários livros sobre o mesmo assunto” (FOLHA DE S.PAULO, 1984). Para Marques de Melo (1994, p. 125), é o mesmo que crítica.

A tipologia de gêneros utilizada para realizar a classificação dos textos do corpus, apresentada acima, pretendeu ser abrangente o suficiente para compreender a grande maioria dos casos observados no corpus de estudo, mas é preciso ressaltar, como dito anteriormente, que não existe consenso a respeito dessa nomenclatura. Todavia, a pesquisa aqui relatada abre a possibilidade indireta de testar essa lista, indicando a existência de eventuais desvios ou distorções, seja pelo peso dado às menções de gênero presentes na lista, seja pela utilização da opção “outros”.

Como dito anteriormente, as citações reunidas acima servem de substrato teórico para uma classificação manual de gêneros do corpus de estudo. O procedimento irá dividi-lo em subcorpora de gêneros (cf. 4.6), base de estudo para a fase final da pesquisa.

No capítulo seguinte, por meio da apresentação e discussão dos resultados da pesquisa serão detalhados os aspectos que descreverão a mecânica para a obtenção de uma solução satisfatória da Análise Fatorial, prática que envolve várias extrações. Será explanada também a etapa de interpretação dos fatores, processo de rotulagem que os transforma em dimensões.

## 4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Serão apresentados a seguir os resultados da Análise Multidimensional (BIBER, 1988; LEE, 2000) de um corpus que procura representar com rigor uma amostra da imprensa escrita diária em português do Brasil nos dias de hoje. Com o intuito de aplicar os parâmetros técnicos e estatísticos examinados até aqui, será também dado destaque à descrição das fases finais do processo da Análise Fatorial, procedimento chave para a realização da Análise Multidimensional.

Conforme visto no capítulo anterior, a montagem do corpus obedeceu a determinados princípios de representatividade e extensão (cf. TOGNINI-BONELLI, 2001, p. 57-63; HUNSTON, 2002, p. 28-30; BERBER SARDINHA, 2004a, p. 22-27) na sua composição. Com base em uma “semana construída” (KENNEDY, 1998, p. 75), foram compiladas sete edições diárias de 2003, aleatoriamente escolhidas, de um jornal de circulação nacional (“Folha de S.Paulo”), totalizando um corpus de estudo composto por 1.431 textos completos, o correspondente a 493.780 palavras.

O corpus de estudo foi classificado morfossintaticamente, por um processo automático de etiquetagem do português via Internet, o VISL (“Visual Interactive Syntax Learning”; cf. BERBER SARDINHA, 2004a, p. 136-142), que possui um índice de confiabilidade de 99% em termos de classe de palavras e flexão e de 97% a 98% para sintaxe (BICK, 2005).

Os resultados, apresentados e discutidos ao longo deste capítulo, buscam responder às questões de pesquisa, explicitadas na Introdução. Os fatores resultantes da Análise Fatorial, bem como as variáveis que os compõem, estão mostrados nas seções 4.1 a 4.5. Os resultados da análise concernente aos gêneros encontram-se nas seções 4.6 e 4.7. Em 4.8 e 4.9, antes da discussão final, será examinado o processo de interpretação dos fatores em termos de dimensões, que representam funções subjacentes, de caráter comunicativo (BIBER, 1988).

#### 4.1 SELEÇÃO DAS VARIÁVEIS

O ponto de partida para a escolha das variáveis que compõem a Análise Multidimensional foi o jogo de etiquetas gerado pelo etiquetador morfossintático VISL, para a língua portuguesa (BICK, 2005; cf. Anexo A). Em adição ao disponibilizado pelo VISL, foi incluída a variável quantidade de palavras, que mede o tamanho de cada texto pelo número total de palavras. Essa informação foi extraída com a ferramenta Wordlist, da suíte WordSmith Tools, versão 3.0 (número de “tokens”).

A lista das variáveis inicialmente pesquisadas encontra-se no Anexo A. Essa lista inclui as variáveis que foram pesquisadas basicamente por dois motivos: a) são variáveis de alta frequência no corpus de estudo etiquetado; b) são variáveis derivadas de um estudo a respeito das palavras-chave ocorrentes em grupos de textos do corpus de estudo, classificados de acordo com o gênero (KAUFFMANN, 2005).

Com o auxílio de expressões regulares utilizadas em sistemas Unix<sup>104</sup>, as etiquetas foram contadas e os dados resultantes foram arquivados em uma planilha Excel. Algumas etiquetas originais foram somadas, como os lemas que integram a variável verbos públicos.

Com base nos critérios mencionados em 3.2.2, foram selecionadas desse conjunto maior 19 características e categorias morfossintáticas e lexicais teoricamente relevantes para compor o rol de variáveis efetivamente utilizadas na Análise Multidimensional. Elas estão descritas em maior detalhe a seguir, na Tabela 1, acompanhadas de funções detectadas e referências de outros estudos:

---

<sup>104</sup> Foi utilizado para a contagem um emulador Unix para Windows (Cygwin) e, principalmente, a expressão regular grep (cf. BERBER SARDINHA, 2004a, p. 54, 65-70).

FIGURA 2 - INFORMAÇÕES SOBRE AS VARIÁVEIS UTILIZADAS NA EXTRAÇÃO INICIAL DA ANÁLISE FATORIAL

continua

Variável	Etiqueta(s) VISL	Exemplos extraídos do corpus de estudo etiquetado <sup>105</sup>	Funções	Fonte das informações
<b>Adjetivos</b>	ADJ (Adjetivo)	aéreas [aéreo] <b>ADJ</b> F P @N<	Expandir e elaborar a informação presente no texto (Pacheco de Oliveira, 1997, p. 93)	Chafe, 1982; Biber, 1988, p. 237
<b>Advérbios</b>	ADV (Adjunto Adverbial)	geralmente [geralmente] <b>ADV</b> @ <b>ADV</b> >	Expressar uma reação emocional, julgamento ou avaliação do escritor (Pacheco de Oliveira, 1997, p. 93)	Leech e Svartvik, 1980; Pacheco de Oliveira, 1997, p. 93
<b>Aposições</b>	@APP (Aposição do substantivo) [epíteto de identidade]	Walter=Neves, antropólogo “antropólogo” N M S @ <b>APP</b>	Itens lexicais adjacentes com mesmo referente	Quirk et al., 1985, p. 1260, 1300-1301
<b>Aspas</b>	<*1> (Aspa esquerda)	qualificou como decepcionante [decepcionante] <*1> <*2> <b>ADJ</b> M/F S @AS<	Introduz a fala de outros atores no texto; marca de citação	Corpus de estudo
<b>Conjunções coordenativas</b>	KC (Conjunção coordenativa)	mediante ameaça ou [ou] <co-prparg> <b>KC</b> @CO violência	Indicar conexões entre períodos; criar fragmentação (Pacheco de Oliveira, 1997, p. 95)	Lux e Grabe, 1991, p. 146; Biber, 1992, p. 140
<b>Conjunções subordinativas</b>	KS (Conjunção subordinativa)	há quem afirme que [que] <b>KS</b> @SUB @#FS-<ACC ele	Indicar conexões entre períodos; indicar complexidade estrutural (Pacheco de Oliveira, 1997, p. 95)	Grabe, 1987, p. 118; Biber, 1992, p. 140
<b>Futuro do pretérito</b>	COND (Condicional)	não deveria [dever] V <b>COND</b> 3S VFIN @FAUX ser	Exprime referência passada; incerteza, condicionalidade, polidez	Cunha, 1980, p. 441

<sup>105</sup> A relação das etiquetas do VISL está no Anexo A. Foram incluídos fragmentos de palavras que aparecem antes e/ou depois da palavra etiquetada.

FIGURA 2 - INFORMAÇÕES SOBRE AS VARIÁVEIS UTILIZADAS NA EXTRAÇÃO INICIAL DA ANÁLISE FATORIAL

continuação

Variável	Etiqueta(s) VISL	Exemplos extraídos do corpus de estudo etiquetado	Funções	Fonte das informações
<b>Nomes próprios</b>	PROP (Nome próprio)	SP [SP] <b>PROP</b> M S @SUBJ> lidera ranking	Indicar referências extratexto ao contexto situacional (Pacheco de Oliveira, 1997, p. 92)	Pacheco de Oliveira, 1997, p. 92
<b>Números cardinais</b>	NUM (Numeral cardinal)	há 70 [70] <card> <b>NUM</b> M P @>N magistrados	Referência quantitativa	Corpus de estudo
<b>Presente do indicativo</b>	PR IND (Presente do indicativo)	a mudança faz [fazer] <fmc> V <b>PR</b> 3S <b>IND</b> VFIN @FMV parte das discussões	Marcador de situação imediata; foco na informação apresentada	Biber, 1986; Biber, 1988, p. 224
<b>Pretérito imperfeito</b>	IMPF IND (Pretérito imperfeito do indicativo)	três titulares jogavam [jogar] <fmc> V <b>IMPF</b> 3P <b>IND</b> VFIN @FMV na Europa	Utilizado em descrições e narrações de acontecimentos passados	Cunha, 1980, p. 432
<b>Pretérito perfeito</b>	PS, PF / MQF (Pretérito perfeito do indicativo e pretérito perfeito / Mais que perfeito do indicativo <sup>106</sup> )	ele se chocou [chocar] <fmc> V <b>PS</b> 3S <b>IND</b> VFIN @FMV com um relógio de rua	Marcador narrativo; foco na seqüência temporal	Biber, 1986; Biber, 1988, p. 223-224
<b>Pronomes demonstrativos</b>	<dem> (Pronome demonstrativo)	a taxa esperada para este [este] <dem> DET M S @>N mês.	Referência exofórica ou a um referente anteriormente citado no texto; propriedade dêitica	Biber, 1988, p. 226
<b>Quantidade de palavras</b>	Não há	Não há	Avaliar o comprimento do texto; especificidade lexical	Biber, 1988, p. 238; Reid, 1990, p. 195; Pacheco de Oliveira, 1997, p. 98

<sup>106</sup> As duas etiquetas foram somadas porque se verificou que a ocorrência da etiqueta PF/MQF era composta praticamente na totalidade de formas verbais flexionadas de 3ª pessoa do plural do pretérito perfeito do modo indicativo (Exemplos: “fizeram”, “foram”, “disseram”).

FIGURA 2 - INFORMAÇÕES SOBRE AS VARIÁVEIS UTILIZADAS NA EXTRAÇÃO INICIAL DA ANÁLISE FATORIAL

				conclusão
Variável	Etiqueta(s) VISL	Exemplos extraídos do corpus de estudo etiquetado	Funções	Fonte das informações
<b>Substantivos</b>	N	afeta aproximadamente 13% da população [população] N F S @P< mundial.	Índice nominal de um texto; foco informacional abstrato	Biber, 1988, p. 227-228
<b>Verbos públicos</b>	[afirmar], [dizer], [falar], [declarar], [comentar], [perguntar], [responder] (Lemas dos verbos acima)	Piva afirmou <b>[afirmar]</b> <fmc> V PS 3S IND VFIN @FMV ontem durante o seminário	Introduzem o discurso (direto ou indireto)	Quirk et al., 1985, p. 1180-1181; Neves, 2000, p. 47-49
<b>Verbos / pronomes 1ª pessoa singular</b>	1S (1ª pessoa do singular - verbos e pronomes)	agora eu [eu] PERS M/F <b>1S</b> NOM @SUBJ> nem sei mais o gosto que tem	Marcam o envolvimento do escritor e dos atores do texto	Chafe, 1982; Biber, 1988, p. 225; Biber, 1995, p. 373.
<b>Verbos / pronomes 3ª pessoa singular</b>	3S (3ª pessoa do singular - verbos e pronomes)	ele “ele” PERS M <b>3S</b> NOM @SUBJ> encontra Margarida	Marcam referência a pessoas ou coisas fora da interação imediata	Biber, 1986; Biber, 1988, p. 225
<b>Voz passiva</b>	PASS (Agente da passiva)	uma delas foi [ser] <fmc> V PS 3S IND VFIN @FAUX dirigida [dirigir] V PCP F S @IMV @#ICL-AUX< por [por] PRP @< <b>PASS</b> Bernardinho	Apresentação mais abstrata da informação	Biber, 1988, p. 228

Foi feita a seguir uma normalização da contagem (base = 1.000 palavras<sup>107</sup>), cujo resultado, juntamente com a estatística descritiva de cada variável, está mostrado na tabela abaixo:

TABELA 4 - ESTATÍSTICA DESCRITIVA DAS VARIÁVEIS NO CORPUS DE ESTUDO (POR 1.000 PALAVRAS; N = 1.431\*)

<b>Característica / categoria</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Alcance</b>	<b>Média</b>	<b>Desvio padrão**</b>	<b>Assimetria**</b>	<b>Curtose**</b>
<b>Adjetivos</b>	0	166,67	166,67	56,18	23,06	0,58	1,28
<b>Advérbios</b>	0	384,62	384,62	125,67	37,57	0,32	2,21
<b>Aposições</b>	0	81,97	81,97	5,53	7,35	3,28	20,23
<b>Aspas</b>	0	120	120	9,70	12,67	3,39	19,39
<b>Conjunções coordenativas</b>	0	102,04	102,04	27,78	15,02	0,72	1,37
<b>Conjunções subordinativas</b>	0	65,57	65,57	11,88	10,45	0,97	1,11
<b>Futuro do pretérito</b>	0	40,54	40,54	2,79	4,83	2,82	11,10
<b>Nomes próprios</b>	0	363,64	363,64	78,36	37,12	1,06	3,35
<b>Números cardinais</b>	0	540,98	540,98	27,99	26,51	6,34	101,81
<b>Presente do indicativo</b>	0	169,23	169,23	49,58	25,11	0,52	0,53
<b>Pretérito imperfeito</b>	0	93,02	93,02	5,98	8,95	2,97	14,82
<b>Pretérito perfeito</b>	0	114,29	114,29	21,39	17,90	1,21	2,08
<b>Pronomes demonstrativos</b>	0	45,45	45,45	6,55	6,77	1,41	2,91
<b>Quantidade de palavras</b>	22	5.009	4.987	345,06	286,67	4,27	53,44
<b>Substantivos</b>	65,57	470	404,43	230,58	39,60	0,25	2,66
<b>Verbos públicos</b>	0	53,33	53,33	3,95	5,97	2,76	11,87
<b>Verbos / pronomes 1ª pessoa singular</b>	0	107,14	107,14	4,16	10,18	4,20	23,63
<b>Verbos / pronomes 3ª pessoa singular</b>	0	261,54	261,54	72,27	26,59	0,57	2,59
<b>Voz passiva</b>	0	83,33	83,33	3,89	5,44	3,86	35,96

\* N significa “número de textos”. \*\* Cf. definições em 3.4.1.

<sup>107</sup> Exceto a variável quantidade de palavras.

Os dados obtidos na Tabela 4 revelam que as variáveis possuem perfis diversos de distribuição, como diferentes frequências médias, desvios padrão, assimetria e curtose. A grande maioria, no entanto, possui um traço em comum: tem frequência mínima zero – isto é, em pelo menos um texto do corpus de estudo não houve nenhuma ocorrência da variável.

Há algumas variáveis com frequência relativamente rara (entre outras, futuro do pretérito, voz passiva e verbos públicos), a maioria coincidentemente com altos índices de assimetria e curtose. Tais variáveis devem ter, portanto, seu comportamento observado na etapa seguinte, em que é realizada a extração inicial da Análise Fatorial.

Observa-se que algumas características são muito seletivas e só ocorrem em quantidade em poucos textos, causando uma grande curtose (numerais cardinais e número de palavras têm os maiores índices). Isso sugere que certas características lingüísticas fogem bastante de um padrão de distribuição normal, fazendo com que um estudo quantitativo da língua de natureza estatística como o da Análise Multidimensional tenha de admitir essas grandes variações para buscar ser suficientemente abrangente em termos lingüísticos (cf. LEE, 2000, p. 237-238, 394-395).

Uma extração inicial da Análise Fatorial foi feita levando-se em conta as 19 variáveis levantadas. O resultado da Análise Fatorial mostrou um resultado global de MSA de 0,576, e o teste de Bartlett efetuado eliminou a chance de a matriz ser considerada inválida.

Na extração seguinte foram consideradas apenas variáveis com comunalidade mínima de 0,10<sup>108</sup>. Ou seja, as variáveis aposição, aspas, conjunções coordenativas, futuro do pretérito e voz passiva foram excluídas da análise por não estarem aptas a integrar o procedimento estatístico da Análise Fatorial, que se baseia justamente na *co-ocorrência* entre variáveis, em que é condição *sine qua non* possuir

---

<sup>108</sup> Conforme recomendado por Biber (comunicação pessoal).

nível suficiente de comunalidade. Isso significa que as variáveis excluídas, apesar de apresentarem uma frequência positiva no corpus jornalístico, não estiveram associadas na extração a nenhuma outra variável com semelhante perfil de distribuição.

A extração seguinte, com as comunalidades das variáveis modificadas com a exclusão das cinco mais fracas, serve como marco inicial da Análise Fatorial. Foi obtido o resultado global de medida de adequação de amostragem (MSA) de 0,515. Observam-se na Tabela 5 as comunalidades  *finais*  das 14 variáveis que doravante serão utilizadas nas análises fatoriais subseqüentes.

TABELA 5 - COMUNALIDADES FINAIS DA EXTRAÇÃO INICIAL, COM 14 VARIÁVEIS  
(MÉTODO DE EXTRAÇÃO: PRINCIPAL AXIS FACTORING)

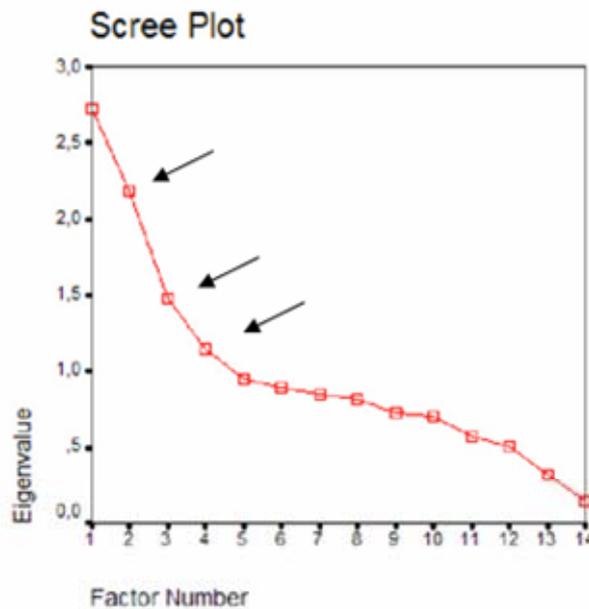
<b>Variável</b>	<b>Comunalidade final</b>
<b>Adjetivos</b>	0,144
<b>Advérbios</b>	0,146
<b>Conjunções subordinativas</b>	0,424
<b>Nomes próprios</b>	0,335
<b>Números cardinais</b>	0,164
<b>Presente do indicativo</b>	0,300
<b>Pretérito imperfeito</b>	0,141
<b>Pretérito perfeito</b>	0,659
<b>Pronomes demonstrativos</b>	0,234
<b>Quantidade de palavras</b>	0,116
<b>Substantivos</b>	0,136
<b>Verbos públicos</b>	0,212
<b>Verbos / pronomes 1ª pessoa singular</b>	0,135
<b>Verbos / pronomes 3ª pessoa singular</b>	0,420

Em esclarecimento à utilização do termo comunalidade “final”, deve-se dizer que toda extração de Análise Fatorial no SPSS traz uma tabela de comunalidades, que se repete se não houver alteração na composição das variáveis. Ela é “final” somente para os fins da pesquisa em curso.

## 4.2 DEFINIÇÃO DO NÚMERO DE FATORES

O gráfico “*scree*” da extração de Análise Fatorial com 14 variáveis, mostrado a seguir, representa o montante de variação atribuído a cada fator. Ele servirá de auxiliar para a decisão sobre o número final de fatores a extrair da Análise Fatorial:

GRÁFICO 1 - GRÁFICO “SCREE” DA EXTRAÇÃO COM 14 FATORES (MÉTODO DE EXTRAÇÃO: PRINCIPAL AXIS FACTORING)



Pela interpretação do gráfico “*scree*”, é possível formular três opções possíveis para o número de fatores a ser definido na Análise Fatorial final: 2, 3 ou 4 fatores (marcados por mim com as setas, sobre o gráfico original do SPSS). A partir dessa análise, foram feitas três extrações testando essas hipóteses e aferindo a sua interpretabilidade, de modo a colher subsídios para a determinação da solução mais adequada para a Análise Fatorial.

A decisão sobre a escolha do número de fatores que compõem a Análise Fatorial, para Lee (2000, p. 241), é o passo que mais influi nos resultados da extração.

Há muito de subjetivo na escolha, que depende de dois parâmetros principais<sup>109</sup>: a leitura do gráfico “*scree*” e a *interpretabilidade* dos fatores resultantes<sup>110</sup> – isto é, em que medida os fatores resultantes da solução podem ser explicados com apoio da teoria existente.

Biber acrescenta um outro componente para julgar a adequação do número de fatores, a quantidade de variáveis por fator: “Em geral, cinco pesos salientes são necessários para uma interpretação significativa do constructo subjacente a um fator” (BIBER, 1988, p. 88<sup>111</sup>). Apesar disso, na solução de AMD de Biber (1988), com sete fatores, o sétimo deles possuía um único peso saliente (posteriormente, esse fator foi eliminado da análise). Sua justificativa, no entanto, foi a cautela em não contabilizar um número excessivamente reduzido de fatores, que poderiam ser o produto da soma de outros fatores potencialmente interessantes. Apóia essa postura a afirmação de que é preferível extrair mais fatores que um número muito pequeno deles, optando-se mais tarde por eliminar os fatores mais fracos (cf. BIBER, 1988, p. 88).

#### 4.3 MATRIZES PADRÃO E ESTRUTURAL DA ANÁLISE FATORIAL

Na presente pesquisa, em que os dados do corpus não possuem a mesma intensidade de variação que o corpus estudado por Biber (1988), dada a própria natureza da origem dos dados – todos de um mesmo modo, o escrito, e provenientes de uma fonte comum, o jornal –, é suposto que a solução de Análise Fatorial definitiva tenha um número menor de fatores que os extraídos em Biber (1988).

---

<sup>109</sup> Lee (2000, p. 182) sugere também a análise paralela, descrita em Sharma (1996, p. 77-79), como meio alternativo de criar um critério mais objetivo para a tomada de decisão, mas ela não será analisada aqui.

<sup>110</sup> Para Biber (comunicação pessoal), a interpretabilidade dos fatores, efetuada pelo pesquisador, é um critério fundamental para essa escolha.

<sup>111</sup> No original, “In general, five salient loadings are required for a meaningful interpretation of the construct underlying a factor”.

Foram tentadas três alternativas de extração a partir da leitura do gráfico “*scree*” – determinando-se previamente por soluções com dois, três ou quatro fatores na Análise Fatorial. Optou-se, por mostrar-se mais *interpretável* (cf. 4.1), pela solução com dois fatores<sup>112</sup>, cujas matrizes padrão e estrutural são mostradas a seguir.

TABELA 6 - MATRIZ PADRÃO DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES (PESOS ACIMA DE 0,30)

Variável	Fator 1	Fator 2
<b>Pretérito perfeito</b>	0,73	-0,38
<b>Verbos / pronomes 3ª pessoa singular</b>	0,64	
<b>Conjunções subordinativas</b>	0,51	0,39
<b>Verbos públicos (falar, afirmar, dizer, etc.)</b>	0,44	
<b>Pretérito imperfeito</b>	0,37	
<b>Advérbios</b>	0,34	
<b>Verbos / pronomes 1ª pessoa singular</b>	0,33	
<b>Substantivos</b>	-0,35	
<b>Adjetivos*</b>		
<b>Presente do indicativo</b>		0,55
<b>Pronomes demonstrativos</b>		0,48
<b>Quantidade de palavras</b>		0,34
<b>Nomes próprios</b>		-0,58
<b>Números cardinais</b>		-0,34

\* A variável obteve peso menor que 0,30.

A matriz padrão (Tabela 6) apresenta as variáveis com pesos salientes (acima de 0,30 absoluto; ou seja, com valor positivo ou negativo) agrupadas nos dois fatores. Os campos em branco correspondem a pesos menores que 0,30, considerados pouco importantes na análise (BIBER, 1988, p. 87). Com isso, a variável adjetivos fica excluída da Análise Fatorial (os mesmos pesos pouco expressivos repetem-se na matriz estrutural, como será visto na Tabela 7).

---

<sup>112</sup> Decisão que deve ser inteiramente responsabilizada a este pesquisador, mas que foi norteadada pela crucial colaboração do próprio Prof. Douglas Biber, que gentilmente examinou os dados desta pesquisa nas extrações iniciais da Análise Fatorial.

Duas variáveis, pretérito perfeito e conjunções subordinativas, tiveram um segundo peso saliente em mais de um fator. Seguindo o procedimento utilizado por Biber (1988), apenas os pesos mais salientes em valores absolutos (nesse caso, ambos concentrados no Fator 1) prevalecem para o cálculo dos escores, em detrimento dos que obtiveram peso menor, no segundo fator.

Oito variáveis foram agrupadas no Fator 1 (pretérito perfeito, verbos/pronomes da 3ª pessoa singular, conjunções subordinativas, verbos públicos, pretérito imperfeito, advérbios, verbos/pronomes da 1ª pessoa singular e substantivos) e cinco no Fator 2 (presente do indicativo, pronomes demonstrativos, quantidade de palavras, nomes próprios e números cardinais).

Os valores negativos apresentados têm tanta importância quanto os positivos. Eles representam uma correlação negativa entre os pesos presentes nos fatores. Por exemplo, um texto que tipicamente represente o pólo positivo do Fator 2 teria relativamente muitos verbos no presente do indicativo, pronomes demonstrativos e grande quantidade de palavras; *ao mesmo tempo*, seriam observados poucos nomes próprios e numerais cardinais. O inverso é observado nos textos que têm escores no pólo negativo da dimensão: muitos nomes próprios e números, associado a poucos verbos no presente e pronomes demonstrativos, em textos relativamente pequenos em extensão.

A matriz estrutural da solução de dois fatores (Tabela 7), com 14 variáveis, apresenta poucas diferenças em relação à matriz padrão. As duas variáveis que apresentaram pesos salientes secundários na matriz padrão, pretérito perfeito e conjunções subordinativas, repetiram esse comportamento na matriz estrutural.

A semelhança das duas matrizes apresentadas indica que a solução da Análise Fatorial é estável, proporcionando maior segurança ao pesquisador quanto à definição dos fatores.

TABELA 7 - MATRIZ ESTRUTURAL DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES (PESOS ACIMA DE 0,30)

Variável	Fator 1	Fator 2
<b>Pretérito perfeito</b>	0,72	-0,36
<b>Verbos / pronomes 3ª pessoa singular</b>	0,64	
<b>Conjunções subordinativas</b>	0,52	0,40
<b>Verbos públicos (falar, afirmar, dizer, etc.)</b>	0,44	
<b>Pretérito imperfeito</b>	0,37	
<b>Advérbios</b>	0,35	
<b>Verbos / pronomes 1ª pessoa singular</b>	0,33	
<b>Substantivos</b>	-0,35	
<b>Adjetivos*</b>		
<b>Presente do indicativo</b>		0,55
<b>Pronomes demonstrativos</b>		0,48
<b>Quantidade de palavras</b>		0,34
<b>Nomes próprios</b>		-0,58
<b>Números cardinais</b>		-0,35

\* A variável obteve peso menor que 0,30.

Antes de passar à fase de interpretação dos fatores e ao processo que envolve a sua rotulagem, para convertê-los em dimensões, é pertinente mostrar a matriz de correlação entre os fatores obtidos, como segue:

TABELA 8 - MATRIZ DE CORRELAÇÃO ENTRE OS FATORES DA EXTRAÇÃO COM 14 VARIÁVEIS, 2 FATORES

Fator	1	2
1	1,00	0,02
2	0,02	1,00
Método de extração: Principal Axis Factoring (Promax)		

A matriz que mostra a correlação existente entre os fatores indica que há um valor insignificante obtido entre os dois fatores (0,02), sugerindo um alto grau de independência nos perfis de distribuição – ou seja, mesmo estando sempre presentes em todos os textos do corpus, não há evidência, nos dados observados no corpus, de co-ocorrência entre os dois fatores.

Se considerarmos a variância total explicada pelos dois fatores, o resultado

obtido corresponde a 25,5% de toda a variação lingüística encontrada no corpus. Esse resultado é a soma da variância explicada pelo Fator 1 (14,6%) e pelo Fator 2 (10,8%).

#### 4.4 CÁLCULO DE ESCORES DE TEXTOS POR FATOR

O resultado da Análise Fatorial, mostrado na matriz padrão apresentada na seção anterior, dá condição para a realização de cálculos envolvendo a atribuição de escores das variáveis sobre os textos do corpus de estudo. Tais escores de variáveis nos textos permitem outro cálculo, o do escore do fator por texto. Desse modo, é possível ter acesso ao perfil estatístico de distribuição de cada texto do corpus de estudo em relação aos parâmetros abordados na pesquisa: qual é o peso de cada variável no texto, qual é o posicionamento do texto em relação a cada fator e qual é a relação que mantém com outros textos que fazem parte de um mesmo subgrupo de gênero.

Biber (1988, p. 93-97) propõe um sistema de atribuição de escores aos fatores por meio do cálculo de escalas somadas (LEE, 2000, p. 205), que seriam produto da contagem padronizada da frequência das variáveis salientes do fator, com base nos seus desvios padrões e médias<sup>113</sup>. Matematicamente, os escores que representarão cada texto individual nos fatores obtidos da solução da Análise Fatorial são o resultado do seguinte cálculo, a partir da frequência normalizada por 1.000 palavras:

$$\begin{array}{l} \text{Escore} \\ \text{padronizado} \\ \text{do fator} \end{array} = \frac{\text{(frequência da variável no texto - média da variável no corpus)}}{\text{desvio padrão da variável}}$$

Para calcular os escores padronizados por variável no ambiente do SPSS,

---

<sup>113</sup> Os desvios padrões e médias de cada variável normalizada estão na Tabela 4.

deve-se utilizar a janela “Compute variable”, por meio do menu “Transform” e, a seguir, “Compute”. Na janela “Numeric expression”, reproduz-se a fórmula acima apresentada, em que “a frequência da variável no texto” é representada pelo nome da variável e a média e o desvio padrão, pelos valores normalizados encontrados na estatística descritiva. O resultado do cálculo virá na forma de uma nova variável agregada ao arquivo de dados utilizado.

Uma vez obtido o escore individual de cada texto, variável a variável, o passo seguinte é calcular o escore do texto por fator, que leva em conta os escores padronizados das variáveis salientes de cada fator. Para isso, basta realizar a somatória dos escores padronizados das variáveis com valor positivo e diminuir esse valor da somatória dos escores padronizados das variáveis com valor negativo. Por exemplo, encontramos no texto CP1\_0002.TXT<sup>114</sup> os seguintes escores padronizados das variáveis integrantes do Fator 2:

TABELA 9 - ESCORES PADRONIZADOS DAS VARIÁVEIS DO FATOR 2 EM TEXTO SELECIONADO DO CORPUS DE ESTUDO

<b>Texto</b>	<b>Variável</b>				
	<b>PR</b> (Presente do indicativo)	<b>DEM</b> (Pronomes demonstrativos)	<b>TOKENS</b> (Quantidade de palavras)	<b>PROP</b> (Nomes próprios)	<b>NUM</b> (Números cardinais)
<b>CP1_0002.TXT</b>	-0,77	-0,97	-0,86	-0,48	+1,61

O cômputo do escore médio do Fator 2 é assim calculado:

$$(PR + DEM + TOKENS) - (PROP + NUM)$$

Ou, com os valores do texto em exame, temos:

---

<sup>114</sup> Cf. Anexo B.

$$(-0,77-0,97 -0,86) - (-0,48 +1,61) = -3,73$$

A mesma ferramenta utilizada no cálculo anterior, a janela “Compute variable” do SPSS, é capaz de realizar essa tarefa automaticamente e criar uma nova variável, a de escore médio do fator.

O método descrito acima será o utilizado no presente trabalho, mas é importante mencionar a crítica pertinente de Lee (2000, p. 205): o cálculo proposto por Biber desconsidera os diferentes *pesos* de cada variável no fator, tratando-os igualmente como se tivessem a mesma correlação<sup>115</sup>.

#### 4.5 DESCRIÇÃO DOS FATORES

A atribuição de escores aos textos individuais do corpus de estudo permite que se observem de forma direta as características manifestadas pelas variáveis dos dois fatores resultantes. Um meio de adquirir conhecimento sobre os fatores é avaliar os textos que estão nos extremos dos escores dos fatores e os que estão com escores com valores próximos de zero, que marcariam a posição “neutra” da manifestação dos fatores. É o que será feito nesta seção, com a apresentação de três exemplos de textos em cada fator: o primeiro texto representa a pontuação mínima do fator no corpus; o segundo possui escore próximo de zero no fator e representa a média; e o terceiro tem a pontuação máxima do escore de fator. Posteriormente os resultados serão comentados.

##### 4.5.1 Textos nos extremos e na média do Fator 1

O texto a seguir obteve o maior escore negativo no Fator 1 no corpus de

---

<sup>115</sup> Como alternativa, Lee (2000, p. 207) adota uma abordagem de contagem de escores pelo método de regressão, que, no entanto, não será utilizada aqui.

estudo (-13,83):

TEXTO 1 - MÁXIMO ESCORE NEGATIVO - FATOR 1

Sintomas aparecem em situações sociais

\* Sintomas físicos: ruborização ou palidez; tremores; sudorese; sensação simultânea de calor e frio; tontura; taquicardia; tensão muscular; diarreia; alterações de apetite; insônia; dor de cabeça; sensação de aperto no tórax (“nó no peito”).

\* Sintomas emocionais: ansiedade antecipatória (a pessoa se sente ansiosa muito tempo antes de algum evento social); sensação de fracasso; insegurança; perda de memória; medo diante de situações banais; preocupação excessiva, desproporcional ao problema; ataques de pânico; mal-estar indefinido, sem motivo aparente; dificuldade de atenção e concentração; apreensão; inquietação; irritabilidade.

Fontes: Instituto de Psiquiatria da UFRJ e Ambulatório de Ansiedade do HC (CP4\_0740.TXT)

O Texto 1 praticamente não traz a marca verbal, exceto por dois verbos, um deles no título. Ele é uma lista de substantivos e nominalizações. Alguns estudiosos talvez não o aceitem como um texto jornalístico. Não é por certo um texto “clássico”, porque lhe faltam sujeito, ação verbal e seus complementos.

O texto abaixo obteve escore de valor zero no Fator 1:

TEXTO 2 - ESCORE MÉDIO - FATOR 1

PSDB e PFL garantem a vitória do governo

DA SUCURSAL DE BRASÍLIA

O governo não conseguiria aprovar o texto principal da reforma da Previdência ontem sem o apoio dos dois partidos de oposição, o PFL e o PSDB, que deram, juntos, 62 votos favoráveis à aprovação da proposta. Ou seja, sem os opositoristas, o governo teria 296 votos, 12 a menos do que os 308 votos necessários para aprovação de emendas à Constituição.

A atuação pró-governo entre os opositoristas ajudou a amenizar as baixas na base aliada ao governo, que contabilizou 56 votos contrários ao relatório do deputado José Pimentel. Os votos contra o governo se reuniram em 8 dos 10 partidos da base.

Proporcionalmente, o PC do B –partido do líder do governo na Casa, Aldo Rebelo (SP)– foi o aliado que mais “traiu” o Planalto. Quatro de seus onze deputados (36,4%) votaram contra o governo. No PT, três dos 93 deputados votaram contra –Luciana Genro (RS), Babá (PA) e João Fontes (SE)– e oito se abstiveram. No PMDB os votos contra o governo contabilizaram 26,9% da bancada. No PTB, esse índice foi de 16%, no PP, 27,1%, no PSB, 17,2%, no PDT, 28,6%, e no PV, 16,7%. O PL e o PPS foram os únicos aliados que não registraram votos contra o Planalto.

A ajuda que veio da oposição teve participação decisiva nos bastidores do senador Antonio Carlos Magalhães (PFL-BA), que conseguiu levar 33 dos 69 pefelistas a votar com o Planalto.

O emissário direto do senador na articulação da votação foi o deputado ACM Neto (PFL-BA), que, antes da votação, já afirmava que o partido teria cerca de 30 votos favoráveis à reforma. No PSDB, o contador de votos pró-governo foi o deputado Custódio Mattos (MG), aliado do governador Aécio Neves (PSDB-MG). O tucano contabilizava antes da votação o número de 28 votos pró-governo. Na verdade, 29 tucanos votaram a favor da proposta.

### **Juízes**

O presidente da AMB (Associação dos Magistrados Brasileiros), desembargador Claudio Baldino Maciel, disse que a entidade pedirá ao Supremo que derrube o subteto salarial do Judiciário estadual de 85,5% do que ganham ministros do Supremo caso o governo insista em manter esse limite.

“A AMB vai propor uma ação direta de inconstitucionalidade no primeiro dia após a promulgação da emenda [da reforma da Previdência] se isso não for corrigido, porque é motivado por mesquinhez”, disse. Ele acusou o governo de fazer “picuinhas”.

Baldino Maciel disse que os representantes da magistratura tentariam convencer deputados e senadores da necessidade de elevação do limite para 90,25% do que ganham ministros do STF.

O presidente da AMB afirmou que os Estados correm o risco de ficar sem nenhum subteto, porque o Supremo certamente derrubará o limite de 85,5%. “Como magistrado, posso afirmar com toda a tranqüilidade que o STF vai declarar inconstitucional.”

Para ele, o governo pode ter insistido no limite menor por vaidade ou para não dar o braço a torcer. “Aparentemente, pelo que se sabe, foi o presidente da República que bateu o pé”, disse.

Segundo Baldino Maciel, haveria uma inconstitucionalidade: muitos Estados fixaram o salário do desembargador em 90,25%: o valor menor implicaria violação do princípio constitucional da irredutibilidade dos salários. (RB)

(CP3\_0360.TXT)

No Texto 2, que representa a média do Fator 1, encontram-se distribuídos de maneira mais uniforme os dois elementos – verbal e nominal. O texto é formado, na verdade, pela soma de dois textos: a primeira parte trata sobre um assunto (aprovação de emenda da Previdência), diferente do encontrado na segunda parte (pleito da classe jurídica), iniciada pelo subtítulo “Juízes”. Isso ocorre às vezes no texto jornalístico e não causa estranheza ao leitor. A sua estrutura lembra a de um texto-colônia (HOEY, 2001), de ocorrência comum no jornal.

O texto com maior escore positivo do Fator 1 (20,18) é mostrado abaixo:

TEXTO 3 - MÁXIMO ESCORE POSITIVO - FATOR 1

Rojas sofre 1ª contestação no São Paulo  
DA REPORTAGEM LOCAL

Invicto como técnico, Roberto Rojas não se intimida em trocar os titulares do São Paulo e, ontem, assistiu à primeira revolta contra um ato seu. O lateral Leonardo foi barrado no coletivo e não perdoou a atitude do interino.

Na vaga na direita jogou Thiago, que estreou contra o Grêmio. O novato só deixou o posto porque sentiu uma fisgada na coxa.

“Não entendo por que ele fez isso. Ele disse que me colocou nos reservas porque ainda tenho que fazer exames. E se não der nada? Encerro a carreira? Se não estou bom para jogar, também não posso treinar”, disparou Leonardo, que não enfrentou o Grêmio porque teve uma indisposição gástrica de causa desconhecida.

Rojas lamentou o fato. “Quem só está com o grupo como titular nunca fez parte do grupo”, afirmou, confirmando que Thiago será o titular caso se recupere.

(CP5\_0878.TXT)

Existe no Texto 3 um relato jornalístico e percebe-se que a ênfase é dada à ação de seus atores. O texto é eminentemente narrativo. Traz ainda, em boa parte de sua extensão, declarações de fontes do texto jornalístico, que representam a transcrição para o modo escrito de enunciados originalmente provenientes do modo oral.

Os três textos apresentados representam o arco de variação representado pelo Fator 1, do extremo negativo ao extremo positivo, passando pelo ponto médio. A tensão verbal versus nominal é a peça fundamental para a compreensão do Fator 1.

Em suma, o Fator 1 revela a oposição entre o pólo verbal – representado pelas variáveis de tempos verbais pretérito perfeito e imperfeito do indicativo, das marcas de primeira e terceira pessoas do singular em verbos e pronomes, dos verbos públicos e dos complementos verbais advérbios e conjunções subordinativas – e, no outro extremo, o pólo nominal – representado pela categoria dos substantivos.

#### 4.5.2 Textos nos extremos e na média do Fator 2

Texto com maior escore negativo no Fator 2 (-22,29):

##### TEXTO 4 - MÁXIMO ESCORE NEGATIVO - FATOR 2

Confira os resultados dos jogos de sábado

A Mega Sena sorteou 04, 15, 31, 34, 37 e 38; a Quina, 07, 24, 41, 52 e 77. Loteria Federal: 54.437, 16.908, 12.124, 10.448 e 72.988. Lotomania: 02, 04, 06, 23, 41, 46, 50, 51, 55, 57, 59, 64, 67, 76, 80, 81, 88, 93, 96 e 98.

(CP7\_1253.TXT)

Chama a atenção, no Texto 4, a abundância de numerais, em razão da necessidade de informar os resultados de loterias. Dificilmente também esse seria um exemplar “típico” ou “clássico” de um texto jornalístico, apesar de possuir um encadeamento sintático básico, de coordenação, diferentemente do que ocorre no texto 1.

Exemplo de texto com escore zero no Fator 2:

TEXTO 5 - ESCORE MÉDIO - FATOR 2

CNI cobra ações emergenciais anti-recessão

DA SUCURSAL DE BRASÍLIA

No dia em que o governo reduziu o IPI dos veículos para tentar estimular as vendas de carros e evitar novas demissões de trabalhadores, a CNI (Confederação Nacional da Indústria) divulgou ontem um documento cobrando ações “emergenciais” do governo contra o “aprofundamento do processo recessivo”.

O documento foi divulgado após reunião do Fórum Nacional da Indústria e traz um diagnóstico da atual situação econômica do país com sugestões para o governo federal.

A CNI avalia que, no atual ambiente, “o investimento não prospera e as empresas não contratam”. Para a entidade, o governo está agindo de forma lenta na hora de reduzir a taxa de juros, e deveria também reduzir as alíquotas do depósito compulsório recolhido ao BC.

O presidente da CNI, deputado federal Armando Monteiro Neto (PTB-PE), disse que essas medidas precisam ser tomadas no “curtíssimo prazo, amanhã [hoje], se fosse possível”.

A CNI sugeriu também que o governo use recursos do FGTS, do FAT e da Cide (contribuição sobre o consumo de combustíveis) para ampliar projetos na área de infraestrutura.

(CP3\_0417.TXT)

O Texto 5, com escore neutro no Fator 2, é um texto jornalístico sobre a divulgação de um documento analítico sobre o país. Há a presença de verbos no presente, utilizados para avaliar o quadro (“o investimento não *prospera* e as empresas não *contratam*”) e sugerir soluções (“*disse* que essas medidas *precisam* ser tomadas”). Além disso, há a ocorrência de pronome demonstrativo (“essas”).

Texto com escore positivo máximo no Fator 2 (19,75):

## TEXTO 6 - MÁXIMO ESCORE POSITIVO - FATOR 2

A SOCIEDADE LÍQUIDA<sup>116</sup>

Maria Lúcia Garcia Pallares-Burke

Especial para a Folha

Um renomado periódico espanhol referiu-se recentemente a Zygmunt Bauman (1927) como um dos poucos sociólogos contemporâneos “nos quais ainda se encontram idéias”. Opinião semelhante é frequentemente exposta por críticos de várias partes do mundo quando refletem sobre o pensamento desse intelectual polonês radicado na Inglaterra desde 1971 e empenhado, há meio século, em “traduzir o mundo em textos”, como diz um deles. Indiferente às fronteiras disciplinares, Bauman é um dos líderes da chamada “sociologia humanística”.

De um lado, não se encontram em suas obras abstrações ou análises e levantamentos estatísticos, e, de outro, são ali aproveitadas quaisquer idéias e abordagens que possam ajudá-lo na tarefa de compreender a complexidade e diversidade da vida humana. Essa é uma das razões pelas quais Bauman tem muito a dizer para uma gama de leitores muito maior do que normalmente se espera de um trabalho de sociologia mais convencional, o que condiz com suas próprias ambições de atingir um público composto de pessoas comuns “se esforçando por ser humanas” num mundo mais e mais desumano. Como ele gosta de insistir, seu objetivo é mostrar a seus leitores que o mundo pode ser diferente e melhor do que é.

Autor prolífico e de renome internacional, pode-se dizer que sua fama e prolixidade aumentaram significativamente após sua aposentadoria, em 1990: 16 de seus 25 livros foram publicados após essa data e cinco obras dedicadas ao estudo de seu pensamento foram escritas nos últimos anos.(...)

Nascido na Posnânia em 1925, Bauman escapou dos horrores do Holocausto que aguardavam os judeus poloneses na Segunda Guerra, ao fugir com sua família para a Rússia em 1939. De lá voltou após a guerra, quando se filiou ao Partido Comunista, estudou na Universidade de Varsóvia e conheceu Janina, com quem está casado há 55 anos e com quem teve três filhas: Anna (matemática), Lydia (pintora) e Irena (arquiteta).

Confiantes e animados pelo sonho de criar uma sociedade mais justa e igualitária, Zygmunt e Janina ali estiveram a construir suas carreiras (ele como professor da Universidade de Varsóvia e, ela, como editora de enredos cinematográficos) e criar sua família, até que uma nova onda de anti-semitismo e repressão esmagou os seus sonhos e os forçou ao exílio. Após três anos em Israel, o convite a Bauman para ser chefe do departamento de sociologia da Universidade de Leeds os trouxe à Inglaterra, onde permanecem até hoje.

Bauman recebeu o Mais! em Leeds, na confortável casa onde mora desde que ali chegou, há mais de 30 anos. “Naquela época achei a cidade horrível, imunda”, me disse Janina, comentando a mudança que ocorreu nos últimos tempos e que transformou Leeds, de um sujo centro industrial, em uma cidade bonita, verdejante e cheia de vida.

O senhor já foi descrito como um “profeta da pós-modernidade” e os termos “pós-moderno” e “pós-modernidade” aparecem em títulos de quatro de seus livros. Estaria

---

<sup>116</sup> O texto foi reduzido a uma amostra com cerca de um quarto de sua extensão original. O texto integral, com 5.002 palavras, encontra-se no Anexo F.

sugerindo que ocorreu uma mudança cultural e social significativa na última geração suficientemente grande para que falemos de um novo período da história?

Uma das razões pelas quais passei a falar em “modernidade líquida” em vez de “pós-modernidade” (meus trabalhos mais recentes evitam esse termo) é que fiquei cansado de tentar esclarecer uma confusão semântica que não distingue sociologia pós-moderna de sociologia da pós-modernidade, entre “pós-modernismo” e “pós-modernidade”. No meu vocabulário, “pós-modernidade” significa uma sociedade (ou, se se preferir, um tipo de condição humana), enquanto que “pós-modernismo” se refere a uma visão de mundo que pode surgir, mas não necessariamente, da condição pós-moderna.

Procurei sempre enfatizar que, do mesmo modo que ser um ornitólogo não significa ser um pássaro, ser um sociólogo da pós-modernidade não significa ser um pós-modernista, o que definitivamente não sou. Ser um pós-modernista significa ter uma ideologia, uma percepção do mundo, uma determinada hierarquia de valores que, entre outras coisas, descarta a idéia de um tipo de regulamentação normativa da comunidade humana e assume que todos os tipos de vida humana se equivalem, que todas as sociedades são igualmente boas ou más; enfim, uma ideologia que se recusa a fazer julgamentos e a debater seriamente questões relativas a modos de vida viciosos e virtuosos, pois, no limite, acredita que não há nada a ser debatido. Isso é pós-modernismo.(...)

O senhor tem sempre enfatizado a necessidade de todos nós “questionarmos ostensivamente as premissas de nosso modo de vida”. Teria alguma sugestão a nos dar sobre as respostas a esses questionamentos?

Maurice Blanchot [escritor e crítico francês, 1907-2003] disse certa vez, em palavras que ficaram famosas, que as respostas são a má sorte das perguntas.

De fato, cada resposta implica fechamento, fim da estrada, fim da conversa. Também sugere nitidez, harmonia, elegância; enfim, qualidades que o mundo narrado não possui. Tenta forçar o mundo numa camisa-de-força na qual ele definitivamente não cabe. Corta as opções, a multidão de sentidos e possibilidades que toda condição humana implica a cada momento. Promete falsamente uma solução simples para um busca provocada e impelida pela complexidade. Também mente, pois declara que as contradições e incompatibilidades que provocam as questões são fantasmas –efeitos de erros lingüísticos ou lógicos, em vez de qualidades endêmicas e irremovíveis da condição humana.

Creio que a experiência humana é mais rica do que qualquer de suas interpretações, pois nenhuma delas, por mais genial e “compreensiva” que seja, pode exauri-la. Aqueles que embarcam numa vida de conversação com a experiência humana deveriam abandonar todos os sonhos de um fim tranqüilo de viagem. Essa viagem não tem um final feliz –toda sua felicidade se encontra na própria jornada.

O senhor descreveu modestamente um de seus livros mais recentes como um “discussion paper”. Diria que é por acaso ou propositadamente que tem se dedicado a escrever ensaios? No curso de meio século de estudos e de escrita nunca consegui adquirir a habilidade de terminar um livro... Com o passar do tempo, eu reconheço que todos os meus livros foram entregues ao editor inacabados. Em regra, antes mesmo que o manuscrito seja impresso, fica claro para mim que o que me parecia havia pouco como “o fim” era, de fato, um começo com uma seqüência desconhecida, mas tremendamente necessária. Por trás de cada resposta percebo que novas questões estão piscando; que mais, muito mais, restou a ser explorado e compreendido e quão pouco, de fato, foi revelado pela “acabamento bem-sucedido” das explorações passadas. As perguntas mais intrigantes e provocantes emergem, via de regra, após as respostas.

No decurso dos anos aprendi a apreciar a queixa de Adorno [filósofo alemão, 1903-69]

sobre a convenção linear da nossa escrita: por causa dessa convenção nós não conseguimos transmitir a lógica do pensamento que, diferentemente da escrita, se move em círculos e está invariavelmente forçada, pelo seu próprio progresso, a fazer perpétuos retornos.

Maria Lúcia Garcia Pallares-Burke é professora aposentada da USP e pesquisadora associada do Centro de Estudos Latino-Americanos da Universidade de Cambridge (Reino Unido). É autora de “As Muitas Faces da História” (ed. Unesp).

(CP7\_1293.TXT)

O Texto 6 é, em extensão, o maior texto encontrado no corpus de estudo. Ele tem um nível de articulação lógica muito alto, possivelmente pelo fato de ser o depoimento de um filósofo (observe-se que, apesar de o formato ser o de uma entrevista, a linguagem utilizada lembra pouco o discurso oral). O texto apresenta também grande número de verbos no tempo presente.

O três textos apresentados são muito diferentes entre si. Um grande traço contrastante no Fator 2 é a extensão, mensurada pela variável quantidade de palavras. A apresentação dos textos que cobrem o espectro de variação do Fator 2 sugere que há duas forças atuantes e contrárias: uma deixa mais articulado o discurso (com a presença das variáveis presente do indicativo, pronomes demonstrativos e quantidade de palavras e, de outra parte, a ausência de nomes próprios e numerais); outra preocupa-se com a precisão em apontar pessoas e dados (presença das variáveis nomes próprios e numerais e correspondente ausência das demais variáveis do fator).

Essa foi a descrição dos dois fatores identificados na Análise Fatorial, a partir de um corpus jornalístico. Como dito na Introdução, os fatores não guardam nenhuma relação apriorística com os gêneros a que pertencem os textos do corpus de estudo. Ou, dito de outra maneira, os fatores 1 e 2 (e, por extensão, as dimensões, de que se originam) podem balizar outras classificações de gêneros do jornal, diferentes da que será utilizada a seguir.

#### 4.6 CLASSIFICAÇÃO DOS TEXTOS EM GÊNEROS

Com os escores individuais dos textos por fator identificados, é possível realizar pesquisas em nível microlingüístico sobre a atuação do fator em textos

determinados. Uma extensão desse raciocínio é verificar em que medida os fatores incidem sobre grupos homogêneos ou variedades de textos – no caso aqui apresentado, gêneros textuais/discursivos que são encontrados no corpus jornalístico.

Para a realização dessa etapa, o corpus foi classificado manualmente de acordo com o seu gênero, agrupando os textos em conjuntos definidos previamente, conforme a definição estabelecida após uma pesquisa na literatura (cf 3.5.1). A classificação foi realizada por este autor, cuja experiência profissional no ramo jornalístico (mais de dez anos, no Grupo Folha) pode ser qualificada como a de um especialista. A Tabela 10 mostra o número de textos classificados em cada gênero no corpus de estudo.

TABELA 10 - NÚMERO DE TEXTOS DO CORPUS DE ESTUDO, CLASSIFICADO POR GÊNERO

<b>Gênero</b>	<b>Número de textos</b>
<b>Artigo</b>	<b>86</b>
<b>Carta</b>	<b>84</b>
<b>Chamada</b>	<b>73</b>
<b>Comentário</b>	<b>11</b>
<b>Crítica</b>	<b>47</b>
<b>Crônica</b>	<b>12</b>
<b>Editorial</b>	<b>20</b>
<b>Entrevista</b>	<b>24</b>
<b>Coluna de notas</b>	<b>73</b>
<b>Notícia</b>	<b>296</b>
<b>Reportagem</b>	<b>655</b>
<b>Resenha</b>	<b>18</b>
<b>Outros</b>	<b>32</b>
<b>Total</b>	<b>1.431</b>

O gênero reportagem, com 655 textos selecionados, responde por quase a

metade dos textos do corpus de estudo. O segundo gênero com maior número de textos é a notícia, com o correspondente a 20% do total. Os dois gêneros somam dois terços do total em número de textos e cerca de 65% em número de palavras, o que justifica serem estudados com atenção. É um valor semelhante ao obtido por Chaparro (1997, p. 139), que mensurou o espaço ocupado pelos gêneros em quatro jornais brasileiros de prestígio<sup>117</sup> e verificou que a participação somada dos gêneros reportagem e notícia corresponde a 70% do corpus (em cm<sup>2</sup>).

#### 4.7 CÁLCULO DOS ESCORES MÉDIOS DOS GÊNEROS

Nesse ponto da pesquisa, reunimos as duas informações básicas para a análise dos gêneros do jornal: de um lado, os escores individuais dos textos, de acordo com cada fator resultante da Análise Fatorial; de outro, como explicado acima, o corpus de estudo dividido em grupos de textos de acordo com o gênero. Interessa observar como os escores médios dos grupos de gêneros distribuem-se ao longo dos *continua* que representam os dois fatores obtidos e as características estatísticas dessa distribuição.

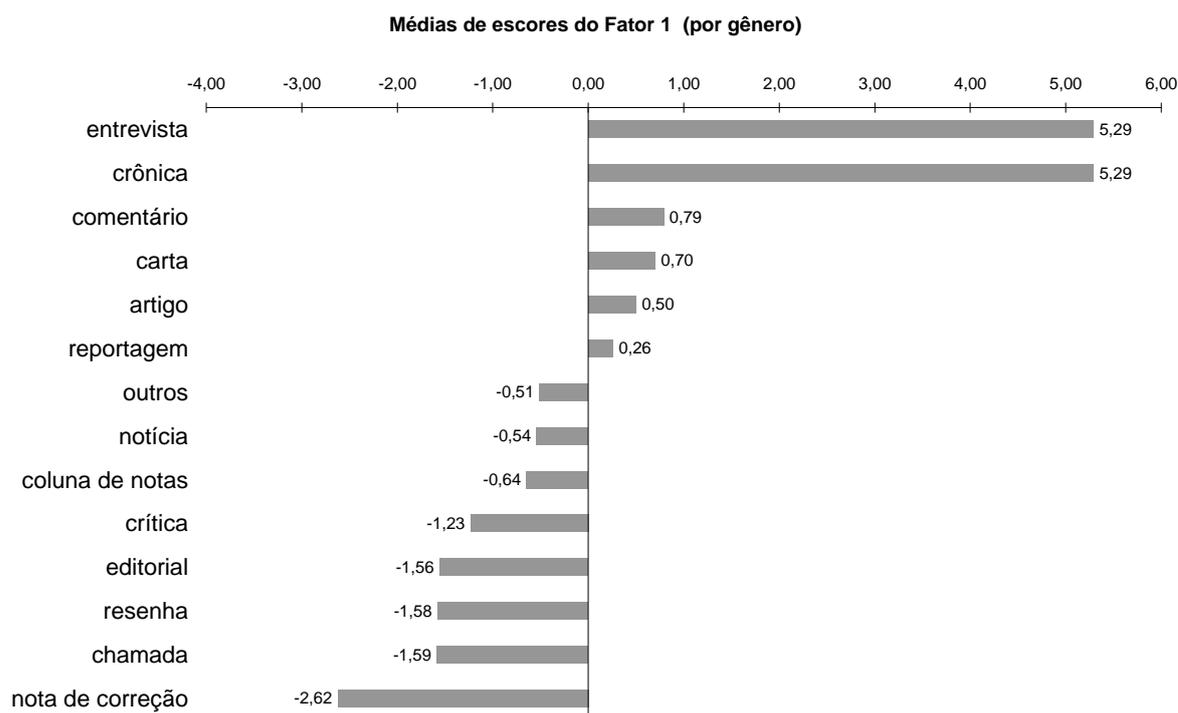
A obtenção do escore médio por gênero é feita com o auxílio de outro módulo do SPSS, a janela “Univariate” (menu “Analyze” / “General Linear Model”). Os parâmetros a serem definidos são os seguintes: inserir o escore médio do fator (um por vez) no campo “Dependent Variable”; inserir a variável de gênero no campo “Fixed Factor(s)”; e, pressionando o botão “Options”, selecionar “Descriptive Statistics”. Os resultados aparecerão no SPSS Viewer. Importa analisar as médias obtidas por fator (coluna “Mean” do quadro “Descriptive Statistics”), que servirão para criar uma escala na qual estarão dispostos os diversos gêneros, conforme será mostrado a seguir.

---

<sup>117</sup> “Folha de S. Paulo”, “Jornal do Brasil”, “O Estado de S. Paulo” e “O Globo”.

O Gráfico 2 mostra com se distribuem os gêneros no Fator 1. Os maiores valores positivos foram alcançados pelos gêneros entrevista e crônica (5,29 para ambos). Em seguida, três gêneros – comentário, carta e artigo – têm média semelhante, igual ou pouco superior a 0,50. Reportagem vem logo abaixo, com 0,26. No pólo negativo do Fator 1, o gênero que ocupa a posição mais extrema é a nota de correção (-2,62). Chamada, resenha e editorial têm média semelhante, ao redor de -1,50. Pouco abaixo vem crítica, com -1,23. Coluna de notas, notícia e outros gêneros formam um outro grupo, com valores próximos a -0,50.

GRÁFICO 2 - ESCORES MÉDIOS DO FATOR 1, POR GÊNERO

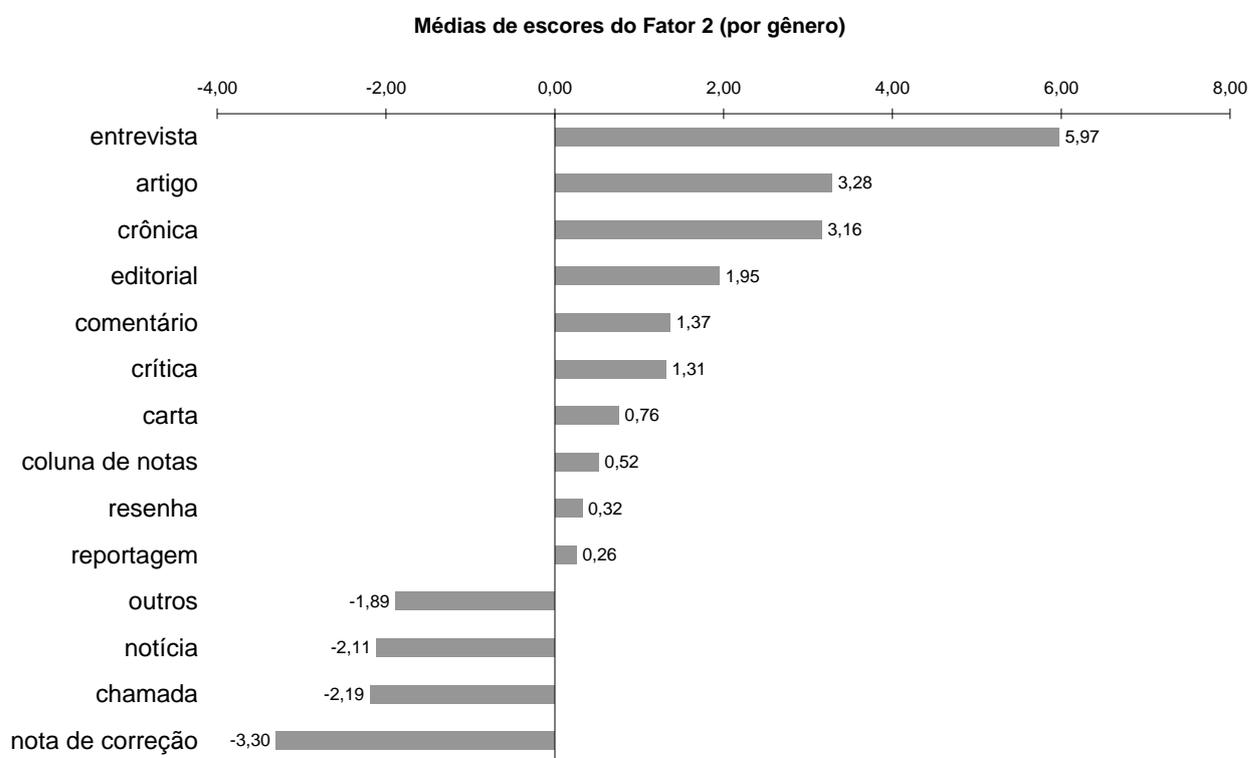


O Gráfico 3 traz os escores médios correspondentes dos gêneros no Fator 2. Ele concentra mais gêneros no pólo positivo (10) que no negativo (4) e apresenta a entrevista como o gênero de maior valor positivo (5,97). Artigo e crônica, também no

campo positivo, têm médias parecidas (3,28 e 3,16). Editorial vem em seguida, com 1,95. Comentário e crítica formam um grupo intermediário, acima de um grupo maior, com escores fracamente positivos, próximos a zero – carta, coluna de notas, resenha e reportagem. No pólo negativo, sobressai um grupo com três gêneros, outros (-1,89), notícia (-2,11) e chamada (-2,19); no extremo do pólo negativo, novamente, está a nota de correção (-3,30).

Para uma correta avaliação dos resultados obtidos nos gráficos mencionados, é necessário verificar em que medida as distribuições são significativamente diferentes, sob o ponto de vista estatístico. Para isso, são realizados nessa fase testes *post-hoc*, mostrados a seguir.

GRÁFICO 3 - ESCORES MÉDIOS DO FATOR 2, POR GÊNERO



#### 4.7.1 Testes estatísticos *post-hoc*

São examinados os valores de F, p e R<sup>2</sup>, testes estatísticos<sup>118</sup> que revelam, respectivamente: se, com base nos dados observados, há diferenças significantes entre os gêneros; se a diferença encontrada entre os grupos não é um resultado fortuito, do acaso; e se o escore médio obtido pelos gêneros nos fatores é relativamente importante<sup>119</sup>. A Tabela 11 mostra os resultados obtidos por esses três testes no cômputo dos dois fatores.

Os valores de F obtidos – de 6,7 para o Fator 1 e de 50,4 para o Fator 2 – revelam que ambos os fatores são considerados como estatisticamente relevantes para indicar que são capazes de diferenciar os grupos de gêneros. Há uma capacidade maior, porém, de os gêneros diferenciarem-se no Fator 2, pelo valor de F ser bem superior que o encontrado no Fator 1.

O valor de p é zero em ambos os fatores. Isso significa que o resultado obtido pela distribuição dos agrupamentos de gêneros ao longo dos fatores não pode ser atribuído ao acaso.

TABELA 11 - TESTES ESTATÍSTICOS DA VARIAÇÃO ENTRE GÊNEROS, POR FATOR (VALORES DE F, p E R<sup>2</sup>)

Fator	F	p	R <sup>2</sup>
<b>Fator 1</b>	6,7	0,00000	5,8%
<b>Fator 2</b>	50,4	0,00000	31,6%

<sup>118</sup> No ambiente do SPSS Viewer, obtidos no item “Tests of Between-Subjects Effects”.

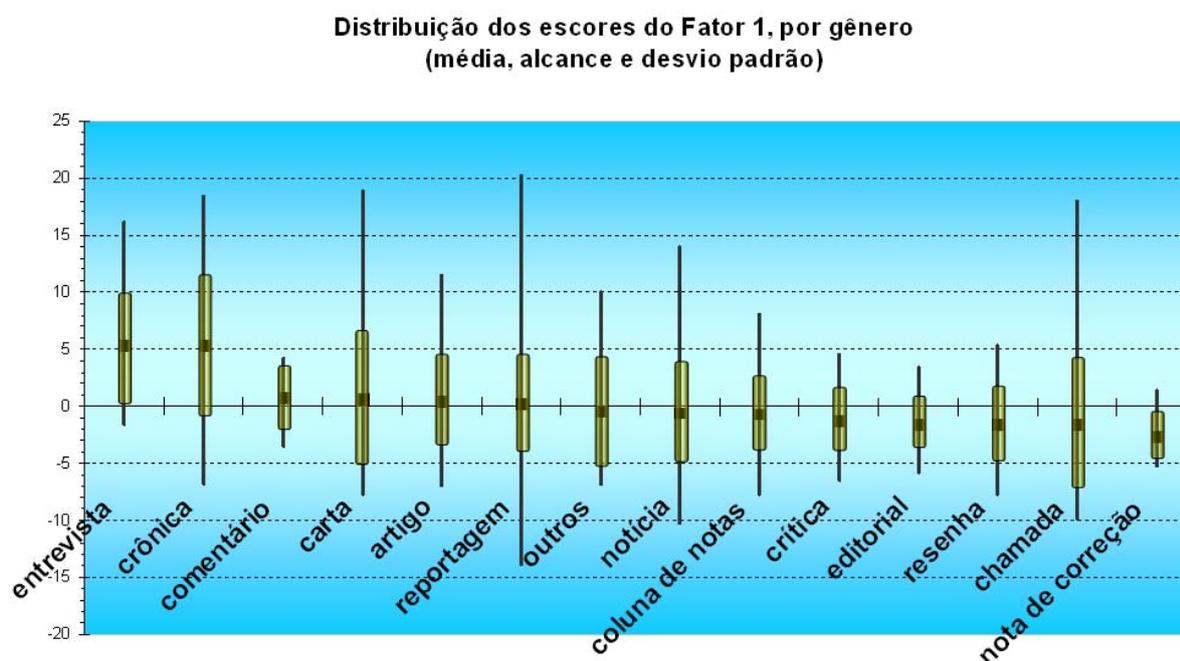
<sup>119</sup> O valor de p (mostrado como “Sig.” no SPSS) deve ser inferior a 0,05 (isto é, com um grau de confiança maior que 95%). R<sup>2</sup> deve ter seu valor multiplicado por 100 e ter seu resultado apresentado como percentual (%).

O quadrado de R, de sua parte, mostra que o Fator 2, sendo o componente diferenciador de maior importância, tem um poder maior de predição que o Fator 1.

Os testes estatísticos efetuados, que avaliam a qualidade da diferenciação entre grupos, foram claros e mostram que a variação lingüística expressa pelos fatores influencia significativamente os gêneros. Uma interpretação adequada dos fatores deve levar em conta, portanto, a influência que o fator tem sobre cada gênero.

#### 4.7.2 Focos dimensionais nos gêneros

GRÁFICO 4 - DISTRIBUIÇÃO DOS ESCORES DO FATOR 1, POR GÊNERO (MÉDIA, DESVIO PADRÃO E ALCANCE)



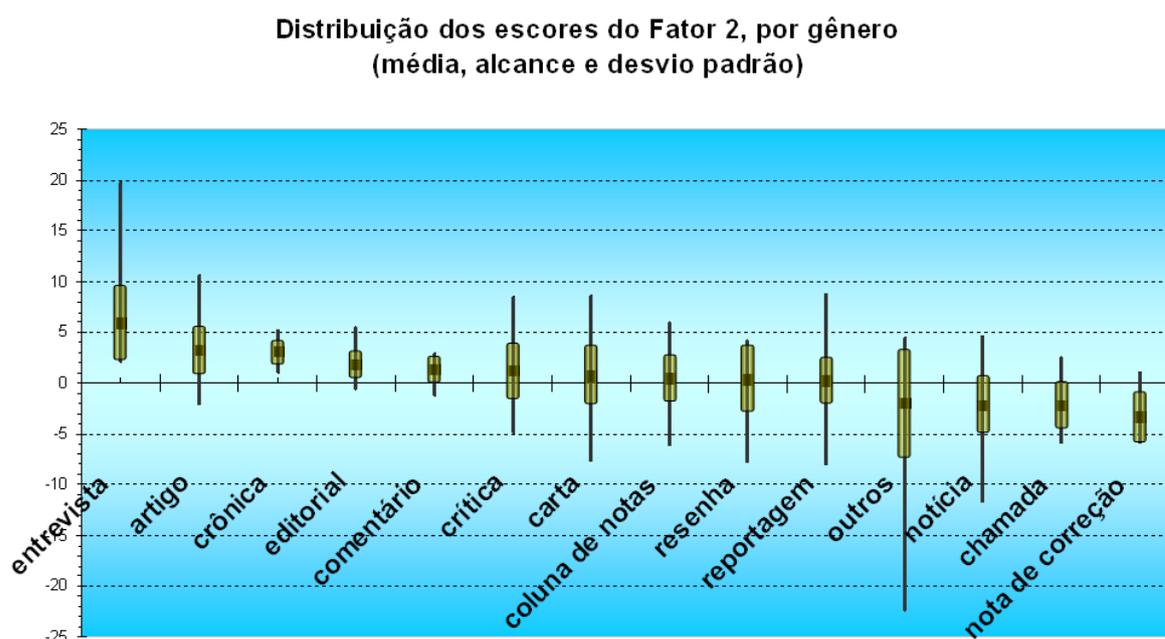
São apresentados aqui dois gráficos (Gráficos 4 e 5) que mostram complementarmente os resultados exibidos nos Gráficos 2 e 3 (os dados originais de todos eles estão no Anexo C). Eles propiciam a visão não só das médias obtidas pelos gêneros, mas também da faixa que concentra a distribuição da maioria dos textos de cada gênero (indicada pela barra que atravessa verticalmente os pontos médios; a medida baliza dois desvios padrões, onde se concentram cerca de dois terços dos textos do grupo) e dos limites máximo e mínimo obtidos pelos escores no gênero (mostrados pela linha fina vertical; ela marca, portanto, o alcance de cada gênero).

As três medidas – média, alcance e desvio padrão – dão uma noção mais detalhada da concentração dos escores e das faixas em que se situam os gêneros, em relação às médias exibidas nos Gráficos 2 e 3.

O Gráfico 4 indica que o alcance é muito amplo em quase todos os gêneros. Há textos no gênero reportagem, por exemplo, que atingiram tanto o valor máximo positivo como o máximo negativo. Observa-se também que não há nenhum gênero que não possua textos apenas em um pólo do Fator 1. Ao que parece, há dentro dos gêneros grande variação interna do Fator 1 entre os textos. E, com o olhar mais distanciado, a conclusão que se tira é que todos os gêneros parecem funcionar em uma faixa de variação muito sutil no Fator 1.

A análise do Gráfico 5 mostra que os gêneros se distinguem mais claramente nos escores do Fator 2. O “foco” representado pela barra de desvio padrão está visivelmente mais estreito que no gráfico do Fator 1. O alcance está menor também, o que demonstra haver uma maior homogeneidade do fator entre os textos dentro dos grupos de gêneros.

GRÁFICO 5 - DISTRIBUIÇÃO DOS ESCORES DO FATOR 2, POR GÊNERO (MÉDIA, DESVIO PADRÃO E ALCANCE)



Os quatro gráficos mostrados nesta seção dão subsídios adicionais para interpretar os fatores de variação que incidem sobre o corpus jornalístico. Podemos apurar quais são as características retóricas que mais influem na variação dos fatores, examinando exemplares dos gêneros com cargas positivas, neutras ou negativas dos fatores.

Outro exercício interessante é o de buscar textos “típicos” de cada gênero com base nos fatores resultantes, ou seja, selecionar um exemplar de texto em cada gênero que possua escores os mais aproximados possível das médias obtidas pelo gênero nos dois fatores. O Anexo D traz exemplares “típicos” de cada gênero do corpus de estudo com base nessa metodologia.

#### 4.8 DIMENSÕES DO TEXTO JORNALÍSTICO

Esta seção apresenta a proposta de interpretação dos fatores resultantes da Análise Fatorial. Como visto anteriormente, Biber (1988) chama de dimensões os fatores que são interpretados no âmbito das funções comunicativas que exercem, como responsáveis pela variação lingüística do corpus.

Serão utilizadas como base para a proposta de denominação das dimensões três informações apresentadas neste capítulo: a) as variáveis que compõem cada fator (cf. 4.3); b) na forma com que os gêneros se distribuem ao longo dos fatores (cf. 4.7); e c) em exemplares de texto do corpus selecionados, de acordo com os escores que obtiveram (cf. 4.5 e Anexo D).

A análise do Fator 1 mostra a preponderância de variáveis ligadas aos verbos (pretérito perfeito, verbos/pronomes da 3ª pessoa singular, conjunções subordinativas, verbos públicos, pretérito imperfeito, advérbios e verbos/pronomes da 1ª pessoa singular) no pólo positivo e, no pólo oposto, encontra-se a variável nominal por excelência, a da categoria dos substantivos. Os exemplos colhidos mostram que um texto com valor alto no pólo negativo do Fator 1 (Texto 1) não é fluente do ponto de vista sintático e lembra mais uma lista. No outro extremo, o texto com escore alto no

pólo verbal (Texto 3) é rico em contar detalhes do conflito entre o jogador de futebol e o seu técnico.

Entrevista e crônica – os textos mais coloquiais – são os gêneros que obtiveram maior escore médio no pólo verbal, enquanto o maior escore do pólo nominal foi o gênero nota de correção. Em um patamar mediano negativo estão praticamente juntos a chamada, a resenha e o editorial. Os demais gêneros estão concentrados a pouca distância um do outro, localizados um pouco acima ou um pouco abaixo do valor zero. A força atuante na variação lingüística do Fator 1 compõe uma massa relativamente homogênea, com poucos casos de escores nas áreas mais extremas.

Ante esse quadro a respeito do Fator 1, a hipótese aventada neste trabalho é a de que a **Dimensão 1** concentra um pólo **narrativo** (pólo verbal, dos escores positivos) e um pólo **expositivo** (pólo nominal, dos escores negativos). A nomenclatura da Dimensão 1, portanto, será “Narrativo versus Expositivo”.

Em relação ao Fator 2, as variáveis co-ocorrentes no pólo positivo – presente do indicativo, pronomes demonstrativos e quantidade de palavras – são uma reunião de diversas categorias, diferentemente do que ocorreu no Fator 1. O pólo negativo, em oposição, traz as variáveis nomes próprios e números cardinais. Ambos os pólos não formam combinações intuitivas. As variáveis merecem ser observadas atentamente, porque são as responsáveis por explicar a maior parte da variação lingüística entre os gêneros, muito superior à alcançada pela Dimensão 1 (cf. 4.7.1).

Deve também ser lembrado que a variável quantidade de palavras mede a extensão dos textos. Esse é um elemento certamente influente na produção dos textos jornalísticos, que sofrem pressão constante para garantir seu espaço no conjunto.

O presente do indicativo é o componente verbal do Fator 2, naturalmente apartado dos tempos verbais do pretérito do Fator 1 porque é de utilização excludente. Ressalte-se o uso disseminado de títulos do texto jornalístico no presente do indicativo, que faz com que a distribuição dessa variável não seja tão contrastiva no universo pesquisado. O presente não é utilizado no texto jornalístico em notícias e

reportagens, que costumam preferir a forma do pretérito perfeito (LAGE, 1998). O presente do indicativo parece ser mais usado em textos **argumentativos**, descolados do factual e mais voltados para a reflexão, como nos gêneros entrevista, artigo e crônica, que obtiveram os maiores escores positivos no Fator 2.

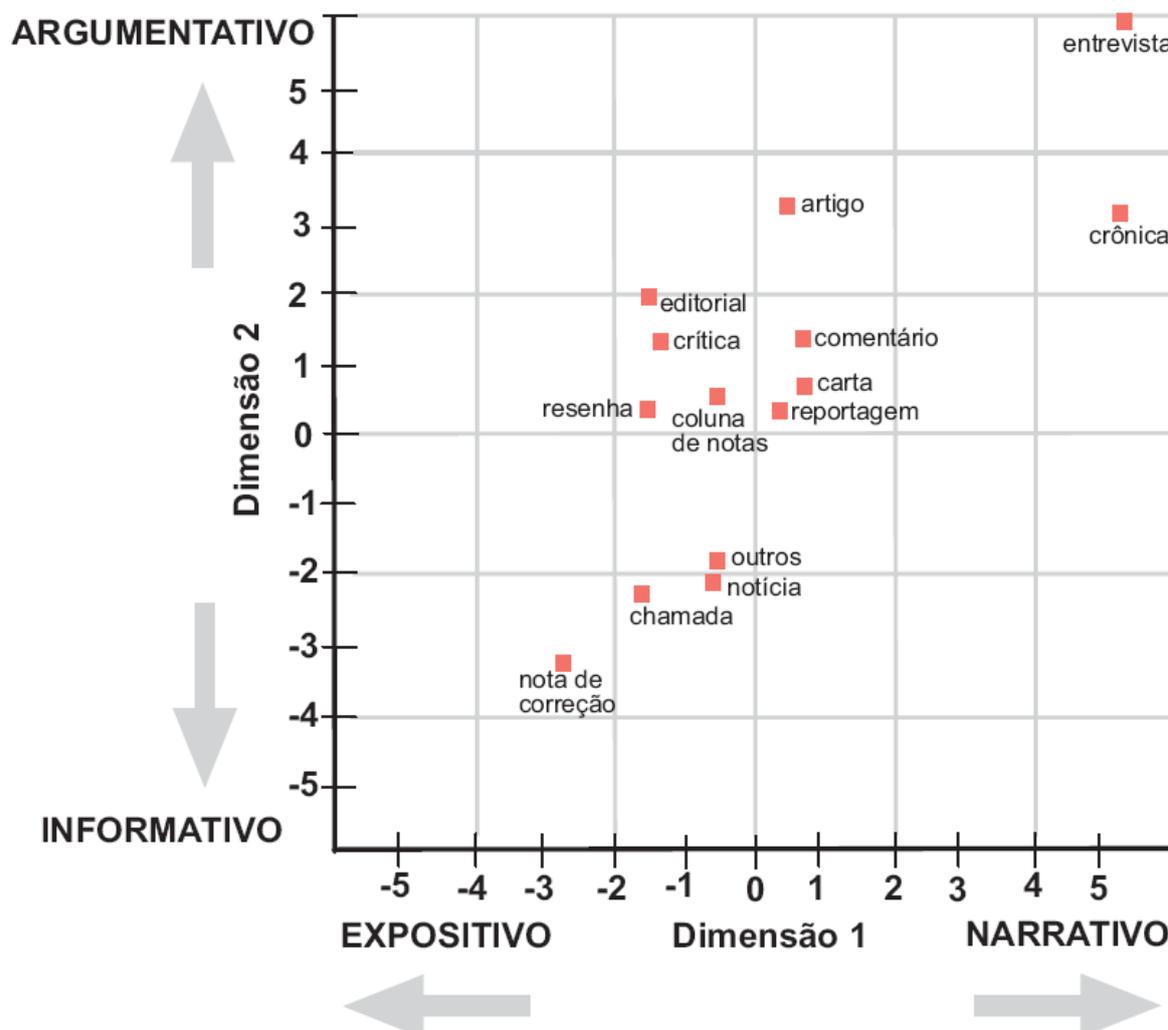
Por fim, o pólo positivo do Fator 2 é composto pelos pronomes demonstrativos, teoricamente associados à classe dos dêiticos. Para Bechara (1999), “a posição indicada pelo demonstrativo pode referir-se ao espaço, ao tempo (demonstrativos dêiticos, espaciais e temporais) ou ao discurso (demonstrativo anafórico)”. Os demonstrativos são, em suma, marcas de argumentação que reforçam a coesão e coerência do texto.

O pólo oposto do Fator 2, que concentra as variáveis nomes próprios e números cardinais, é um pólo nominal, como o do Fator 1. Há um empilhamento de informação impressionante na incidência máxima desse escore, como visto no Texto 4. Os gêneros que mais obtiveram escores nesse pólo foram nota de correção, chamada e notícia, gêneros **informativos**. A ênfase está na exatidão de nomes e números, e não nas outras variáveis do pólo positivo (extensão do texto, dêiticos e tempo presente, que, inversamente, estão ausentes). Denominaremos, portanto, a **Dimensão 2** de “Argumentativo versus Informativo”.

A Análise Multidimensional de Biber (1988) subsidiou os resultados deste trabalho, em que se obteve uma bidimensionalidade funcional da língua portuguesa no corpus de estudo, que pretendeu representar a imprensa brasileira contemporânea. As dimensões “Narrativo versus Expositivo” e “Argumentativo versus Informativo” seriam os eixos sobre os quais podem ser acomodados os diferentes gêneros textuais/discursivos presentes nas páginas do jornal. A próxima seção faz um resumo gráfico que exprime tal bidimensionalidade.

## 4.9 APRESENTAÇÃO COMBINADA DAS DIMENSÕES: UMA SÍNTESE

GRÁFICO 6 - DISTRIBUIÇÃO BIDIMENSIONAL DOS ESCORES MÉDIOS, POR GÊNERO



O Gráfico 6 mostra a relação espacial ocupada pelos gêneros ao longo das dimensões, a partir dos escores médios apresentados anteriormente (cf. Gráficos 2 e 3). Os gêneros ocupam três quadrantes da área do gráfico, delimitados pelos eixos da coordenada e da abscissa, que se cruzam no valor zero de ambos. A maioria dos gêneros se concentra no quadrante “**Narrativo-Argumentativo**”: entrevista, artigo, crônica, comentário, carta e reportagem. O quadrante oposto a ele, o “**Expositivo-Informativo**”, abriga os gêneros nota de correção, chamada, notícia e outros. Entre os dois quadrantes está o quadrante “**Expositivo-Argumentativo**”, área em que se

encontram os gêneros editorial, resenha, crítica e coluna de notas.

O mapeamento realizado permite aferir instigantes relações de semelhança e diferença entre os gêneros. A reportagem, gênero mais freqüente, está situada no ponto mais próximo ao ponto médio do gráfico (valor zero nas duas dimensões). É o gênero, assim, mais equilibrado do corpus, em termos das dimensões que atuam na língua em uso do jornal. Já o gênero mais afastado dos demais é a entrevista – não por acaso, o gênero que é o mais próximo no corpus do modo oral. A notícia revela ter um “parentesco” com a chamada, ambas no quadrante “Expositivo-Informativo”. A resenha está muito próxima da crítica, o suficiente para questionar por que os dois gêneros não estão juntos, como propõe Marques de Melo (1994). A nota de correção aparece como o gênero com as maiores cargas informativas e expositivas do corpus. O editorial, um gênero consensualmente considerado como opinativo, tem uma dimensão argumentativa mais contida que o artigo, a crônica ou a entrevista.

Mesmo sem esgotar todas as possibilidades de interpretação, acreditamos que o gráfico apresentado nesta seção tenha cumprido o objetivo de resumir os resultados obtidos na pesquisa.

A seguir será feito um balanço dos resultados, das perspectivas de análise e de novos rumos que podem ser propostos para a pesquisa de gêneros jornalísticos.

#### 4.10 DISCUSSÃO

Uma vez descritos e interpretados os fatores, transformados em dimensões, conclui-se a Análise Multidimensional proposta por esta pesquisa. Como última etapa do trabalho, contudo, ainda se faz necessário expor algumas reflexões a respeito da relação estabelecida entre este trabalho (e seus resultados) e o de outros autores, no sentido de ampliar os horizontes teóricos e metodológicos das duas principais áreas envolvidas, a Lingüística de Corpus e os Estudos de Comunicação.

Foi discutido no capítulo de Fundamentação Teórica que a tipologia dos

gêneros encontrados no jornal não é consensual entre os autores que os descreveram (NABANTINO RAMOS, 1970; MARQUES DE MELO, 1987, 1994, 1998a; CHAPARRO, 1997; RABAÇA; BARBOSA, 1998; BONINI, 2003, 2004). Até onde os resultados obtidos aqui podem indicar, por meio da AMD foi possível identificar semelhanças e diferenças entre os diversos grupos de textos e oferecer parâmetros para delimitar os gêneros do jornal. Este estudo, assim, contribui para as áreas de Lingüística Aplicada e de Estudos de Comunicação no sentido de proporcionar uma perspectiva inovadora de observação do corpus jornalístico.

Um aspecto a destacar, nesse sentido, é o de que a metodologia utilizada nesta pesquisa permite também que os grupos de textos possam ser reorganizados – dado que a classificação em subcorpora de gêneros só foi empregada *após* a Análise Fatorial do corpus de estudo como um todo. É possível, em suma, *replicar* o experimento de acordo com outras bases teóricas de classificação de gêneros e inseri-los novamente nas duas dimensões encontradas. Desse modo, a moldura teórica que procura explicar a variação lingüística do corpus de estudo mantém-se como suporte para o mapeamento de outros grupos de textos que se queira testar. Essa é uma evidência de utilização dos resultados que converge com a discussão, no âmbito da Lingüística de Corpus (cf. 2.1), relativa à importância de considerar novos critérios de classificação de um corpus, incluindo os critérios lingüísticos (EAGLES, 1996a, 1996b; BIBER, 1993, 1994; LEE, 2001).

Outra reflexão que deve ser feita diz respeito à maneira como as dimensões encontradas podem ser comparáveis à literatura que trata do estudo de variação lingüística por meio da AMD (BIBER, 1988; LEE, 2000). Em relação às dimensões encontradas em Biber (1988), a maioria das variáveis utilizadas aqui, com algumas diferenças metodológicas (já que lidam com línguas e ferramentas diferentes), está presente na Dimensão 1 ou na Dimensão 2 do estudo de Biber (exceto pelos advérbios,

que estão na Dimensão 3 de Biber<sup>120</sup>). Há um número relevante de variáveis presentes na Dimensão 2 de Biber, que trata justamente da narratividade, na composição da Dimensão 1 examinada aqui: pretérito perfeito, verbos / pronomes da 3ª pessoa do singular, verbos públicos – e, talvez, pretérito imperfeito. Do mesmo modo, pode-se dizer que as variáveis da Dimensão 2 utilizadas aqui estão direta ou indiretamente relacionadas à Dimensão 1 de Biber (1988): presente do indicativo, pronomes demonstrativos, quantidade de palavras (medida de especificidade lexical semelhante à variável de Biber razão forma / palavra, que ocorre na D1), nomes próprios e números cardinais – as duas últimas relacionadas ao domínio nominal. A semelhança dos eixos dimensionais “Narrativo versus Expositivo” e “Argumentativo versus Informativo” com as dimensões 1 e 2 de Biber (1988) pode indicar que a solução fatorial encontrada nesta pesquisa reflete, ao menos em parte, as forças atuantes das dimensões consideradas mais consistentes, do ponto de vista teórico, do estudo de Biber (1988) (cf. LEE, 2000).

Uma crítica feita por Lee (2000) a Biber (1988) é com relação à representatividade do corpus que Biber utilizou. É, por definição, impossível determinar quais são as exatas proporções de gêneros que compõem o idioma inglês; é, portanto, arriscado generalizar a partir do corpus estudado pelo autor (LEE, 2000, p. 391). Não seria possível dizer o mesmo em relação ao trabalho aqui exposto, uma vez que o corpus de estudo, definido como uma amostra de um jornal impresso brasileiro de prestígio, tem um foco muito mais preciso que o de Biber – e pode, portanto, ser mais rigoroso em termos estatísticos. É factível supor que a reunião de sete edições completas de um jornal logrou capturar todas as mais frequentes variedades de linguagem da imprensa diária escrita, na proporção atestada pela prática e uso.

Em relação às dimensões identificadas nesta pesquisa – os eixos “Narrativo

---

<sup>120</sup> Lee (2000), no entanto, argumenta que há uma correlação relevante entre a D1, a D3 e a D5 de Biber e que, em muitas soluções estatísticas alternativas à adotada por Biber (1988), a D3 e a D5 tendem a entrar em colapso e, às vezes, se fundir com a D1 (LEE, 2000, p. 395).

versus Expositivo” e “Argumentativo versus Informativo” –, é notável a sua semelhança com algumas propostas teóricas da área de Estudos de Comunicação que estabeleceram uma categorização dos gêneros jornalísticos (MARQUES DE MELO, 1994; CHAPARRO, 1997; cf. 2.3.6). Elas serão abordadas abaixo.

Marques de Melo (1994) identifica uma distinção formal inicial básica, formada pelo jornalismo opinativo e o jornalismo informativo. A distribuição dos gêneros nos dois grupos efetuada pelo autor é bastante análoga à encontrada ao longo da Dimensão 2, eixo dimensional denominado “Argumentativo versus Informativo”, excetuando-se os gêneros reportagem e entrevista (ambos citados por Marques de Melo como pertencentes ao grupo informativo). O pólo Argumentativo estaria associado ao jornalismo opinativo, enquanto o pólo Informativo estaria relacionado com o jornalismo informativo. Os resultados obtidos nesta pesquisa, partindo de abordagem diversa à de Marques de Melo, confirmam em larga medida a taxonomia proposta pelo autor. A diferença é que, diferentemente da tipologia de Marques de Melo, a metodologia utilizada permite caracterizar com maior refinamento se os gêneros estão a maior ou menor distância de seus pólos dimensionais.

Chaparro (1997), por seu turno, refuta a terminologia (mas não a tipologia<sup>121</sup>) proposta por Marques de Melo e propõe, inspirado nos esquemas de superestruturas Van Dijk para o discurso jornalístico, a existência de dois esquemas básicos de organização dos gêneros: o esquema narrativo e o esquema argumentativo, que relatam e comentam os acontecimentos, respectivamente (CHAPARRO, 1997, p. 60). A esses dois esquemas, que o autor chama de gêneros, correspondem agrupamentos de espécies – argumentativas, narrativas e práticas<sup>122</sup> –, das quais derivariam os gêneros propriamente ditos (que o autor chama de espécies). A proposta inicial de Chaparro, de

---

<sup>121</sup> A esse respeito, Chaparro parece aceitar os argumentos de Marques de Melo quando escreve: “Os fazeres da ação jornalística são dois: **relatar** a atualidade; **comentar** a atualidade [grifos do autor]. Com **opinião** e **informação** [grifos nossos]. Nada além disso” (CHAPARRO, 1997, p.62).

<sup>122</sup> A quarta espécie é não-verbal, denominada gráfico-artísticas.

buscar uma formulação alternativa e matizada, em contraposição aos pólos tipológicos de Marques de Melo, acaba caindo no mesmo tipo de oposição bipolar, com a exceção do gênero “híbrido” coluna (que o autor inclui tanto como uma espécie argumentativa como uma espécie narrativa). Os componentes teóricos introduzidos por Chaparro na discussão são, portanto: a) a proposta de fugir de uma polarização conceitual dicotômica e b) a idéia de *narração* como um dos componentes básicos da linguagem do jornal. Em certa medida, ambas as idéias lograram ser utilizadas e desenvolvidas neste trabalho. Os resultados apresentados aqui criaram, conforme o desejo de Chaparro, uma percepção mais sutil dos gêneros que estão presentes em cada pólo das dimensões, pois estão dispostos em um contínuo. Em relação à segunda contribuição de Chaparro, deve ser dito que o eixo dimensional que opõe narração a exposição foi o maior responsável pela variação lingüística encontrada no corpus jornalístico estudado (cf. 4.3) – apesar de a distribuição da Dimensão 1 entre os gêneros ter-se revelado menos contrastada que a da Dimensão 2. A instância da narração, assim, deve ser também levada em conta na descrição das categorias mais gerais das quais os gêneros do jornal fazem parte<sup>123</sup>.

Foram apresentados e discutidos neste capítulo os resultados do processo da Análise Multidimensional de um corpus jornalístico, conduzido fundamentalmente segundo os princípios teórico-metodológicos propostos em Biber (1988). A apresentação dos dados implicou a *descoberta* – uma vez que a pesquisa teve um sentido exploratório e empírico – de duas dimensões possuidoras de determinadas funções comunicativas que perpassam todos os textos analisados, em maior ou menor medida, e são responsáveis por uma significativa parcela da variação lingüística observada no corpus. As dimensões resultantes revelaram, por fim, a capacidade de identificar semelhanças e diferenças de natureza lingüística entre os gêneros, como mostraram os resultados estatísticos *post-hoc* (cf. 4.7.1).

---

<sup>123</sup> Lage (1998, p. 35-38) também identifica no texto jornalístico uma ordem narrativa, oposta a uma ordem expositiva, mas focado no gênero notícia.

O capítulo a seguir encerra a dissertação com a sumarização da pesquisa, sua crítica e suas limitações, bem como com o exame de alguns aspectos ligados às possibilidades de aplicação pedagógica dos resultados aqui obtidos.

## 5 CONCLUSÃO

À guisa de conclusão, inicialmente este capítulo irá rever de forma resumida a problemática que originou este trabalho, os objetivos daí derivados e as questões de pesquisa que especificaram as suas metas. Serão também retraçados brevemente o caminho teórico-metodológico percorrido e os resultados atingidos – sempre que possível, com um olhar mais distanciado que a forma tratada anteriormente. A crítica ao experimento merece, nesse ponto, ser explicitada, seguida pela menção de alguns limites metodológicos identificados durante a consecução do experimento. Por fim, serão abordadas a perspectiva de novos estudos que poderiam ser criados a partir dos resultados obtidos, as aplicações pedagógicas imaginadas e as possíveis contribuições deste trabalho nas áreas de Lingüística de Corpus, Lingüística Aplicada e Estudos de Comunicação.

O jornal é uma mídia influente na sociedade. Ele veicula textos que atendem a necessidades social e culturalmente estabelecidas, em formatos lingüísticos reconhecidos por uma comunidade discursiva. Os textos do jornal são aqui a matéria-prima e o principal objeto de análise. O fato de o conjunto de textos do jornal possibilitar a formação de subgrupos relativamente homogêneos, de acordo com vários critérios de seleção, criou, em conseqüência, um número expressivo de tipologias, por vezes conflitantes e com raras sistematizações na literatura. Este trabalho volta o foco para a descrição lingüística de um corpus, observando, em primeiro lugar, a variação lingüística encontrada nessa importante amostra de língua do modo escrito do português brasileiro atual. Em segundo lugar, procurou-se evidenciar de que maneira os gêneros convencionados na cultura criam padrões ou diferenças de caráter lingüístico. As questões de pesquisa, que deram concretude aos referidos objetivos, apontam três propósitos básicos: identificar quais características e categorias lingüísticas mais influenciaram na variação lingüística do corpus; que fatores ocorreram no corpus e como essas variáveis foram neles agrupadas; e, por fim, quais funções comunicativas, ou dimensões, representaram os fatores encontrados.

Para fundamentar teoricamente este trabalho, recorreu-se a autores da corrente neofirthiana, que postulam: a) a indissociação entre texto e contexto; b) o texto como unidade de significado; e c) a crença de que o significado está distribuído em vários níveis na linguagem. A língua é vista, assim, como um sistema probabilístico e a variação lingüística existente no corpus pode ser reconhecida como objeto de uma exploração empírica. A Análise Multidimensional de Biber, principal metodologia utilizada nesta pesquisa, propiciou a análise e interpretação dos dados obtidos em termos de dimensões, que cumprem um papel comunicativo e geral no corpus.

Com o intuito de subsidiar a questão dos gêneros encontrados no jornal, foram examinadas várias abordagens teóricas. Concluiu-se que a base teórica adequada para caracterizar a natureza desta pesquisa é uma combinação que alia a perspectiva geral de Swales (1990) a respeito do gênero (sem, porém, considerar os passos retóricos intratextuais propostos pelo autor, uma vez que nosso interesse está voltado para a variação intertextual) às considerações de Lee (2001) sobre as diferenças entre gênero e registro e à conceituação de tipo de texto, proposta por Biber (1988). Foi procurado mostrar aqui também o papel importante que ocuparam os estudos sobre gêneros do jornal no âmbito dos Estudos de Comunicação, a partir dos quais foi elaborada uma tipologia de gêneros para este trabalho.

Foram selecionadas sete edições diárias de um jornal de prestígio nacional, a “Folha de S.Paulo”, para reunir uma amostra representativa do corpus jornalístico. O ciclo de uma semana representa ao mesmo tempo para o jornal o que há de mais típico e rotineiro em suas páginas e, simultaneamente, revela particularidades de um jornalismo mais especializado, de publicação semanal. A estratégia de compor uma semana artificialmente construída foi simples e eficaz no sentido de diluir a repetição de palavras seguidamente utilizadas no noticiário diário.

Com o corpus de estudo selecionado, o passo seguinte foi etiquetá-lo morfossintaticamente, pelo programa VISL, de forma automática e com alta precisão. Ele gerou uma grande quantidade de etiquetas, que possibilitou uma extensa pesquisa

de seleção, segundo critérios teóricos e estatísticos. Para permitir a comparação entre os textos do corpus, de diferentes tamanhos, a frequência das variáveis selecionadas passou por um processo de padronização, que os igualou arbitrariamente como se todos os textos tivessem a extensão de 1.000 palavras.

O trabalho pioneiro de Biber (1988) e a sua contradita crítica, Lee (2000), foram as principais obras de referência utilizadas aqui para a aplicação da metodologia da Análise Multidimensional<sup>124</sup>. O procedimento estatístico chave, a Análise Fatorial, pelas dificuldades técnicas encontradas, foi considerado a fase mais crítica na pesquisa – daí a ênfase, no decorrer do trabalho, em explicitar didaticamente seus passos metodológicos e as operações que envolveram a operação do software SPSS, bastante conhecido entre estatísticos, mas pouco divulgado entre lingüistas.

Os dois fatores resultantes da Análise Fatorial realizada foram o produto de uma série de análises que abrangeram o uso, por parte deste pesquisador, de critérios objetivos e subjetivos. Os critérios objetivos estão relacionados às características dos procedimentos estatísticos, como a exigência de um valor mínimo de frequência e de comunalidade, à atribuição de escores médios a cada texto individual, de acordo com o peso que tem nas variáveis e nos fatores, e aos resultados dos vários testes estatísticos aplicados. Os critérios subjetivos existem em menor número no experimento, mas são decisivos: relacionam-se a decisões fundamentais, como a escolha do número final de fatores – ou, antes ainda, com a seleção das variáveis que compuseram a extração fatorial – e a sua interpretação em dimensões. Outro critério subjetivo a ser levado em conta é a classificação dos textos de acordo com o gênero, efetuada com uma precisão significativa, mas sem a pretensão de ter sido a classificação definitiva (cf. abaixo).

Os testes *post-hoc* fortaleceram a defesa da solução fatorial escolhida como meio diferenciador dos gêneros. O conseqüente desenvolvimento do constructo dimensão, que implica a interpretação dos fatores, foi realizado sob duas abordagens

---

<sup>124</sup> Berber Sardinha, o divulgador da Lingüística de Corpus de maior expressão em língua portuguesa, é também referência para as pesquisas feitas com AMD no país.

básicas: a partir do que se observou nos textos situados no fator – em seus dois pólos e em sua zona neutra –, de cujas forças atuantes são representantes (cf. 4.5), e em relação ao posicionamento dos gêneros, dispostos graficamente (cf. 4.8 e 4.9) por meio de seus escores médios nos fatores. Ambas as formas de analisar os fatores possibilitaram fundamentar a proposição final da pesquisa, ou seja, identificar a variação lingüística existente no corpus do jornal em termos de eixos dimensionais denominados “Narrativo versus Expositivo” e “Argumentativo versus Informativo”.

Por definição, as dimensões expressam características funcionais comunicativas e situacionais (BIBER, 1988). Pôde-se, a partir dessa suposição, ambicionar associar as dimensões obtidas neste trabalho com a teoria existente sobre os gêneros do jornal, no âmbito dos Estudos de Comunicação. As dimensões resultantes reforçaram a tese de Marques de Melo, síntese de uma leitura profunda da literatura, de que existe nos gêneros uma cisão básica entre opinião e informação (expressa pela Dimensão 2, o eixo “Argumentativo versus Informativo”). Concomitantemente, os dados obtidos também puderam corresponder, em alguma medida, à reflexão de Chaparro, já que apontaram para a existência de uma instância narrativa também atuante – a mais forte, ainda que menos diferenciadora, Dimensão 1, o eixo dimensional “Narrativo versus Expositivo”.

A exposição do experimento, sumarizada até aqui, permite agora que se exponham alguns comentários críticos acerca da AMD, como metodologia, e de alguns processos da pesquisa. Em relação às críticas que podem ser feitas à metodologia de Biber (1988), está em Lee (2000, p. 391-396) o elenco das mais importantes. Foram citadas aqui, expressa ou lateralmente, a crítica à representatividade do corpus (cf. 4.9), ao cálculo de escores médios dos fatores (cf. 4.7), à arbitrária escolha das variáveis (cf. acima), à diversidade exagerada de gêneros (cf. 3.3.3 e 3.4.1) e à desconsideração com as variáveis estatisticamente inaptas para a análise (cf. 3.2.2). Em um balanço geral, considerando que a pesquisa aqui conduzida convencionou-se, desde o princípio, exploratória, a AMD, apesar dos problemas mencionados, foi uma metodologia que se prestou à formulação perspicaz de hipóteses

que puderam ter a oportunidade de serem comparadas, em termos funcionais, com outras teorias.

Três críticas podem ser mencionadas em relação a outros processos: a) o etiquetador morfossintático VISL, apesar da apregoada precisão, apresentou erros de etiquetagem, principalmente quando uma frase do texto (como em títulos e subtítulos) terminava sem ponto final (cf. Anexo B). Deve-se admiti-los abertamente e dizer que, neste estudo, eles não foram corrigidos, medida que exigiria uma revisão manual impraticável. É de supor que os erros, porém, estejam distribuídos uniformemente pelo corpus, mas é preciso, em novos estudos, verificar se não há diferenças significativas de precisão entre os gêneros; b) a tipologia de gêneros efetuada, que resultou de uma pesquisa extensiva sobre os gêneros presentes no jornal (cf. 3.5.1), pode ser falha e alvo de crítica. Há diferenças entre a tipologia criada nesta pesquisa e as tipologias propostas em outros estudos, como Vélez (1985), Marques de Melo (1994), Chaparro (1997) e Bonini (2003), o que, por si só, é motivo para polêmica; c) a classificação manual pode estar a sujeita a erro porque foi efetuada por apenas um especialista. Apesar de considerá-la largamente consensual (85% de coincidência em relação a um grupo de controle; cf. 3.5), é necessário reconhecer que há diferença relevante de opiniões sobre determinados gêneros.

Além das críticas, é preciso apresentar o que pareceu ser um limite da pesquisa: sendo o corpus de estudo uma amostra muito pequena em relação ao universo lingüístico da imprensa, é preciso relativizar os resultados obtidos antes de afirmar que sejam realmente generalizações que representem dimensões comunicativas da imprensa escrita – e muito menos do modo escrito como um todo. Sempre haverá dúvidas, que merecem ou não ser minimizadas em razão de alguma vantagem teórica adquirida na análise. Em todo caso, nossa ambição de representar o corpus jornalístico foi muito menor que a de Biber, que quis abarcar todo o idioma inglês – e esse objeto mais preciso materialmente proporcionou mais foco e, possivelmente, mais argumentos favoráveis à realização de uma AMD. O outro limite do experimento é o pequeno destaque dado à análise qualitativa do corpus, ainda que a AMD, como

pesquisa quantitativa, inclua a observação de variáveis no nível microlingüístico; novos estudos de AMD deveriam desenvolver maior nível de diálogo entre as duas abordagens.

O presente trabalho, a esse respeito, espera ter aberto novos horizontes de pesquisa, em várias áreas de atuação. Ele possibilita imaginar AMDs de outros corpora especializados da língua portuguesa e variedades sociolingüísticas, entre outras abordagens e aplicações (os dados do Anexo E, com as estatísticas gerais de frequência dos gêneros, poderiam ser utilizados no desenvolvimento de materiais pedagógicos, por exemplo). A pesquisa aqui descrita também pode ser reorganizada em termos de outras classificações de gêneros, tipologias e verificação de divergências cognitivas<sup>125</sup>, no sentido de buscar um maior grau de refinamento e consistência teórica. Adicionalmente, poderiam ser estudados os diferentes tópicos do jornal, seus assuntos, editorias e cadernos, morfologicamente, além de abrir a pesquisa à exploração de outras variáveis, como sugerem os trabalhos de Shimazumi (1998), Pacheco de Oliveira (1997), Conde (2002) e Shergue (2003), que consideraram a inclusão de etiquetas de caráter funcional e extralingüístico. Por fim, inspirado em Biber (1993), espera-se que seja possível subsidiar a construção de um modelo de classificação automática de gêneros jornalísticos escritos em português, de cunho probabilístico, capaz de ser aplicada sistematicamente em arquivos de textos não anotados armazenados em formato digital.

Conforme dito na Introdução, um dos problemas do ensino de língua portuguesa na Educação Fundamental é a sua forma assistemática de abordar o texto jornalístico no tocante ao gênero, freqüentemente acompanhado de certa frouxidão terminológica, apesar de iniciativas padronizadoras como os PCNs. Desse modo, e extrapolando o terreno do desenvolvimento de material didático básico, é possível vislumbrar diversas aplicações pedagógicas para o modelo que aqui foi apresentado,

---

<sup>125</sup> Cf. Bonini (2002) é uma interessante abordagem cognitiva que analisa o reconhecimento dos gêneros pela comunidade de jornalistas, embora a partir de fragmentos textuais, não de textos completos.

em conformidade com os melhores princípios da Lingüística Aplicada: a) na formação contínua de professores de Ensino Fundamental; b) como ferramenta de análise do texto jornalístico, no Ensino Médio e Superior; c) auxiliando a produção de determinados gêneros do texto jornalístico, em disciplinas da especialização universitária; d) em programas de formação de quadros profissionais das empresas jornalísticas; e) por fim, divulgar a metodologia da AMD, que pode ser utilizada por mais pesquisadores em Lingüística de Corpus.

Este estudo considera que as dimensões do corpus jornalístico discutidas aqui puderam contribuir com a literatura que busca descrever por meio de um sistema ou estrutura os diversos tipos de gêneros encontrados no jornal. Isso demonstra que a abordagem, desenvolvida na Lingüística de Corpus, pode contribuir para questões ligadas a gêneros na Lingüística Aplicada e em áreas como a de Estudos de Comunicação. Espera-se igualmente que todo o esforço em cumprir a AMD com um alto nível de rigor metodológico possa sugerir caminhos para que os futuros estudos que a utilizem tornem-se mais cuidadosos e menos generalistas.

Finalmente, vale desejar que este trabalho, de certa forma pioneiro<sup>126</sup> por realizar uma Análise Multidimensional em um corpus exclusivamente do português do Brasil, expanda também o conhecimento do jornalista a respeito de suas práticas lingüísticas, como autor principal da imprensa diária escrita contemporânea.

---

<sup>126</sup> Até onde a literatura consultada pôde corroborar.

## REFERÊNCIAS

- AARTS, B. Intuition-based and observation-based grammars. In: AIJIMER, K.; ALTENBERG, B. (Orgs.). **English Corpus Linguistics** – Studies in Honour of Jan Svartvik. London: Longman, 1991.
- AKINNASO, F. On the differences between spoken and written language. **Language and Speech**, n. 25, p. 97-125, 1982.
- ALUÍSIO, S. M. et al. The Lacio-Web Project: overview and issues in Brazilian Portuguese *corpora* creation. In: **Corpus Linguistics**, Lancaster, v. 16, p. 14-21, 2003.
- ARISTÓTELES. **Arte retórica e arte poética**. Rio de Janeiro: TecnoPrint, 1991.
- BAKHTIN, M. Os gêneros do discurso. In: \_\_\_\_\_. **Estética da criação verbal**. 3. ed. São Paulo: Martins Fontes, 2000.
- BECHARA, E. **Moderna gramática portuguesa**. 37. ed. Rio de Janeiro: Lucerna, 1999.
- BELL, A. **The Language of News Media**. Oxford: Basil Blackwell, 1991.
- BERBER SARDINHA, A. P. Usando o WordSmith Tools na investigação da linguagem. **DIRECT Papers**, São Paulo e Liverpool, n. 40, 1999.
- \_\_\_\_\_. **Linguística de Corpus: histórico e problemática**. **D.E.L.T.A.**, São Paulo, n. 16, v. 2, p. 323-367, 2000 (a).
- \_\_\_\_\_. **Análise Multidimensional**. **D.E.L.T.A.**, São Paulo, n. 16, v. 1, p. 99-127, 2000 (b).
- \_\_\_\_\_. **Linguística de corpus**. Barueri: Manole, 2004 (a).
- \_\_\_\_\_. Informatividade, interatividade e narratividade na reunião de negócios – Análise Multidimensional e palavras-chave. **DIRECT Papers**, São Paulo e Liverpool, n. 52, 2004 (b).
- \_\_\_\_\_. (org.). **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005.
- BHATIA, V. **Analysing Genre: Language Use in Professional Settings**. London:

- Longman, 1993.
- BIBER, D. Investigating macroscopic textual variation through multifeature / multidimensional analyses. **Linguistics**, n. 23, p. 337-360, 1985.
- \_\_\_\_\_. Spoken and written textual dimensions in English: resolving the contradictory findings. **Language**, 62, p. 384-414, 1986.
- \_\_\_\_\_. **Variation Across Speech and Writing**. Cambridge: Cambridge University Press, 1988.
- \_\_\_\_\_. On the complexity of discourse complexity: A multidimensional analysis. **Discourse Processes**, n. 15, p. 133-163, 1992.
- \_\_\_\_\_. Using Register-Diversified Corpora for General Language Studies. **Computational Linguistics**, v. 19, n. 2, p. 219-241, 1993.
- \_\_\_\_\_. Representativeness in Corpus Design. In: ZAMPOLLI, A.; CALZOLARI, N.; PALMER, M. (Ed.). **Current Issues in Computational Linguistics: In Honour of Don Walker**. *Linguistica Computazionale IX.X*. Pisa e Dordrecht: Giardini e Kluwer Academic Publishers, 1994.
- \_\_\_\_\_. **Dimensions of Register Variation: A Cross-linguistic Comparison**. Cambridge: Cambridge University Press, 1995.
- BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus Linguistics: Investigating Language Structure and Use**. Cambridge: Cambridge University Press, 1998.
- BIBER, D. et al. **Longman Grammar of Spoken and Written English**. London: Longman, 1999.
- BONINI, A. **Gêneros textuais e cognição**. Florianópolis: Insular, 2002.
- \_\_\_\_\_. Página, seção, notícia, nota: critérios de identificação do gênero no jornal. **Boletim da Associação Brasileira de Lingüística (ABRALIN)**, 26. Fortaleza, v. 1, p. 616-618, 2003.
- \_\_\_\_\_. Metodologias para o estudo dos gêneros textuais: como estudar o encaixe dos gêneros no jornal?. In: MAGALHÃES CAVALCANTE, M.; PAIVA BRITO, M. A. (Org.). **Gêneros textuais e referenciação**. Fortaleza, v. 1, 2004.
- BRAIT, B. (org.). **Bakhtin, dialogismo e construção do sentido**. Campinas:

Unicamp, 1997.

BRASIL/SEF/MEC. **Parâmetros Curriculares Nacionais – 5ª a 8ª série do Ensino Fundamental**. Brasília: SEF/MEC, (1997 e 1998).

CARROLL, J. B. Vectors of prose style. In: SEBEEK, T. A. (Org.), **Style in Language**. Cambridge, Mass.: MIT Press, 1960. p. 283-292.

CHAFE, W. Integration and involvement in speaking, writing, and oral literature. In: TANNEN, D. (Ed.), **Spoken and Written Language: Exploring Orality and Literacy**. Norwood, N.J.: Ablex, 1982. p. 35-54.

CHAFE, W.; DANIELEWICZ, J. Properties of spoken and written language. In: HOROWITZ, R.; SAMUELS, S. J. (Org.), **Comprehending Oral and Written Language**. San Diego: Academic Press, 1987. p. 83-113.

CHAPARRO, M. C. **Jornalismo, discurso em dois gêneros**. São Paulo, 1997. Tese (Livre docência de Jornalismo e Editoração) – Escola de Comunicações e Artes, USP, São Paulo.

\_\_\_\_\_. **Sotaques d'aquém e d'além mar**. Santarém: Jortejo, 1998.

CHARAUDEAU, P.; MAINGUENEAU, D. **Dicionário de análise do discurso**. São Paulo: Contexto, 2004.

COLLINS COBUILD ENGLISH GUIDES SERIES. **Prepositions** (v. 1). London: HarperCollins, 1991.

CONDE, H. M. A. **Escolhas lexicais em composições de alunos avançados de inglês originários de instituições de ensino bilíngües e monolíngües** - Um estudo multidimensional baseado em corpus. São Paulo, 2002. Dissertação (Mestrado em Lingüística Aplicada e Estudos da Linguagem) – LAEL, PUC/SP, São Paulo.

CUNHA, C. F. **Gramática da língua portuguesa**. 7. ed. Rio de Janeiro: FENAME, 1980.

DE BEAUGRANDE, R. **Linguistic Theory: The Discourse of Fundamental Works**. London: Longman, 1991.

\_\_\_\_\_. The 'pragmatics' of doing language science: The 'warrant' for large-corpus

- linguistics. **Journal of Pragmatics**, n. 25, p. 503-535, 1996.
- DUCROT, O.; TODOROV, T. **Dicionário das ciências da linguagem**. 2. ed. Lisboa: Dom Quixote, 1974.
- EAGLES. **Preliminary recommendations on corpus typology**. EAG-TCWG-CTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale, 1996 (a). [Disponível no site <http://www.ilc.cnr.it/EAGLES96/corpusyp/corpusyp.html>; consultado em 21/08/2005]
- \_\_\_\_\_. **Preliminary recommendations on text typology**. EAG-TCWG-TTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale, 1996 (b). [Disponível no site <http://www.ilc.cnr.it/EAGLES/texttyp/texttyp.html>; consultado em 21/08/2005]
- EDITORA ABRIL. **Manual de estilo**. Rio de Janeiro: Nova Fronteira, 1990.
- EGGINS, S.; MARTIN, J. R. Genres and registers of discourse. In: VAN DIJK, T. A. (Ed.). **Discourse as Structure and Process – Discourse Studies: A Multidisciplinary Introduction**, v. 1. London, Thousand Oaks, New Delhi: Sage Publications, 1997.
- FAIRCLOUGH, N. **Media Discourse**. London: Edward Arnold, 1995.
- FIRTH, J. R. The technique of Semantics. **Transactions of the Philological Society**, p. 36-72, 1935.
- \_\_\_\_\_. **Papers in Linguistics – 1934-1951**. Oxford: Oxford University Press, 1957.
- \_\_\_\_\_. **Selected Papers of J. R. Firth – 1952-1959**. Palmer, F. R. (Org.). London: Longman, 1968.
- FOLHA DE S.PAULO. **Manual geral da redação**. São Paulo, 1984.
- \_\_\_\_\_. \_\_\_\_\_. 2. ed. São Paulo, 1987.
- \_\_\_\_\_. **Novo manual da redação**. São Paulo, 1992.
- \_\_\_\_\_. **Manual da redação**. São Paulo: Publifolha, 2001.
- FOWLER, R. **Language in the News – Discourse and Ideology in the Press**. London: Routledge, 1991.

- GRABE, W. Contrastive rhetoric and text type research. In: CONNOR U.; KAPLAN, R. (Ed.). **Writing Across Languages: Analysis of L2 Texts**. Reading, MA: Adisson-Wesley, 1987. p. 113-137.
- HALLIDAY, M. A. K. **An Introduction to Functional Grammar**. 2. ed. London: Edward Arnold, 1994.
- HALLIDAY, M. A. K.; HASAN, R. **Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective**. Oxford: Oxford University Press, 1989.
- HOEY, M. **Patterns of Lexis in Text**. Oxford: Oxford University Press, 1991.
- HUNSTON, S. **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press, 2002.
- HUNSTON, S.; FRANCIS, G. **Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English**. Amsterdam e Philadelphia: John Benjamins, 1999.
- JONES, R. L.; CARTER JR., R. E. Some procedures for estimating 'news hole' in content analysis. **Public Opinion Quarterly**, n. 23, p. 399-403, 1959.
- KARLSSON, F. et al. **Constraint grammar: A language-independent system for parsing unrestricted texts**. Berlin: Mouton de Gruyter, 1995.
- KAUFFMANN, C. H. Variação de palavras-chave dentre gêneros de um corpus jornalístico. **Intercâmbio**, São Paulo, v. XIV, 2005.
- \_\_\_\_\_. Gêneros presentes na imprensa diária escrita no Brasil: um estudo baseado em corpus. Artigo em fase de elaboração.
- KENNEDY, G. D. **An Introduction to Corpus Linguistics**. London: Longman, 1998.
- LABOV, W. **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press, 1972.
- LAGE, N. **Ideologia e técnica da notícia**. 2. ed. Petrópolis: Editora Vozes, 1998.
- LEE, D. Y. W. **Modelling variation in spoken and written language: the Multi-Dimensional Approach revisited**. Lancaster, 2000. Tese (Doutoramento em Linguística) – Dept. of Linguistics and Modern English Language, Lancaster

University.

- \_\_\_\_\_. Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle. **Language Learning & Technology**, n. 3, v. 5, p. 37-72, 2001.
- LEECH, G. The state of the art in corpus linguistics. In: AIJMER, K.; ALTENBERG, B. (Org.). **English Corpus Linguistics – Studies in Honour of Jan Svartvik**. London: Longman, 1991.
- \_\_\_\_\_. Corpora and theories of linguistic performance. In: SVARTVIK, J. (Org.). **Directions in Corpus Linguistics**. Proceedings of Nobel Symposium, 82. Stockholm, 4-8 August 1991, p. 105-122. Berlin: Mouton de Gruyter, 1992.
- LEECH, G.; SVARTVIK, J. **A Communicative Grammar of English**. London: Longman, 1980.
- LONGMAN CONTEMPORARY ENGLISH DICTIONARY. 2. ed. London: Longman, 1987.
- LUX, P.; GRABE, W. Multivariate approaches to contrastive rhetoric. **Linguas Modernas**, n. 18, p. 133-160, 1991.
- MALINOWSKI, B. The problem of meaning in primitive languages. In: **The Meaning of Meaning**. London: Kegan Paul, 1923.
- MARQUES DE MELO, J. (org.). **Gêneros jornalísticos na Folha de S.Paulo**. São Paulo: FTD/IPCJE, 1987.
- \_\_\_\_\_. **A opinião no jornalismo brasileiro**. 2. ed. Petrópolis: Vozes, 1994.
- \_\_\_\_\_. (org.). Gêneros e formatos na comunicação massiva periodística: um estudo do jornal “Folha de S.Paulo” e da revista “Veja”. **Anais do Congresso Brasileiro de Ciências de Comunicação (Intercom)**, 21. Recife, 1998 (a).
- \_\_\_\_\_. (org.). **Identidade da imprensa brasileira no final do século**. São Bernardo: UNESCO/UMESP, 1998 (b).
- MONAGHAN, J. **The Neo-Firthian Tradition and its Contribution to General Linguistics**. Tübingen: Max Niemeyer Verlag, 1979.
- McENERY, T.; WILSON, J. **Corpus Linguistics**. Edinburgh: Edinburgh

- University Press, 1996.
- NABANTINO RAMOS, J. **Jornalismo** – dicionário enciclopédico. São Paulo: Ibrasa, 1970.
- NEVES, M. H. M. **Gramática de usos do português**. São Paulo: Editora Unesp, 2000.
- NOBLAT, R. **A arte de fazer um jornal diário**. 2. ed. São Paulo: Contexto, 2002.
- OCHS, E. Planned and unplanned discourse. In: T. GIVÓN (org.), **Discourse and Syntax**, p. 51-80. New York: Academic Press, 1979.
- O ESTADO DE S. PAULO. **Manual de redação e estilo**. MARTINS, E. (Ed.). São Paulo: O Estado de S. Paulo, 1990.
- OXFORD COLLOCATIONS DICTIONARY FOR STUDENTS OF ENGLISH. Oxford: Oxford University Press, 2002.
- PACHECO DE OLIVEIRA, L. **Variação intercultural na escrita**: contrastes multidimensionais em inglês e português. São Paulo, 1997. Tese (Doutoramento em Linguística Aplicada e Ensino de Línguas) – LAEL, PUC/SP.
- QUIRK, R. S. et al. **A Comprehensive Grammar of the English Language**. Harlow: Longman, 1985.
- RABAÇA, C. A.; BARBOSA, G. **Dicionário de comunicação**. 3. ed. São Paulo: Ática, 1998.
- REID, J. Responding to different topic types: a quantitative analysis from a contrastive rhetoric perspective. In: KROLL, B. (Ed.). **Second language writing**: research insights for the classroom. New York: Cambridge University Press, 1990.
- RÓNAI, P. Um gênero brasileiro; a crônica. In: HOWER, A.; PRETO-RODAS, R. (Org.). **Crônicas brasileiras**. Center for Latin American Studies, University of Florida, 1971.
- SANCHEZ, A. Definición e historia de los corpus. In: SANCHEZ, A. et al. (Org.) **CUMBRE – Corpus Lingüístico del Español Contemporáneo**. Madrid: SGEL, p. 7-24, 1995.
- SANTOS, V. B. M. P. As características léxico-gramaticais de um manual de gestão a

- partir da Análise Multidimensional de Biber. **The ESpecialist**, v. 24, n. 2, p. 201-227, 2003.
- SCOTT, M. PC Analysis of key words – and key key words. **System**, n. 25, p. 233-245, 1997.
- SHERGUE, O. **Dimensão de variação no discurso médico acadêmico**: o artigo de pesquisa e a apresentação de trabalhos científicos em congressos. São Paulo, 2003. Dissertação (Mestrado em Lingüística Aplicada e Estudos da Linguagem) – LAEL, PUC/SP.
- SINCLAIR, J. **A Course in Spoken English: Grammar**. London: Oxford University Press, 1972.
- \_\_\_\_\_. **Looking Up: An Account of the COBUILD Project in Lexical Computing**. London: Collins, 1987.
- \_\_\_\_\_. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.
- SHIMAZUMI, M. Investigating EFL writing: A Multidimensional analysis. In: CONVENÇÃO BRAZ-TESOL, 6. Recife, 13-16 de julho de 1998. Comunicação.
- STANTON, C. **What is This Thing Called Genre?** Nottingham: University of Nottingham, 1996.
- STUBBS, M. British traditions in text analysis. In: BAKER, M.; FRANCIS, G.; TOGNINI-BONELLI, E. (Org.), **Text and Technology** – In Honour of John Sinclair. Philadelphia e Amsterdam: John Benjamins, 1993.
- \_\_\_\_\_. **Text and Corpus Analysis**. Oxford: Blackwell, 1996.
- \_\_\_\_\_. **Words and Phrases**. Oxford: Blackwell, 2001.
- SWALES, J. **Genre Analysis: English in Academic and Research Settings**. Cambridge: Cambridge University Press, 1990.
- \_\_\_\_\_. Re-thinking genre: another look at discourse community effects. In: **Re-Thinking Genre Colloquium**. Ottawa: Carleton University, 1992.
- TODOROV, T. **Os gêneros do discurso**. São Paulo: Martins Fontes, 1980.
- TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam e Philadelphia:

- John Benjamins, 2001.
- VÉLEZ, M. M. L. **As folhas do diário**. São Paulo, 1985. Dissertação (Mestrado em Jornalismo e Editoração) – Escola de Comunicações e Artes, USP.
- VAN DIJK, T. A. Critical News Analysis. **Critical News**. Amsterdam e New York: Rodopi, 1989. p. 103-126.
- \_\_\_\_\_. **La notícia como discurso**. Barcelona/Buenos Aires: Paidós, 1990.
- VIAN JR., O. **Conceito de gênero e análise de textos de vídeos institucionais**. São Paulo, 1997. Dissertação (Mestrado em Lingüística Aplicada e Ensino de Línguas) – LAEL, PUC/SP, São Paulo.
- \_\_\_\_\_. Sobre o conceito de gêneros do discurso: diálogos entre Bakhtin e a lingüística sistêmico-funcional. In: BRAIT, B. (Org.). **Estudos Enunciativos no Brasil – Histórias e Perspectivas**. Campinas: Pontes/FAPESP, 2001.
- VIVALDI, G. M. **Gêneros Periodísticos**. Madrid: Paraninfo, 1973.
- WEST, M. **A General Service List of English Words**. London: Longman, 1953.
- WESTIN, I. **Language Change in English Newspaper Editorials**. Amsterdam e New York: Rodopi, 2002. (Language and Computers Series, v. 44)
- ZERO HORA. **Manual de ética, redação e estilo**. Porto Alegre: L&PM, 1994.

## LISTA DE ANEXOS

ANEXO A - Lista de Etiquetas do VISL para o Português.....	139
ANEXO B - Exemplo de Texto Etiquetado pelo VISL.....	145
ANEXO C - Estatística Descritiva dos Fatores, por Gênero.....	148
ANEXO D - Textos “típicos” dos gêneros.....	149
ANEXO E - Estatística Descritiva de Frequência das Variáveis, por Gênero.....	167
ANEXO F - Versão Integral do Texto 6.....	181

## ANEXO A - Lista de Etiquetas do VISL para o Português

São apresentadas a seguir a lista de etiquetas produzidas pelo “*parser*” VISL pesquisadas neste trabalho e, como referência, a relação geral das etiquetas mais utilizadas por esse sistema na versão para a língua portuguesa.

TABELA 12 - ETIQUETAS DO VISL PESQUISADAS ORIGINALMENTE NO TRABALHO

Variável
ADJ (adjetivos)
@>N (prenominal adjunct)
@N< (postnominal adjunct)
PR 3S IND (3ª pessoa singular, presente do indicativo)
@PASS (voz passiva)
@FAUX (verbo auxiliar finito)
@IMV (verbo principal não-finito)
@<ADVL (adjunto adverbial)
Standardised type / token ratio (base = 50 tokens)
ADV (advérbio)
GER (gerúndio)
<rel> (pronome relativo)
“ser” (lema SER)
<SUBJ (sujeito pós-posicionado)
COND (condicional)
@CO (conjunção coordenada)
@SUB (conjunção subordinada)
@<ACC (objeto direto)
@ACC> (acusativo, objeto direto)
PR 1S IND (1ª pessoa singular, presente do indicativo)
PS 3S IND (3ª pessoa singular, pret. perfeito do indicativo)
@IAUX (verbo auxiliar não-finito)
@FMV (verbo principal finito)
@ADVL (nú principal adverbial)
@ADVL> (adjunto adverbial)
N (substantivo)
V (verbo)
DET (determinante)
KC (conjunção coordenativa)
KS (conjunção subordinativa)
PROP (nome próprio)
ARTD (artigo definido)
ARTI (artigo indefinido)
IND (modo indicativo)
PR (tempo verbal presente)
PS (tempo verbal pretérito perfeito)
INF (infinitivo)
PCP (particípio passado)
<dem> (pronome demonstrativo)

<poss> (pronome possessivo)
[estar] (lema ESTAR)
[dizer] (lema DIZER)
[ter] (lema TER)
[haver] (lema HAVER)
[fazer] (lema FAZER)
[poder] (lema PODER)
3S (3ª pessoa singular)
3P (3ª pessoa plural)
SUBJ> (sujeito pré-posicionado)
<quant> (pronome quantificador)
NUM (numerais cardinais)
<sam-> (contração morfológica)
<*1> (aspa esquerda)
PF / MQF (pret. Perfeito / Mais que Perfeito do indicativo)
IMPF 3S IND (3ª pessoa sing., pret. imperfeito do indicativo)
IMPF (tempo verbal pretérito imperfeito)
1S (1ª pessoa singular)
PERS (pronome pessoal)
SUBJ (modo subjuntivo)
[afirmar] (lema AFIRMAR)
[dar] (lema DAR)
[dever] (lema DEVER)
[querer] (lema QUERER)
1P (1ª pessoa plural)
PS 1S IND (1ª pessoa singular, pret. perfeito do indicativo)
IMPF 3P IND (3ª pessoa pl., pret. imperfeito do indicativo)
vpublic (lemas dizer, afirmar, falar, declarar, comentar)
Não (advérbio de negação)
PRP (preposição)
ponfinal (ponto final)
virgula (vírgula)
@# (qualquer tipo de “subclause”)
KOMP_ (comparativo)

A lista a seguir foi extraída tal como apresentada na página do VISL na internet<sup>127</sup> e traz as principais etiquetas geradas pela classificação morfossintática da Gramática Construtiva na língua portuguesa:

#### WORD CLASS TAGS

N Nouns

PROP Proper nouns (names)

SPEC Specifiers (defined as non-inflecting pronouns, that can't be used as prenominals):

---

<sup>127</sup> Disponível em <http://visl.hum.sdu.dk/visl/pt/portsymbol.html>. Consultado em 21/08/2005.

e.g. indefinite pronouns, nominal quantifiers, nominal relatives  
 DET Determiners (defined as inflecting pronouns, that can be used as pronominals):  
 e.g. articles, attributive quantifiers  
 PERS Personal pronouns (defined as person-inflecting pronouns)  
 ADJ Adjectives (including ordinals, excluding participles which are tagged V PCP)  
 ADV Adverbs (both ‘primary’ adverbs and derived adverbs ending in ‘-mente’)  
 V Verbs (full verbs, auxiliaries)  
 NUM Numerals (cardinals)  
 KS Subordinating conjunctions  
 KC Coordinating conjunctions  
 IN Interjections  
 EC Hyphen-separated prefix (“elemento composto”, category being phased out)

#### INFLECTION TAGS

Gender: M (male), F (female), M/F [for: N’, PROP’, SPEC’, DET, PERS, ADJ, V PCP, NUM]

Number: S (singular), P (plural), S/P [for: N, PROP’, SPEC’, DET, PERS, ADJ, V PCP, V VFIN, INF, NUM]

Case: NOM (nominative), ACC (accusative), DAT (dative), PIV (prepositional), ACC/DAT, NOM/PIV [for: PERS]

Person: 1 (first person), 2 (second person), 3 (third person), 1S, 1P, 2S, 2P, 3S, 3P, 1/3S, 0/1/3S [for: PERS, V VFIN, V INF]

Tense: PR (present tense), IMPF (imperfecto), PS (perfeito simples), MQP (mais-que-perfeito), FUT (futuro), COND (condicional) [for: V VFIN]

Mood: IND (indicative), SUBJ (subjunctive), IMP (imperative) [for: V VFIN]

Finiteness: VFIN (finite verb), INF (infinitive), PCP (participle), GER (gerund) [for: V]

#### SYNTACTIC TAGS

@SUBJ> @<SUBJ subject

@ACC> @<ACC accusative (direct) object

@DAT> @<DAT dative object (only pronominal)

@PIV> @<PIV prepositional (indirect) object

@ADVS> @<ADVS adverbial object (place, time, duration, quantity)

@ADVO> @<ADVO adverbial object (object-related)

@SC> @<SC subject predicative

@OC> @<OC object predicative

@ADVL> @<ADVL adverbial

@PASS> @<PASS agent of passive

(All above clause arguments [@SUBJ, @ACC, @DAT, @PIV, @ADVS, @ADVO, @SC, @OC, @PASS] and the adverbial complements [@ADVL] attach to the nearest main verb to the left [<] or right [>].)

@ADVL ‘free’ adverbial phrase (in non-sentence expression)

- @NPHR 'free' noun phrase (in non-sentence expression without verbs)
- @VOK 'vocative' (e.g. 'free' addressing proper noun in direct speech)
- @>N prenominal adjunct (attaches to the nearest NP-head to the right, that is not an adnominal itself)
- @N< postnominal adjunct (attaches to the nearest NP-head to the left, that is not an adnominal itself)
- @N<PRED postnominal (in-group predicative) or predicate in small clause introduced by 'com/sem' (rare, e.g. 'com a mão na bolsa', 'sem o pai ajudando, não conseguiu')
- @APP identifying apposition (always after NP + komma)
- @>A prepositioned adverbial adjunct  
(attaches to the nearest ADJ/PCP/ADV or attributive used N to the right)
- @A< postpositioned adverbial adjunct (rare, e.g. 'caro demais') or dependent/argument of attributive participle (with function tag attached, e.g. @A<ADVL or @A<SC)
- @PRED> 'forward' free predicative (refers to the following @SUBJ, even when this is incorporated in the VP)
- @<PRED 'backward' free predicative (refers to the nearest NP-head to the left, or to the nearest @SUBJ to the left)
- @P< argument of preposition
- @S< sentence anaphor ('não venceu o que muito o contrariou')
- @FAUX finite auxiliary (cp. @#ICL-AUX<)
- @FMV finite main verb
- @IAUX infinite auxiliary (cp. @#ICL-AUX<)
- @IMV infinite main verb
- @PRT-AUX< verb chain particle (preposition or "que" after auxiliary)
- @CO coordinating conjunction
- @SUB subordinating conjunction
- @KOMP< argument of comparative (e.g. "do que" referring to 'melhor')
- @COM direct comparator without preceding comparative
- @PRD role predicator (e.g. "work as", "function as")
- @FOC> @<FOC focus marker ("gosta é de peixe.")
- @TOP topic constituent ("Esse negócio, não gosto dele.")
- @#FS- finite subclause (combines with clausal role and intraclausal word tag, e.g. @#FS-<ACC @SUB for "não acredito que seja verdade")
- @#ICL- infinite subclause (combines with clausal role and intraclausal word tag, e.g. @#ICL-SUBJ> @IMV in "consertar um relógio não é fácil")
- @#ICL-AUX< argument verb in verb chain, refers to preceding auxiliary (the verb chain sequence @FAUX - @#ICL-AUX< is used, where both verbs have the same subject, @FMV - @#ICL-<ACC is used where the subjects are different)
- @#AS- averbal (i.e. verb-less) subclause (combines with clausal role and intraclausal word tag, e.g. @#AS-<ADVL @ADVL> in "ajudou onde possível")
- @AS< argument of complementiser in averbal subclause

## SECONDARY TAGS

## Subclass tags

- <artd> definite article (DET)
- <arti> indefinite article (DET)
- <quant> quantifier pronoun (DET: <quant1>, <quant2>, <quant3>, SPEC: <quant0>) or intensifier adverb
- <dem> demonstrative pronoun (DET: <dem>, SPEC: <dem0>)
- <poss> possessive pronoun (DET)
- <refl> reflexive personal pronoun (“se” PERS ACC, “si” PERS PIV)
- <si> reflexive use of 3. person possessive
- <reci> reciprocal use of reflexive pronoun (= “um ao outro”)
- <coll> collective reflexive (“reunir-se”, “associar-se”)
- <diff> differentiator (DET) (e.g. “e outros temas”, “a mesma diferença”)
- <ident> identator (DET) (e.g. “o próprio usuário”, “a si mesmo”)
- <rel> relative pronoun (DET, SPEC)
- <interr> interrogative pronoun (DET, SPEC)
- <post-det> typically located as post-determiner (DET @N<)
- <post-attr> typically post-positioned adjective (ADJ @N<)
- <ante-attr> typically pre-positioned adjective (ADJ @>N)
- <adv> can be used adverbially (ADJ @ADVL)
- <ks> relative adverb used like a subordinating conjunction
- <kc> conjunctive adverb (pois, entretanto)
- <det> determiner usage/inflection of adverb (“ela estava toda nua.”)
- <foc> focus marker adverb (also forms of “ser”)
- <prp> relative adverb used like a preposition
- <KOMP> <igual> “equalling” comparative (ADJ, ADV) (e.g. “tanto”, “tão”)
- <KOMP> <corr> correlating comparative (ADJ, ADV) (e.g. “mais velho”, “melhor”)
- <komp> <igual> “equalling” particle referring to comparative (e.g. “como”, “quanto”)
- <komp> <corr> “correlating” particle referring to comparative (e.g. “do=que”)
- <SUP> superlative
- <setop> operational adverb (eg. “não”, “nunca”, “já”, “mais” in “não mais”)
- <dei> discourse deictics (e.g. “aqui”, “ontem”)
- <card > cardinal (NUM)
- <NUM-ord> ordinal (ADJ)
- <NUM-fract> fraction-numeral (N)
- <cif> cipher (<card> NUM, <NUM-ord> ADJ)
- <sam-> first part of morphologically fused word pair (“de” in “dele”)
- <-sam> last part of morphologically fused word pair (“ele” in “dele”)
- <\*> 1. letter capitalized
- <\*1> left quote attached
- <\*2> right quote attached
- <hyfen> hyphenated word
- <ABBR> abbreviation

<prop> noun, adjective etc. used as name (upper case initial in mid-sentence)

<n> adjective or participle used as a “noun”, typically as head of a nominal phrase

<fmc> finite main clause heading verb

<co-acc>, <co-advl>, <co-app>, <co-dat>, <co-fmc>, <co-ger>, <co-inf>, <co-oc>, <co-pcv>, <co-postad>, <co-postnom>, <co-pred>, <co-prenom>, <co-prparg>, <co-sc>, <co-subj>, <co-vfin> co-ordinator tags indicating what is co-ordinated: @ACC, @ADVL, @APP, @DAT, main clauses, GER, INF, @OC, PCP-@IMV, @A<, @N<, @PRED, @>N, @P<, @SC, @SUBJ, VFIN (ordered list matching the <co-...> tags)

## ANEXO B - Exemplo de Texto Etiquetado pelo VISL

TEXTO 7 - AMOSTRA DE TEXTO ETIQUETADO PELO VISL (TEXTO: CP1\_0002.TXT)

S.=Paulo [S.=Paulo] PROP M S @SUBJ>  
 sobe [subir] <fmc> V PR 3S IND VFIN @FMV  
 para [para] PRP @<ADVS  
 1º [1º] <NUM-ord> ADJ M S @>N  
 lugar [lugar] N M S @P<  
 em [em] PRP @<ADVL  
 roubos [roubo] N M P @P<  
 em [em] <sam-> PRP @<ADVL  
 o [o] <-sam> <artd> DET M S @>N  
 Estado=A [Estado=A] PROP M S @P<  
 cidade [cidade] N F S @<ACC  
 de [de] PRP @N<  
 São=Paulo [São=Paulo] PROP M S @P<  
 assumiu [assumir] <fmc> V PS 3S IND VFIN @FMV  
 o [o] <artd> DET M S @>N  
 topo [topo] N M S @<ACC  
 de [de] <sam-> PRP @N<  
 o [o] <-sam> <artd> DET M S @>N  
 ranking [ranking] N M S @P<  
 de [de] <sam-> PRP @N<  
 os [o] <-sam> <artd> DET M P @>N  
 municípios [município] N M P @P<  
 paulistas [paulista] ADJ M P @N<  
 em [em] <sam-> PRP @N<  
 a [o] <-sam> <artd> DET F S @>N  
 relação [relação] N F S @P<  
 roubos [roubo] N M P @N<  
 / [/] <co-acc> KC @CO  
 população [população] N F S @<ACC  
 ,  
 sem "sem" PRP @<ADVL  
 contar [contar] V INF @IMV @#ICL-P<  
 roubos [roubo] N M P @<ACC  
 de [de] PRP @N<  
 veículos [veículo] N M P @P<  
 ,  
 pela=primeira=vez "pela=primeira=vez" ADV @<ADVL  
 em [em] PRP @<ADVL

quatro [quatro] <card> NUM M P @>N  
 anos [ano] N M P @P<  
 .  
 em "em" PRP @ADVL>  
 2002 [2002] <card> NUM M/F S @P<  
 ,  
 a "o" <artd> DET F S @>N  
 capital [capital] N F S @SUBJ>  
 paulista [paulista] ADJ F S @N<  
 teve [ter] <fmc> V PS 3S IND VFIN @FMV  
 1.139,52 [1.139,52] <card> NUM M P @>N  
 roubos [roubo] N M P @<ACC  
 por [por] PRP @<PIV  
 100 [100] <card> NUM M/F P @>N  
 mil [mil] <card> NUM M/F P @>N  
 habitantes [habitante] N M/F P @P<  
 ,  
 índice "índice" N M S @N<PRED  
 6,73 [6,73] <card> NUM M P @>N  
 % [%] N M P @APP  
 superior [alto] <KOMP> <SUP> ADJ M S @N<  
 a [a] <sam-> PRP @A<  
 o [o] <-sam> DET M S @P<  
 de [de] PRP @N<  
 2001 [2001] <card> NUM M/F P @P<  
 .  
 o "o" <artd> DET M S @>N  
 ranking [ranking] N M S @SUBJ>  
 foi [ser] <fmc> V PS 3S IND VFIN @FAUX  
 feito [fazer] V PCP M S @IMV @#ICL-AUX<  
 com=base=em [com=base=em] PRP @<ADVL  
 números [número] N M P @P<  
 de [de] <sam-> PRP @N<  
 a [o] <-sam> <artd> DET F S @>N  
 Secretaria=da=Segurança [Secretaria=da=Segurança] PROP F S @P<  
 .  
 especialistas "especialista" N M/F P @SUBJ>  
 atribuem [atribuir] <fmc> V PR 3P IND VFIN @FMV  
 o [o] <artd> DET M S @>N  
 aumento [aumento] N M S @<ACC  
 de [de] <sam-> PRP @N<  
 os [o] <-sam> <artd> DET M P @>N  
 crimes [crime] N M P @P<

a [a] <sam-> PRP @<PIV  
 a [o] <-sam> <artd> DET F S @>N  
 falta [falta] N F S @P<  
 de [de] PRP @N<  
 planejamento [planejamento] N M S @P<  
 de [de] <sam-> PRP @N<  
 o [o] <-sam> <artd> DET M S @>N  
 Estado [estado] <prop> N M S @P<  
 .  
 para "para" PRP @ADVL>  
 o [o] <artd> DET M S @>N  
 governo [governo] N M S @P<  
 ,  
 o "o" <artd> DET M S @>N  
 crescimento [crescimento] N M S @SUBJ>  
 se [se] <refl> PERS M/F 3S ACC @ACC>-PASS  
 deve [dever] <fmc> V PR 3S IND VFIN @FMV  
 a [a] <sam-> PRP @<PIV  
 a [o] <-sam> <artd> DET F S @>N  
 desigualdade [desigualdade] N F S @P<  
 social [social] ADJ F S @N<  
 e [e] <co-piv> KC @CO  
 a [a] <sam-> PRP @<PIV  
 o [o] <-sam> <artd> DET M S @>N  
 comércio [comércio] N M S @P<  
 irregular [irregular] ADJ M S @N<  
 .  
 pág. "pág." N F S/P @NPHR  
 .  
 C1 "C1" PROP F S @N<  
 .

## ANEXO C - Estatística Descritiva dos Fatores, por Gênero

TABELA 13 - ESTATÍSTICA DESCRITIVA DO FATOR 1, POR GÊNERO

<b>Gênero</b>	<b>Média</b>	<b>Desvio padrão</b>	<b>Nº de textos</b>
<b>Artigo</b>	0,50	3,52	86
<b>Carta</b>	0,70	5,71	84
<b>Chamada</b>	-1,59	5,60	73
<b>Comentário</b>	0,79	2,72	11
<b>Crítica</b>	-1,23	2,85	47
<b>Crônica</b>	5,29	6,11	12
<b>Editorial</b>	-1,56	2,35	20
<b>Entrevista</b>	5,29	4,72	24
<b>Nota de correção</b>	-2,62	1,99	7
<b>Coluna de notas</b>	-0,64	3,27	73
<b>Notícia</b>	-0,54	4,46	296
<b>Outros</b>	-0,51	4,88	25
<b>Reportagem</b>	0,26	4,22	655
<b>Resenha</b>	-1,58	3,37	18
<b>Total</b>	0,00	4,44	1.431

TABELA 14 - ESTATÍSTICA DESCRITIVA DO FATOR 2, POR GÊNERO

<b>Gênero</b>	<b>Média</b>	<b>Desvio padrão</b>	<b>Nº de textos</b>
<b>Artigo</b>	3,28	2,37	86
<b>Carta</b>	0,76	2,84	84
<b>Chamada</b>	-2,19	2,22	73
<b>Comentário</b>	1,37	1,24	11
<b>Crítica</b>	1,31	2,71	47
<b>Crônica</b>	3,16	1,08	12
<b>Editorial</b>	1,95	1,28	20
<b>Entrevista</b>	5,97	3,71	24
<b>Nota de correção</b>	-3,30	2,34	7
<b>Coluna de notas</b>	0,52	2,18	73
<b>Notícia</b>	-2,11	2,65	296
<b>Outros</b>	-1,89	5,40	25
<b>Reportagem</b>	0,26	2,20	655
<b>Resenha</b>	0,32	3,21	18
<b>Total</b>	0,00	2,98	1.431

## ANEXO D - Textos “Típicos” dos Gêneros

A título de exemplificação dos gêneros do texto jornalístico estudados neste trabalho, a seguir é apresentado o que poderia ser chamado de textos “típicos” de cada gênero, à luz dos fatores de variação lingüística obtidos nos resultados desta pesquisa. Com base nas pontuações dos escores médios dos Fatores 1 e 2 nos gêneros (cf. 4.7 e Anexo C), foi montada uma lista combinada de ambos os fatores em cada subcorpus de gêneros, resultando na seleção dos textos seguintes, cujos escores possuem simultaneamente a maior aproximação possível com os escores médios dos Fatores 1 e 2 obtidos na pesquisa, em todos os gêneros estudados.

O método proposto pode ser visto também como uma tentativa de estabelecer um critério lingüístico objetivo para a seleção de um texto “típico” de determinado gênero dentre um grupo de textos.

## TEXTO 8 - EXEMPLO “TÍPICO” DE ARTIGO

Nível de bem-estar material está menor do que no ano passado

FRANCISCO PESSOA FARIA

ESPECIAL PARA A FOLHA

PARA ENTENDER a discrepância entre a percepção (negativa) comum sobre o estado da economia brasileira e o crescimento do PIB é preciso levar em consideração algumas questões metodológicas e estatísticas, além do perfil do crescimento revelado pelo IBGE.

Começando pelos aspectos metodológicos, vale destacar os subsetores Aluguéis e Administração Pública, que têm grande peso no PIB (respectivamente, 11% e 16%). Uma boa parte do subsetor Aluguéis corresponde, seguindo a metodologia recomendada pela ONU, ao chamado aluguel imputado –um “serviço de habitação” proporcionado por todo domicílio, mesmo que seja de propriedade de quem o habita. Ora, a não ser no caso de uma catástrofe, o número de domicílios no país não teria por que cair; na verdade, devido ao aumento populacional, ele costuma registrar todo ano um crescimento em torno de 2% –o que ajuda a “segurar” a queda do PIB quando outras atividades não vão bem. Já no caso da Administração Pública ocorre algo semelhante: devido às dificuldades de obtenção de estatísticas de todos os níveis de governo, o IBGE calcula parte importante da evolução desse subsetor por meio do crescimento da população (além de utilizar dados sobre saúde e educação públicas), adotando a hipótese de que a utilização de serviços per capita permanece igual. Como a população cresce, uma parte do PIB da Administração Pública sempre cresce, independentemente dos gastos públicos. Note-se, portanto, que o impacto registrado provavelmente é bem menor do que vem acontecendo na prática, já que as autoridades estão promovendo um forte ajuste fiscal.

Quanto ao aspecto estatístico, lembremos que no período de comparação (primeiro trimestre de 2002) a economia também não estava bem, ainda sob impactos diretos e indiretos do apagão de 2001. Como o nível de atividade estava baixo, crescer em relação a ele não significa que as empresas estejam em uma boa situação, mas apenas que estejam em uma situação menos desconfortável.

Finalmente, é preciso destacar a diferença entre a evolução do mercado doméstico e a do comércio exterior. Os números do IBGE mostraram que o consumo privado caiu em relação ao primeiro trimestre de 2002, o que corrobora a percepção de que o nosso nível de bem-estar material está realmente menor do que no ano passado. Mas mostram também que as importações recuaram mais do que o registrado pelo consumo privado, e, ainda, que as exportações subiram. Isso significa que parte do aumento da produção serviu para substituir bens importados e parte serviu para atender o consumo do exterior. Nesse sentido não há contradição entre o fato de estar ocorrendo, simultaneamente, um aumento da produção nacional (e do PIB) e uma queda do consumo interno.

Quanto ao resto do ano, espera-se que o consumo privado comece a melhorar, caso não haja deterioração do cenário externo. Isso porque os bons resultados das exportações poderiam ser mantidos e haveria uma maior disponibilidade de capital externo para nos financiar. Isso, por sua vez, ajudaria a manter o câmbio e a inflação sob controle, o que permitiria a redução da taxa de juros, diminuiria a deterioração do salário real e aumentaria a confiança dos consumidores.

Francisco Pessoa Faria, 35, é economista sênior da LCA Consultores.

(CP5\_0830.TXT, escores: Fator 1, 0,8; Fator 2, 3,52)

## TEXTO 9 - EXEMPLO "TÍPICO" DE CARTA

## Transtornos mentais

“Foi muito interessante o destaque que nosso presidente deu, na última quarta-feira, ao campo da saúde mental (‘Todos temos ‘um pouco’ de loucos, diz Lula’, Brasil, pág. A4, 29/5).

Não foi à toa que ele frisou para olharmos para os nossos últimos dez anos, pois estudos epidemiológicos apontam que cerca de 25% da população de São Paulo desenvolve pelo menos um transtorno mental ao longo de sua vida, que a prevalência anual desses transtornos chega a 19% e que, do ponto de vista da incapacitação para o trabalho, os transtornos mentais superam todos os outros diagnósticos, segundo dados da Secretaria da Saúde de São Paulo.

Isso nos leva a pensar numa política mais ampla, que substitua o modelo hospitalocêntrico-manicomial por uma rede de ações em saúde mental que passe pelos postos de saúde, pelo PSF, pelos centros de atenção psicossociais e pelas enfermarias de psiquiatrias em hospitais gerais e que forme políticas de intersetorialidade com a cultura, com a assistência social e com os demais setores da saúde.”

Pedro Carlos Carneiro, médico psiquiatra, psicólogo e membro do Fórum Paulista da Luta Antimanicomial (São Paulo, SP)

(CP2\_1383.TXT, escores: Fator 1, -0,79; Fator 2, 0,77)

## TEXTO 10 - EXEMPLO "TÍPICO" DE CHAMADA

Controladoria Geral da União encontra problemas em 92 das 100 localidades fiscalizadas; empresas negam irregularidades

Teles descumprem meta, diz governo

ELVIRA LOBATO

DA SUCURSAL DO RIO

Grandes teles descumpriam metas em 92% das cidades fiscalizadas em julho e agosto.

A conclusão é da Controladoria Geral da União, que analisou metas de qualidade e universalização em cem localidades atendidas por Telemar, Brasil Telecom e Telefônica.

As metas tratam de qualidade e universalização dos serviços, como oferta de orelhões, instalação em escolas e postos de saúde e acesso gratuito a serviços de emergência.

A CGU, órgão do governo federal, coloca em dúvida a certificação dada à Telemar e à Telefônica pela Anatel no ano passado pelo cumprimento antecipado das metas de 2003.

Como prêmio, elas foram autorizadas a operar DDD, DDI, telefonia celular e chamadas locais fora das áreas de concessão. A Brasil Telecom diz que atingiu as metas em fevereiro.

O subcontrolador-geral da União, Jorge Hage, qualifica de "muito graves" as conclusões.

A Anatel só falará após analisar os relatórios. As teles afirmam cumprir as regras.

Págs. B1 e B4

(CP7\_1121.TXT; escores: Fator 1, -1,87; Fator 2, -2,18)

## TEXTO 11 - EXEMPLO "TÍPICO" DE COLUNA DE NOTAS:

'Vírus do Big Brother 4' dribla a Globo

DANIEL CASTRO

COLUNISTA DA FOLHA

A TV Globo contratou uma empresa especializada em segurança na internet para tentar descobrir quem é o hacker que está distribuindo e-mails que convidam internautas a participarem de um falso processo de seleção para o "reality show" "Big Brother Brasil 4". Há três meses, técnicos do portal Globo.com tentam, sem sucesso, rastrear o hacker.

O e-mail diz que "a Globo.com está fazendo uma alto (sic) seleção de candidatos via internet" e convida o usuário a baixar em seu computador um arquivo com um formulário para a inscrição. O arquivo contém um vírus, que trava a máquina. Recentemente, a Globo descobriu que o vírus "rouba" senhas bancárias caso o usuário faça transações desse tipo na internet. A emissora diz desconhecer a existência de vítimas.

No último final de semana, uma nova onda de e-mails foi disparada. Anteontem, a Globo divulgou o terceiro comunicado oficial à imprensa informando que se trata de uma fraude. Pela primeira vez desde 18 de junho (data do primeiro comunicado) a emissora admite que o internauta "tem ameaçada a segurança de suas futuras operações na internet".

Segundo a Globo.com, o hacker ainda não foi descoberto porque muda de computador rapidamente. As inscrições para o quarto "BBB", próximo "reality show" da Globo, em janeiro, começam em breve, segundo o diretor do programa, J.B. de Oliveira.

OUTRO CANAL

Ozzy 1

Já tem data de estréia a série de 12 pequenos "reality shows" da MTV, inspirada na norte-americana "The Osbournes". Será em 20 de outubro. Segundo Zico Goes, diretor de programação, a série será apresentada durante três semanas, de segundas a quintas, às 23h30, em comemoração dos 13 anos da emissora.

Ozzy 2

Cada um dos 12 artistas ou "famílias de artistas", todos músicos brasileiros, terão um programa de meia hora. O primeiro, de João Gordo, foi captado por duas câmeras em dois dias, semana passada. "Ficou muito legal, tem uma pegada intimista", diz Goes.

Internacional

O canal pago A&E Mundo exhibe no próximo dia 24, às 21h, na série "Biography", um documentário sobre o presidente Lula – o primeiro do conceituado selo dedicado a um chefe de Estado brasileiro. O programa, que tem imagens inéditas do arquivo pessoal do presidente, inicialmente passará em 20 países da América Latina. E ficará disponível aos demais países em que o canal atua.

Fama

A quase esquecida e pouco conhecida atriz Bianca Castanho está cotadíssima para ser a protagonista da próxima novela do SBT, "Canavial das Paixões". Ela almoçou ontem com diretores na sede da emissora. A atriz teve algum destaque em "Terra Nostra", como Florinda, e papéis secundários em "Malhação", "Uga Uga" e "O Beijo do Vampiro", todas na Globo.

E-mail - daniel.castro@uol.com.br

(CP3\_0497.TXT; escores: Fator 1, -1,21; Fator 2, 0,23)

TEXTO 12 - EXEMPLO “TÍPICO” DE COMENTÁRIO

Revide faz EUA buscarem opções

Coalizão tenta nova alternativa de cerco a Bagdá, enquanto sofre contra-ataque

RICARDO BONALUME NETO

DA REPORTAGEM LOCAL

A tempestade de areia não causou apenas uma pausa no avanço americano. Pode ter sido o principal motivo de os iraquianos terem contra-atacado com forças mecanizadas –cerca de mil veículos de tipos variados em torno de Bagdá–, segundo relatos divulgados ontem pelos correspondentes junto da 3ª Divisão de Infantaria dos EUA.

Se de fato se confirmar que os iraquianos têm procurado atacar em massa com veículos de combate, isso significa uma boa notícia para os americanos. Em movimento, eles passam a ser alvos identificáveis pela aviação dos EUA, e podem ser atacados longe de áreas civis.

Mas justamente por isso o contra-ataque estaria acontecendo agora. A areia impediu os helicópteros Apache de voarem. Também dificultou a localização de alvos pelos caça-bombardeiros. Mesmo os modernos sistemas de visão noturna e infravermelho (detecção de calor) dos veículos blindados ficam inúteis em meio da tempestade de areia.

O poder de fogo dos EUA na ofensiva até Bagdá é altamente dependente da aviação, que age como se fosse uma “artilharia voadora”. Em 1991 os EUA só atacaram depois que a aviação destruiu o que imaginavam ser 50% da capacidade de combate iraquiana (no pós-guerra se descobriu que muito do que fora destruído eram alvos falsos).

É uma tática que pode ser batizada de “abraço de tamanduá”. Em campo aberto, o tamanduá não tem como vencer uma onça. Sua única chance seria fincar suas garras no dorso da onça em um abraço mortal.

Procurando se aproximar dos americanos, os iraquianos conseguiriam impedir os EUA de bombardeá-los, dado o risco de atingir suas próprias tropas (o curiosamente chamado “fogo amigo”, também conhecido como “azul contra azul”, a cor usada nos mapas para representar as forças próprias –o inimigo fica em vermelho).

A resistência iraquiana não era de todo imprevisível. Praticamente todos analistas especularam que na defesa do regime os iraquianos mais fiéis a Saddam Hussein lutariam.

“Nunca acue um tigre, senão ele salta sobre você”; “o inimigo tem sempre três opções, e ele geralmente escolhe a quarta”, costumavam dizer os antigos filósofos chineses da guerra, dos quais Sun Tzu é o mais famoso –e particularmente bem citado no livrinho de dois estrategistas americanos chamado “Choque e Pavor”.

Uma correspondente do “The Washington Post” entrevistou um piloto de helicóptero Apache, Lance McElhiney, veterano tanto da Guerra do Vietnã como da Guerra do Golfo de 1991. McElhiney foi alvo de tiros de fuzis em 91 e agora de novo no fracassado ataque em que pelo menos um helicóptero foi forçado a pousar.

“Você não deve subestimar o inimigo”, disse ele à repórter Mary Beth Sheridan; “no Vietnã, nós subestimamos o cara nos arrozais”.

O grande problema da estratégia americana foi não ter obtido bases na Turquia. O plano inicial previa um avanço em duas frentes: na direção sul a partir da Turquia, na direção norte a partir do Kuwait.

Esse avanço simultâneo diminuiria muito as opções de resistência dos iraquianos, tornando o cerco de Bagdá bem mais fácil. Um “cerco” só merece esse nome se não deixa nenhuma

saída para o adversário.

Na falta de uma força terrestre grande saída da Turquia, os EUA começaram agora a procurar opções. Ontem se revelou que cerca de mil pára-quedistas da 173ª Brigada Aerotransportada foram lançados sobre uma pista de pouso em território controlado pelos curdos ao norte do Iraque. De posse da pista, podem chegar reforços por aviões de transporte. Mas mesmo o maior deles só pode levar um tanque Abrams por vez.

(CP4\_0598.TXT; escores: Fator 1, 1,35; Fator 2, 1,66)

## TEXTO 13 - EXEMPLO "TÍPICO" DE CRÍTICA:

Série renova e preserva a arte da dança

INÊS BOGÉA

CRÍTICA DA FOLHA

REGISTRO ou criação? É memória documental ou uma nova coreografia que nasce da união entre filme e dança?

Iniciativa rara no Brasil, o "STV na Dança" já apresentou 60 programas do segundo semestre de 2001 para cá. Com direção de Antônio Carlos Rebesco (Pipoca) e direção de fotografia de Carlos Travaglia, a série trouxe companhias de estilo e porte muito diversos: Balé da Cidade de São Paulo (em cartaz amanhã, com "Desatino do Norte Desatino do Sul", coreografia de Jorge Garcia), Quasar, Cisne Negro, Verve, Márcia Milhazes, Cena 11, Balé do Teatro Municipal do Rio de Janeiro, Ballet Stagium e outras tantas.

Apresentados pelos atores Felipe Folgosi e Brian Penido Ross, os programas têm comentários de Ana Francisca Ponzio e incluem também depoimentos dos próprios artistas.

Criação: a dança filmada é uma outra dança. Nosso olhar será guiado, agora, pela maneira como ela foi filmada e montada. Exemplo: se uma câmera acompanha um bailarino que se desloca na cena, ela atenua seu movimento. Se a câmera está do lado oposto à saída do bailarino, a percepção é outra. Assim, todo espaço muda neste movimento que é duplo, da câmera e do bailarino.

Registro: num país que não prima pela preservação da memória, esta série cumpre papel importante pelo simples esforço de mapear a produção atual da dança brasileira.

A dança ao vivo é efêmera. Nada é mais impalpável do que uma coreografia em cena; ao mesmo tempo, nada é mais presente do que os corpos. Depois, fica só a memória que cada um tem do espetáculo. E agora também essa mistura de criação e memória, que renova e preserva a arte da dança, nos modestos milhões de palcos das telas de TV.

STV na Dança: Desatino do Norte Desatino do Sul

Ótimo

Quando: amanhã, às 13h30, na Rede SescSenac

(CP2\_0328.TXT; escores: Fator 1, -2,01; Fator 2, 2,12)

## TEXTO 14 - EXEMPLO "TÍPICO" DE CRÔNICA:

Fome Zero! Bush taxa armas e manda bala!

BUEMBA! Buemba! Macaco Simão urgente! O braço armado da gandaia nacional. Direto da República da Língua Plesa! Nada de novo no front! Ninguém botou botox hoje! E o que o Lula foi fazer na Cúpula de Evian? Tomar água mineral! E quando vai ter a Cúpula da Minalba? Não tem Davos e Porto Alegre? Então tem que ter Evian e Minalba! E sabe por que o Lula e o Bush se deram tão bem? Porque nenhum dos dois sabe falar inglês. E adorei a sugestão do Lula: "Taxar as armas pra dar pro Fome Zero Mundial". Para aquele bando de facínoras? Abalou o mundo da bala!

E eu sei o que o Bush vai mandar pro Fome Zero no mundo: bala. Vai todo mundo ficar chupando bala! Fome Zero urgente! Bush taxa armas e manda bala. Rarárá! E, a cada bomba que o Bush joga, o povo come fogo! E eles não vão dar peixe pros pobres. Eles vão dar pobres pros peixes: joga todo mundo no mar!

E continua a repercussão da grande bomba da semana: "Americanos clonaram uma mula". Grandes novidades! O Bush pai já fez isso há muito tempo. E agora vão fabricar Bush em série? Isso é que é destruição em massa. E como disse um amigo: "Arma química é a minha sogra quando se esquece de tomar Gardenal".

E diz que agora vão lançar um Viagra só pro pessoal acima dos 70: Viagra com memoriol. E um amigo meu diz que tá topando qualquer emprego. Inclusive de piloto de prova de fábrica de supositório. De ITU! E o FMI dá sinais de confiança no Brasil. Errado. O FMI vive à custa do Brasil!

Manual do Contra! Continua a todo o vapor o Manual do Contra da Heloísa Helena. Diz que a Heloísa Helena é tão do contra que só ouve música COUNTRY! E diz que falar que ela é do contra é CONTRAproducente. E um leitor me mandou uma foto da Heloísa Helena fazendo propaganda de celular: "VIVO sendo do contra!". Aliás, sabe o que ela parece? Uma adolescente mimada!

E a penúltima derradeira final do Bestiário Tucanês. É que no condomínio de um amigo meu estava a placa: "Local destinado a rota de fuga em casos de sinistro". Tucanaram a saída de emergência! E aí o prédio pega fogo, e até acabar de ler a placa inteira, morre queimado!

Pior, um outro amigo foi a uma rave, e um segurança falou pro outro: "Vá lá e faça uma busca pessoal nele". Busca pessoal? Tucanaram o baculejo! Socorro! Chama o Oswaldo Cruz pra erradicar o tucanês.

E atenção! Cartilha do Lula. Mais um verbete pro óbvio lulante. "Montadora": companheira que anda a cavalo. Rarárá! Nós sofre, mas nós goza. Hoje só amanhã. Que eu vou pingar o meu colírio alucinógeno!

simao@uol.com.br

(CP2\_0330.TXT; escores: Fator 1, 4,52; Fator 2, 3,0)

## TEXTO 15 - EXEMPLO "TÍPICO" DE EDITORIAL

## LIBERALISMO RADICAL

O NOVO pacote de corte de impostos do governo Bush foi recebido com grande apreensão nos meios financeiros internacionais. O conservador diário londrino "Financial Times" expressou-se em editorial intitulado "Desvario tributário". O receio é de que o corte de impostos, que se estende por dez anos, coloque o déficit fiscal dos EUA em trajetória explosiva.

O governo Bush responde à crítica afirmando que o pacote ajudará a impedir uma recessão nos próximos meses. Afirma, além disso, que o crescimento será estimulado continuamente pela redução de impostos, reforçando a receita tributária e desmentindo as previsões "alarmistas" sobre a trajetória do déficit.

Os críticos retrucam que o pacote será pouco eficaz contra a ameaça de recessão, porque a redução imediata de impostos é pequena. Ao longo do tempo, o estímulo ao crescimento continuaria limitado, porque o corte de impostos beneficia sobretudo os mais ricos, que têm menor propensão a aumentar seus gastos.

As críticas ao novo pacote de Bush vão além. Uma delas já vinha sendo ventilada há tempos: a perspectiva de alta do déficit fiscal pressionaria a taxa de juros de longo prazo, prejudicando o crescimento, e agravaria a fragilidade do dólar ante as demais moedas fortes, podendo provocar uma situação financeira perigosa.

Outra preocupação, que vinha sendo manifestada por analistas ligados ao partido Democrata, como Paul Krugman, acaba de receber o endosso do "Financial Times". Trata-se da suspeita de que o governo Bush avalia que o corte de impostos poderá criar uma crise fiscal, mas pretende aproveitar a oportunidade para aprovar agressivos cortes de gastos sociais. Em condições normais, o Congresso recusaria tais cortes, mas, em meio a uma crise fiscal, eles poderiam se tornar inevitáveis.

O liberalismo radical da administração Bush começa a incomodar até setores conservadores, que se assustam com a perspectiva de desmonte de uma rede de proteção social construída ao longo de décadas.

(CP5\_0758.TXT; escores: Fator 1, -1,81; Fator 2, 1,82)

## TEXTO 16 - EXEMPLO "TÍPICO" DE ENTREVISTA

## Mulheres alteradas vão às compras

Quando não está desenhando mulheres que reclamam de celulite ou discutem a diferença que faz na vida aquele sapato, a cartunista argentina Maitena reclama da celulite, embasbaca-se diante de shorts que a deixam com "pernas lindas" e não consegue entrar numa loja sem sair com uma sandália, como fez nas três que visitou em São Paulo com a Folha.

"Mulheres alteradas sou eu", poderia dizer a cartunista, parodiando a divisa do escritor francês Gustave Flaubert (1821-1880). Age como as personagens que desenha. "Mulher, como dizem os homens, é tudo igual", debocha. O sucesso e o olhar impiedoso sobre o seu sexo, porém, diferenciam Maitena.

Aos 41 anos, ela é um dos maiores fenômenos do humor latino-americano. Seus livros já venderam 1 milhão de exemplares e suas tiras são publicadas em 18 países (da França ao México, da Holanda à Grécia). Nada mal para uma cartunista que, em 1992, planejava abandonar o desenho e abrir um restaurante.

Ela até já havia tido um em Santelmo, em Buenos Aires, em 1982, chamado La Antiga Tasca de San Fermin. Teve também um quiosque de cigarros e foi dona de um bar no Uruguai, numa praia chamada La Pedrera: "Eu era uma péssima empresária. Bebia tudo o que deveria vender". Fama, ela só havia alcançado no obscuro circuito de quadrinhos underground, no qual publicava histórias eróticas em revistas como "El Vibora", de Barcelona. Dinheiro, à época, era pouco mais que miragem.

Até que "Mulheres Alteradas" estourou na Espanha em 1997. Há dois anos, ela abandonou Buenos Aires e foi morar em La Pedrera, num vilarejo onde vivem cem pessoas, com uma filha de quatro anos e o marido e agente, Daniel Kon. Seus dois filhos mais velhos, de 23 e 21 anos, continuam em Buenos Aires. Enquanto fazia compras, ela falou sobre o que diferencia homens de mulheres, teorizou sobre sua obsessão por sapatos e esboçou um tratado geral da celulite.

## ALTERADAS X ESTRESSADOS

"Somos alteradas porque fazemos muitas coisas ao mesmo tempo. Temos muitas frentes de combate abertas: no trabalho, com o corpo, com o amor, com os filhos, como mãe, como amigas. Queremos fazer muitas coisas e não se pode fazer tantas coisas bem. Algo sempre sai mal. É por isso que sempre estamos correndo. É por isso que sentimos culpa quando não cuidamos bem dos filhos. Se descuidamos do corpo, sentimos culpa. Se não somos boas amantes, nos sentimos umas bruxas. A mulher alterada não é histérica. Histérica é aquela que seduz e depois não quer. As mulheres que eu desenho são mulheres que querem ser tocadas, querem casar e ter filhos.

Os homens se estressam, e isso é pior. Nós, mulheres, somos mais expressivas, choramos mais, gritamos mais, rimos mais, colocamos tudo para fora. Os homens engolem tudo e têm mais acidez, ficam doentes. É por isso que os cabelos dos homens caem. Nas mulheres cai tudo, menos os cabelos.

## MACHISMO

"Alguns homens estúpidos me acusam de feminista. Mas não me interessa o que pensam os homens estúpidos. Meus leitores homens são todos inteligentes e sensíveis. Os outros que fiquem vendo televisão."

## ODE AO FEMINISMO

“Se não tivesse existido o feminismo, todas nós estaríamos passando roupa. Acho que o feminismo foi o movimento político mais importante do último século por causa das mudanças sociais que gerou. Foi mais importante que o socialismo. Graças a ele, as mulheres puderam ser socialistas. Antes, as mulheres não eram nada. Quando as meninas de 20 anos dizem que o feminismo é só uma antiguidade, tenho vontade de dizer: ‘Vocês não podem cuspir contra o ventilador’. É muito fácil criticar o feminismo quando se tem todas as liberdades que o movimento conquistou.”

#### SOMOS TODAS MACHISTAS

“Às vezes sou dura com as mulheres, mas há uma gota de ternura nas minhas críticas. Sou dura porque sou crítica comigo mesma e com as mulheres em geral. E, às vezes, sou um pouco machista, e não gosto disso. Lamentavelmente, fomos todos educados por uma geração de pais e mães machistas, e é difícil acabar com isso. Como mãe, tenho atitudes machistas com meus filhos. Quando eram adolescentes, e não voltavam de uma festa, o que me preocupava era minha filha, não meu filho. Isso é machismo. A garota às vezes sabe se cuidar melhor do que o menino. Quando a garota tem três namorados que telefonam a toda hora, a mãe fica preocupada. ‘Você é muito fácil, não pode ...’ Mas quando o filho tem três namoradas, ela fica orgulhosa: ‘Muito bem...’. É um tipo de machismo que todas nós temos.”

#### TRATADO GERAL DA CELULITE

“Câncer só algumas mulheres têm, mas celulite todas têm, cedo ou tarde. É por isso que minhas personagens se preocupam mais com celulite. É horrível. É algo que se apodera do seu corpo como um “alien”. É um ser estranho que te invade, que transforma a tua carne.

Todos os homens dizem que não ligam para celulite, mas na praia olham só para as que não têm celulite. Não sejam mentirosos. Celulite é uma grande injustiça da natureza. Por que os homens não têm celulite? As únicas mulheres que não têm celulite são os travestis.”

bergamo@folhasp.com.br

MARIO CESAR CARVALHO (INTERINO) COM CLEO GUIMARÃES E ALVARO LEME

(CP7\_1273.TXT; escores: Fator 1, 6,71; Fator 2, 6,5)

## TEXTO 17 - EXEMPLO "TÍPICO" DE NOTA DE CORREÇÃO

## ERRAMOS

O primeiro lote de livros da "Biblioteca Folha" será entregue aos assinantes até 12 de julho, e não até 12 de junho, como informou o texto "Coleção reúne 30 clássicos do século 20" (caderno Biblioteca Folha, pág. Especial 1, 1º/6).

\*

Diferentemente do que informou o texto "BNDES vai bancar canal de TV latino" (Dinheiro, pág. B2, 30/5), Gabriel Priolli não exerceu o cargo de presidente da TV Educativa do Rio Grande do Sul.

\*

Diferentemente do publicado na reportagem "Nem antibiótico mata o Anthrax" (Folhateen, pág. 5, 2/6), grafa-se "thrash metal", e não "trash metal".

\*

Por um erro de redação, frase no texto "BC culpa reajuste de salário por manter juro" (Dinheiro, pág. B12, 30/5) ficou com o sentido invertido. Em lugar de "o esforço do comércio e da indústria para recompor as margens de lucro (...) contribuiu para a persistência do recuo da inflação", o certo seria "contribuiu para a persistência da inflação".

\*

Diferentemente do publicado nesta seção no domingo passado, o jogo entre Palmeiras e CRB, pela série B do Campeonato Brasileiro, aconteceu no sábado, e não no domingo. (CP2\_0205.TXT; escores: Fator 1, -2,57; Fator 2, -3,28)

## TEXTO 18 - EXEMPLO "TÍPICO" DE NOTÍCIA

## Incêndio atinge fábrica por 4 horas

Um incêndio destruiu na madrugada de ontem uma fábrica de produtos para tratamento de água de piscinas em Guarulhos (Grande São Paulo). O fogo começou por volta da 1h na sede da Genco, no bairro de Cumbica, onde funciona também o armazém da empresa. Trabalham ali 85 pessoas, mas, na hora do incidente, apenas o vigia estava no local. Os bombeiros só conseguiram controlar o incêndio após quatro horas de trabalho.

(CP1\_0089.TXT; escores: Fator 1, -0,96; Fator 2, -2,03)

## TEXTO 19 - EXEMPLO "TÍPICO" DE "OUTROS GÊNEROS"

HÁ 50 ANOS  
DO BANCO DE DADOS

"Dulles e Rhee chegam a acordo sobre a conferência da paz" foi a manchete da Folha no dia 6 de agosto de 1953.

John Foster Dulles, secretário de Estado dos EUA, e o presidente coreano, Syngman Rhee, entraram em acordo para a realização de uma conferência política internacional, entre 1º e 15 de outubro.

Os dois conferencistas ainda não chegaram a um acordo sobre o lugar indicado para a conferência nem sobre as nações que deverão dela participar. Concordaram, não obstante, que o conclave se dedique exclusivamente à unificação da Coreia.

<http://bd.folha.com.br>

(CP3\_0437.TXT; escores: Fator 1, -242; Fator 2, -3,12)

## TEXTO 20 - EXEMPLO "TÍPICO" DE REPORTAGEM

Câmbio instável preocupa os países ricos

Líderes pressionam BCE para que corte os juros na quinta-feira; novos números reforçam cenário de desaceleração na Europa

MARIA LUIZA ABBOTT

ENVIADA ESPECIAL A EVIAN

Os líderes do G8 (grupo dos sete países mais ricos, mais a Rússia) manifestaram confiança na recuperação da economia mundial, ontem. Mas a preocupação com a valorização do dólar diante do euro também transpareceu.

O chanceler alemão, Gerhard Schröder, chegou a ser veemente ao pedir que o Banco Central Europeu (BCE) corte os juros na reunião da próxima quinta-feira. "Com todo o devido respeito pela independência do BCE, deixamos claro que ainda há espaço para estimular o crescimento", disse.

Um corte de juros poderia conter a valorização do euro e ainda dar impulso à economia alemã, ameaçada pela recessão.

As economias do G8 (EUA, Japão, Alemanha, Reino Unido, França, Itália e Canadá, além da Rússia, convidada por sua importância geopolítica) representam dois terços da renda mundial.

O presidente Jacques Chirac (França), em entrevista no fim do dia, disse que "houve convergência nas análises de que estão reunidas todas as condições para a retomada do crescimento".

Apesar do otimismo, a própria França mostrou preocupação com a relação entre dólar e euro. A porta-voz de Chirac, Catherine Colonna, disse que os líderes consideravam a estabilidade cambial "um elemento muito importante para o crescimento".

Durante as conversas, Bush disse que estava comprometido com um dólar forte, mas minimizou a sua influência sobre o câmbio. Sugeriu que as taxas de juros eram determinantes para a decisão dos investidores de aplicar em uma ou outra moeda. "Os EUA apóiam um dólar forte. Por isso é importante garantir políticas que acelerem o crescimento nos Estados Unidos", disse o porta-voz da Casa Branca, Ari Fleischer.

O euro já subiu 28% nos últimos 12 meses. Ontem, no entanto, o dólar recuperou terreno, porque operadores do mercado começam a acreditar em um corte de juros do BCE. Mas, em Evian, não teria havido discussão sobre intervenção no câmbio.

A valorização do euro está abalando a competitividade das exportações da zona do euro, especialmente da Alemanha, a terceira economia do planeta. O país registra recordes de desemprego, dois trimestres seguidos de estagnação do PIB e inflação baixa, beirando a deflação.

Além disso, dois dados divulgados ontem apontam para uma possível exacerbação da desaceleração econômica na zona do euro.

Pesquisa Reuters/NTC que sonda os pedidos feitos à indústria registrou contração, no mês passado, ainda maior que a apurada em abril. O Índice dos Gerentes de Compra, que foi de 47,8 em abril, chegou a 46,8 em maio, o pior nível em 16 meses. Resultados abaixo de 50 indicam contração.

Outro número que pode indicar fraqueza da atividade econômica divulgado ontem foi a queda da inflação na zona do euro. Em termos anualizados, a taxa caiu de 2,1%, em abril,

para 1,9% em maio

A diretoria do BCE tem resistido às pressões por cortes de juros. Um de seus principais argumentos era, justamente, o de que a inflação média dos países da zona do euro estava acima ou próxima da sua meta anual de 2%.

No mês passado, o BCE flexibilizou a definição de sua meta inflacionária, dizendo que a taxa deveria ficar próxima de 2%, e não abaixo disso. A mudança foi vista como um prenúncio de que, no encontro de depois de amanhã, a autoridade monetária irá, enfim, diminuir os juros básicos.

(CP2\_0225.TXT; escores: Fator 1, 0,25; Fator 2, 0,51)

## TEXTO 21 - EXEMPLO "TÍPICO" DE RESENHA

Índio avalia livro de branco

MAURÍCIO M. KAMAYURÁ

ESPECIAL PARA A FOLHINHA

O livro "Sobrevivendo à Escuridão" é muito legal porque traz muitas informações sobre os seres da natureza, nas cavernas, no subsolo e nos oceanos, que existem aqui no Brasil e nos outros países. Por exemplo, morcego, marmota e poraquê (peixe-elétrico).

O livro é interessante porque tem explicações em detalhes. Por exemplo, como as cavernas se formam e como as toupeiras cavam galerias. É importante aprofundar o que está nos livros de ciências. De Iris Stern (ed. Saraiva). [R\$ 16,00.]

Maurício Mattar Matariwa Kamayurá, 23, é professor da escola indígena municipal Ypawu Kamayurá (MT). A Folhinha levou 150 livros e revistas para lá. A Varig fez abatimento no excesso de peso da bagagem da reportagem.

(CP6\_1118.TXT; escores: Fator 1, -0,52; Fator 2, 0,82)

## ANEXO E - Estatística Descritiva de Frequência das Variáveis, por Gênero

TABELA 15 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO ARTIGO (N\* = 86; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
Adjetivos	72,0	31,6	133,8	102,1	20,2	0,3	-0,2
Advérbios	141,7	65,8	199,0	133,2	28,2	-0,1	-0,3
Aposições	3,9	0,0	34,7	34,7	5,2	3,1	14,0
Aspas	8,9	0,0	102,4	102,4	12,5	5,2	36,7
Conjunções coordenativas	34,3	9,6	70,4	60,7	11,7	0,5	0,9
Conjunções subordinativas	14,6	0,0	37,0	37,0	7,8	0,6	0,2
Futuro do pretérito	3,4	0,0	16,6	16,6	4,0	1,6	2,5
Nomes próprios	47,8	3,0	217,6	214,6	30,6	2,3	10,3
Números cardinais	15,4	0,0	52,5	52,5	11,6	1,2	1,2
Presente do indicativo	55,7	8,0	135,9	128,0	22,8	0,4	1,0
Pretérito imperfeito	7,8	0,0	46,5	46,5	8,6	2,2	6,5
Pretérito perfeito	15,2	0,0	44,3	44,3	11,1	0,9	-0,1
Pronomes demonstrativos	11,2	0,0	33,5	33,5	6,0	1,0	1,7
Quantidade de palavras	646,6	292,0	2582,0	2290,0	320,5	2,8	14,9
Substantivos	226,3	108,8	284,3	175,4	31,1	-0,5	1,3
Verbos públicos	2,4	0,0	13,3	13,3	2,6	1,9	4,6
Verbos / pronomes 1ª pessoa singular	5,2	0,0	63,3	63,3	10,4	4,0	19,1
Verbos / pronomes 3ª pessoa singular	72,2	22,1	126,6	104,6	20,7	0,1	0,2
Voz passiva	3,3	0,0	12,7	12,7	3,0	1,0	0,6

\* N significa “número de textos”.

TABELA 16 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CARTA (N = 84; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	63,5	0,0	131,0	131,0	29,1	0,0	-0,3
<b>Advérbios</b>	125,8	19,6	253,5	233,9	48,1	0,2	-0,2
<b>Aposições</b>	5,2	0,0	26,5	26,5	7,6	1,3	0,5
<b>Aspas</b>	9,2	0,0	43,5	43,5	7,7	1,6	4,5
<b>Conjunções coordenativas</b>	31,6	0,0	76,9	76,9	18,3	0,1	-0,4
<b>Conjunções subordinativas</b>	14,9	0,0	48,8	48,8	13,1	0,5	-0,6
<b>Futuro do pretérito</b>	4,3	0,0	27,8	27,8	6,4	1,5	1,9
<b>Nomes próprios</b>	70,3	2,3	241,4	239,1	39,8	1,3	3,2
<b>Números cardinais</b>	18,5	0,0	102,9	102,9	19,7	1,9	4,7
<b>Presente do indicativo</b>	61,6	0,0	121,3	121,3	29,6	-0,2	-0,4
<b>Pretérito imperfeito</b>	4,3	0,0	51,7	51,7	9,2	2,8	9,2
<b>Pretérito perfeito</b>	15,4	0,0	98,2	98,2	21,9	2,2	5,1
<b>Pronomes demonstrativos</b>	9,3	0,0	45,5	45,5	10,5	1,2	1,1
<b>Quantidade de palavras</b>	141,6	29,0	536,0	507,0	98,8	1,9	4,4
<b>Substantivos</b>	225,0	140,8	317,3	176,5	37,6	0,2	0,2
<b>Verbos públicos</b>	3,7	0,0	53,3	53,3	8,8	3,6	15,8
<b>Verbos / pronomes 1ª pessoa singular</b>	17,2	0,0	107,1	107,1	21,4	1,9	4,7
<b>Verbos / pronomes 3ª pessoa singular</b>	60,7	0,0	153,2	153,2	32,6	0,6	0,4
<b>Voz passiva</b>	2,8	0,0	33,3	33,3	6,0	3,0	10,5

TABELA 17 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CHAMADA (N = 73; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	54,4	0,0	166,7	166,7	33,8	1,0	1,9
<b>Advérbios</b>	97,7	0,0	226,1	226,1	43,8	0,3	0,5
<b>Aposições</b>	6,5	0,0	44,0	44,0	8,7	1,5	3,3
<b>Aspas</b>	14,3	0,0	120,0	120,0	29,7	2,5	5,5
<b>Conjunções coordenativas</b>	27,3	0,0	88,9	88,9	20,7	0,7	0,5
<b>Conjunções subordinativas</b>	9,0	0,0	65,6	65,6	13,9	1,8	3,4
<b>Futuro do pretérito</b>	2,0	0,0	33,3	33,3	6,2	3,5	12,2
<b>Nomes próprios</b>	108,7	13,9	227,3	213,4	39,6	0,0	-0,1
<b>Números cardinais</b>	22,3	0,0	93,8	93,8	23,0	1,0	0,6
<b>Presente do indicativo</b>	49,2	0,0	134,6	134,6	36,0	0,6	-0,2
<b>Pretérito imperfeito</b>	5,3	0,0	93,0	93,0	13,0	4,8	29,3
<b>Pretérito perfeito</b>	22,8	0,0	114,3	114,3	23,3	1,4	2,8
<b>Pronomes demonstrativos</b>	2,1	0,0	19,0	19,0	4,7	2,2	3,7
<b>Quantidade de palavras</b>	81,9	22,0	298,0	276,0	53,5	1,8	4,0
<b>Substantivos</b>	245,5	136,4	354,8	218,5	46,9	0,1	-0,2
<b>Verbos públicos</b>	4,0	0,0	32,8	32,8	8,7	2,2	3,6
<b>Verbos / pronomes 1ª pessoa singular</b>	2,1	0,0	69,8	69,8	11,0	5,8	32,8
<b>Verbos / pronomes 3ª pessoa singular</b>	72,4	0,0	200,0	200,0	40,1	0,3	0,9
<b>Voz passiva</b>	4,3	0,0	83,3	83,3	11,5	5,0	31,4

TABELA 18 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO COMENTÁRIO (N = 11; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	68,6	33,9	88,2	54,4	16,0	-0,9	0,8
<b>Advérbios</b>	132,3	96,5	191,9	95,4	27,4	0,9	0,9
<b>Aposições</b>	4,0	0,0	13,6	13,6	3,9	1,5	3,4
<b>Aspas</b>	8,5	0,0	19,7	19,7	7,2	0,5	-1,2
<b>Conjunções coordenativas</b>	20,0	11,6	30,8	19,2	5,3	0,7	0,7
<b>Conjunções subordinativas</b>	17,6	8,4	34,5	26,0	8,2	1,0	0,6
<b>Futuro do pretérito</b>	7,4	0,0	19,0	19,0	5,7	0,9	0,2
<b>Nomes próprios</b>	59,0	27,1	96,1	69,0	22,4	0,2	-0,8
<b>Números cardinais</b>	14,5	1,9	30,9	29,0	9,2	0,2	-0,7
<b>Presente do indicativo</b>	40,6	8,5	70,7	62,2	19,2	-0,4	-0,5
<b>Pretérito imperfeito</b>	9,6	0,0	27,3	27,3	8,5	0,9	0,3
<b>Pretérito perfeito</b>	21,8	7,6	43,7	36,1	10,4	1,0	0,6
<b>Pronomes demonstrativos</b>	7,7	4,1	13,6	9,5	3,5	0,5	-1,4
<b>Quantidade de palavras</b>	496,3	335,0	711,0	376,0	106,3	0,5	0,5
<b>Substantivos</b>	226,4	192,1	261,9	69,7	24,0	0,1	-1,2
<b>Verbos públicos</b>	2,7	0,0	8,6	8,6	2,6	1,1	1,3
<b>Verbos / pronomes 1ª pessoa singular</b>	0,4	0,0	2,1	2,1	0,8	1,9	2,0
<b>Verbos / pronomes 3ª pessoa singular</b>	75,0	49,4	92,4	43,0	12,5	-0,5	0,4
<b>Voz passiva</b>	5,6	1,7	12,6	10,9	3,0	1,3	2,2

TABELA 19 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CRÍTICA (N = 47; POR 1.000 PALAVRAS)

<b>Variável</b>	<b>Média</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Alcance</b>	<b>Desvio padrão</b>	<b>Assimetria</b>	<b>Curtose</b>
<b>Adjetivos</b>	70,7	14,3	136,0	121,7	26,2	0,3	-0,3
<b>Advérbios</b>	133,7	78,7	261,2	182,5	39,4	1,0	1,4
<b>Aposições</b>	4,3	0,0	17,7	17,7	4,5	0,8	0,0
<b>Aspas</b>	18,8	0,0	72,7	72,7	14,9	1,5	3,1
<b>Conjunções coordenativas</b>	36,4	0,0	62,7	62,7	13,7	-0,3	0,0
<b>Conjunções subordinativas</b>	8,8	0,0	24,0	24,0	7,3	0,4	-0,8
<b>Futuro do pretérito</b>	1,5	0,0	17,5	17,5	3,5	3,4	12,4
<b>Nomes próprios</b>	81,3	27,5	178,1	150,5	36,7	1,0	0,8
<b>Números cardinais</b>	20,3	2,5	55,6	53,1	12,6	1,2	1,4
<b>Presente do indicativo</b>	64,5	25,0	156,7	131,8	22,6	1,4	4,8
<b>Pretérito imperfeito</b>	3,5	0,0	22,6	22,6	5,4	1,8	2,8
<b>Pretérito perfeito</b>	8,9	0,0	36,5	36,5	10,1	1,2	0,7
<b>Pronomes demonstrativos</b>	10,4	0,0	37,6	37,6	8,7	1,2	1,6
<b>Quantidade de palavras</b>	327,6	55,0	1145,0	1090,0	208,4	1,4	4,0
<b>Substantivos</b>	214,4	137,0	265,4	128,4	35,7	-0,7	-0,6
<b>Verbos públicos</b>	1,6	0,0	8,2	8,2	2,5	1,4	0,5
<b>Verbos / pronomes 1ª pessoa singular</b>	1,2	0,0	9,3	9,3	2,3	2,1	3,9
<b>Verbos / pronomes 3ª pessoa singular</b>	74,8	36,4	126,9	90,5	21,0	0,4	-0,1
<b>Voz passiva</b>	2,8	0,0	14,9	14,9	3,6	1,6	2,7

TABELA 18 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO CRÔNICA (N = 12; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	52,7	24,2	92,8	68,7	18,7	0,7	0,7
<b>Advérbios</b>	130,5	57,7	178,2	120,5	33,2	-0,5	1,0
<b>Aposições</b>	2,6	0,0	11,1	11,1	3,4	1,6	2,6
<b>Aspas</b>	9,4	0,0	19,6	19,6	6,6	-0,1	-1,3
<b>Conjunções coordenativas</b>	46,2	30,1	64,2	34,0	10,2	0,1	-0,7
<b>Conjunções subordinativas</b>	19,3	2,5	31,4	28,9	8,0	-0,6	0,5
<b>Futuro do pretérito</b>	2,7	0,0	9,1	9,1	3,7	1,0	-0,7
<b>Nomes próprios</b>	53,2	3,0	126,1	123,1	41,0	0,4	-1,3
<b>Números cardinais</b>	10,6	2,5	29,3	26,8	7,4	1,6	3,0
<b>Presente do indicativo</b>	68,9	27,2	114,3	87,2	27,6	-0,2	-0,7
<b>Pretérito imperfeito</b>	15,2	2,1	57,4	55,3	20,3	1,6	1,3
<b>Pretérito perfeito</b>	22,6	6,3	54,4	48,1	14,5	0,9	0,6
<b>Pronomes demonstrativos</b>	9,6	2,5	19,9	17,4	5,6	0,6	-0,5
<b>Quantidade de palavras</b>	517,8	313,0	901,0	588,0	182,8	1,2	0,5
<b>Substantivos</b>	215,4	161,5	314,9	153,4	44,6	1,2	1,1
<b>Verbos públicos</b>	6,0	0,0	14,7	14,7	5,4	0,6	-1,2
<b>Verbos / pronomes 1ª pessoa singular</b>	24,7	1,3	75,5	74,3	21,1	1,4	2,2
<b>Verbos / pronomes 3ª pessoa singular</b>	88,7	44,0	124,3	80,3	25,0	-0,4	-0,6
<b>Voz passiva</b>	1,9	0,0	9,6	9,6	3,1	1,8	2,7

TABELA 19 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO EDITORIAL (N = 20; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
Adjetivos	94,0	66,0	135,1	69,1	19,7	0,3	-0,3
Advérbios	137,1	67,5	224,4	156,8	32,7	0,4	2,1
Aposições	2,9	0,0	13,3	13,3	3,9	1,5	1,6
Aspas	5,3	0,0	32,1	32,1	7,6	2,6	8,0
Conjunções coordenativas	23,6	10,8	41,5	30,7	9,9	0,5	-1,1
Conjunções subordinativas	14,9	3,4	28,8	25,3	6,2	0,6	0,6
Futuro do pretérito	4,3	0,0	16,1	16,1	4,8	1,0	0,3
Nomes próprios	37,2	11,9	69,7	57,8	15,8	0,5	0,6
Números cardinais	18,0	0,0	48,3	48,3	14,2	0,8	0,1
Presente do indicativo	52,1	27,3	95,8	68,6	17,6	0,7	0,4
Pretérito imperfeito	3,4	0,0	10,0	10,0	3,3	0,6	-0,8
Pretérito perfeito	12,3	0,0	29,6	29,6	8,7	0,5	-0,6
Pronomes demonstrativos	9,2	0,0	18,7	18,7	5,1	0,0	-0,2
Quantidade de palavras	338,5	268,0	541,0	273,0	73,4	2,2	4,2
Substantivos	243,8	194,0	285,2	91,2	29,0	-0,3	-1,0
Verbos públicos	1,0	0,0	3,6	3,6	1,5	0,9	-1,1
Verbos / pronomes 1ª pessoa singular	0,4	0,0	3,2	3,2	0,9	2,4	4,5
Verbos / pronomes 3ª pessoa singular	68,1	29,7	109,1	79,4	20,8	0,1	-0,5
Voz passiva	5,3	0,0	12,6	12,6	4,4	0,1	-1,2

TABELA 20 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO ENTREVISTA (N = 24; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	62,4	34,4	101,8	67,3	16,9	0,2	-0,2
<b>Advérbios</b>	153,7	93,4	198,3	104,9	25,0	-0,4	0,2
<b>Aposições</b>	3,5	0,0	7,8	7,8	2,4	0,2	-1,1
<b>Aspas</b>	7,0	0,0	25,2	25,2	5,7	1,5	3,5
<b>Conjunções coordenativas</b>	29,5	11,5	45,2	33,7	8,8	0,0	-0,6
<b>Conjunções subordinativas</b>	21,3	3,9	40,8	37,0	8,8	0,3	0,2
<b>Futuro do pretérito</b>	3,7	0,0	9,4	9,4	2,8	0,4	-0,8
<b>Nomes próprios</b>	47,8	12,8	66,9	54,1	14,3	-0,9	0,3
<b>Números cardinais</b>	14,0	5,6	41,4	35,8	8,7	1,6	3,0
<b>Presente do indicativo</b>	75,8	33,3	121,4	88,1	21,3	0,4	-0,2
<b>Pretérito imperfeito</b>	12,7	0,0	66,6	66,6	14,1	2,9	9,6
<b>Pretérito perfeito</b>	19,8	6,6	39,0	32,4	10,0	0,6	-0,9
<b>Pronomes demonstrativos</b>	14,6	5,7	24,4	18,7	5,5	-0,1	-1,2
<b>Quantidade de palavras</b>	1028,9	267,0	5009,0	4742,0	961,4	3,4	13,4
<b>Substantivos</b>	208,5	164,9	275,9	111,0	27,9	0,6	0,4
<b>Verbos públicos</b>	4,6	0,0	12,1	12,1	2,9	1,0	0,8
<b>Verbos / pronomes 1ª pessoa singular</b>	23,0	0,0	63,1	63,1	18,8	0,7	-0,6
<b>Verbos / pronomes 3ª pessoa singular</b>	84,9	54,5	136,2	81,7	19,7	0,5	0,5
<b>Voz passiva</b>	2,1	0,0	11,5	11,5	2,5	2,5	8,6

TABELA 21 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO NOTA DE CORREÇÃO (N = 7; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	45,2	21,6	74,5	52,8	16,2	0,7	1,7
<b>Advérbios</b>	108,0	48,8	138,3	89,5	32,3	-1,2	0,7
<b>Aposições</b>	38,9	9,3	73,2	63,9	24,7	0,3	-1,8
<b>Aspas</b>	38,8	10,7	85,1	74,4	23,4	1,4	2,7
<b>Conjunções coordenativas</b>	38,2	24,4	53,2	28,8	10,3	0,0	-0,8
<b>Conjunções subordinativas</b>	1,3	0,0	9,3	9,3	3,5	2,6	7,0
<b>Futuro do pretérito</b>	0,8	0,0	5,4	5,4	2,0	2,6	7,0
<b>Nomes próprios</b>	116,2	46,3	182,9	136,6	42,4	-0,2	1,1
<b>Números cardinais</b>	45,8	21,3	82,8	61,6	25,7	0,8	-1,2
<b>Presente do indicativo</b>	31,8	5,4	64,8	59,4	21,3	0,5	-0,9
<b>Pretérito imperfeito</b>	1,5	0,0	10,6	10,6	4,0	2,6	7,0
<b>Pretérito perfeito</b>	22,3	9,3	41,1	31,8	12,9	0,6	-1,4
<b>Pronomes demonstrativos</b>	5,6	0,0	24,4	24,4	9,1	1,9	3,6
<b>Quantidade de palavras</b>	128,3	73,0	187,0	114,0	50,2	0,3	-2,4
<b>Substantivos</b>	192,9	97,6	259,3	161,7	63,7	-0,5	-1,3
<b>Verbos públicos</b>	2,0	0,0	13,7	13,7	5,2	2,6	7,0
<b>Verbos / pronomes 1ª pessoa singular</b>	0,0	0,0	0,0	0,0	0,0	-	-
<b>Verbos / pronomes 3ª pessoa singular</b>	48,3	27,8	63,8	36,1	12,3	-0,7	-0,1
<b>Voz passiva</b>	5,5	0,0	27,4	27,4	10,0	2,3	5,4

TABELA 22 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO COLUNA DE NOTAS (N = 73; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	54,5	19,3	86,1	66,8	14,3	0,1	-0,2
<b>Advérbios</b>	124,4	60,6	203,4	142,8	28,3	0,4	0,4
<b>Aposições</b>	5,4	0,0	15,3	15,3	3,7	0,6	-0,1
<b>Aspas</b>	9,9	0,0	49,6	49,6	11,0	1,7	2,8
<b>Conjunções coordenativas</b>	26,5	2,7	68,2	65,5	12,0	1,1	1,5
<b>Conjunções subordinativas</b>	9,8	0,0	30,5	30,5	6,5	0,7	0,8
<b>Futuro do pretérito</b>	2,5	0,0	19,3	19,3	3,2	2,5	9,9
<b>Nomes próprios</b>	87,4	27,2	159,8	132,6	26,6	0,2	0,3
<b>Números cardinais</b>	31,3	7,2	79,3	72,1	17,6	0,9	0,3
<b>Presente do indicativo</b>	49,2	18,6	103,3	84,6	17,1	0,7	0,4
<b>Pretérito imperfeito</b>	5,8	0,0	33,0	33,0	6,6	2,1	5,0
<b>Pretérito perfeito</b>	18,5	0,0	48,0	48,0	10,6	0,4	0,2
<b>Pronomes demonstrativos</b>	6,9	0,0	38,2	38,2	5,9	2,3	9,7
<b>Quantidade de palavras</b>	588,6	132,0	1785,0	1653,0	223,3	1,7	10,6
<b>Substantivos</b>	225,6	154,2	306,7	152,5	30,4	0,4	0,1
<b>Verbos públicos</b>	2,7	0,0	9,4	9,4	2,5	0,7	-0,2
<b>Verbos / pronomes 1ª pessoa singular</b>	3,5	0,0	29,8	29,8	5,2	2,7	9,7
<b>Verbos / pronomes 3ª pessoa singular</b>	70,2	18,3	108,5	90,2	17,4	-0,7	0,7
<b>Voz passiva</b>	3,2	0,0	16,3	16,3	2,8	1,6	5,0

TABELA 23 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO NOTÍCIA (N = 296; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	52,5	0,0	161,3	161,3	25,9	0,6	1,1
<b>Advérbios</b>	116,7	20,4	384,6	364,2	44,6	1,2	4,7
<b>Aposições</b>	6,3	0,0	82,0	82,0	9,5	2,9	14,8
<b>Aspas</b>	7,3	0,0	83,3	83,3	11,8	2,4	7,6
<b>Conjunções coordenativas</b>	23,4	0,0	89,3	89,3	16,6	0,7	0,7
<b>Conjunções subordinativas</b>	9,9	0,0	48,4	48,4	11,8	1,1	0,5
<b>Futuro do pretérito</b>	3,2	0,0	40,5	40,5	6,5	2,8	9,2
<b>Nomes próprios</b>	94,8	0,0	363,6	363,6	42,9	1,3	5,2
<b>Números cardinais</b>	34,7	0,0	230,8	230,8	31,8	2,2	7,6
<b>Presente do indicativo</b>	38,9	0,0	116,7	116,7	21,0	0,6	0,3
<b>Pretérito imperfeito</b>	5,0	0,0	76,9	76,9	8,8	3,1	16,4
<b>Pretérito perfeito</b>	24,7	0,0	104,2	104,2	18,9	0,7	0,5
<b>Pronomes demonstrativos</b>	3,9	0,0	34,5	34,5	6,0	1,8	3,4
<b>Quantidade de palavras</b>	175,1	31,0	842,0	811,0	162,1	1,7	2,4
<b>Substantivos</b>	227,2	71,4	461,5	390,1	47,8	0,1	2,2
<b>Verbos públicos</b>	4,4	0,0	48,4	48,4	7,0	2,2	6,6
<b>Verbos / pronomes 1ª pessoa singular</b>	1,4	0,0	53,6	53,6	5,5	6,2	45,9
<b>Verbos / pronomes 3ª pessoa singular</b>	70,6	0,0	187,5	187,5	29,8	0,5	0,5
<b>Voz passiva</b>	4,7	0,0	38,5	38,5	6,7	1,8	3,5

TABELA 24 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO OUTROS (N = 25; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	50,4	18,0	105,8	87,8	23,5	0,6	-0,4
<b>Advérbios</b>	121,6	0,0	175,1	175,1	36,1	-1,4	4,3
<b>Aposições</b>	3,8	0,0	33,3	33,3	7,4	3,0	10,7
<b>Aspas</b>	11,9	0,0	100,0	100,0	19,7	4,0	17,9
<b>Conjunções coordenativas</b>	33,6	0,0	74,1	74,1	18,5	0,4	0,1
<b>Conjunções subordinativas</b>	9,7	0,0	48,5	48,5	13,6	1,6	1,9
<b>Futuro do pretérito</b>	2,4	0,0	19,4	19,4	4,5	2,5	7,6
<b>Nomes próprios</b>	74,1	21,6	133,3	111,8	30,2	0,2	-0,8
<b>Números cardinais</b>	59,0	0,0	541,0	541,0	105,0	4,4	20,4
<b>Presente do indicativo</b>	41,2	0,0	118,5	118,5	32,6	0,8	-0,3
<b>Pretérito imperfeito</b>	6,5	0,0	57,6	57,6	13,5	2,8	8,4
<b>Pretérito perfeito</b>	19,9	0,0	58,8	58,8	16,2	0,7	-0,1
<b>Pronomes demonstrativos</b>	4,8	0,0	33,3	33,3	8,6	2,1	4,2
<b>Quantidade de palavras</b>	277,4	30,0	1266,0	1236,0	306,1	2,0	4,0
<b>Substantivos</b>	226,7	65,6	306,2	240,6	50,1	-1,3	3,3
<b>Verbos públicos</b>	3,0	0,0	30,3	30,3	8,0	2,8	7,3
<b>Verbos / pronomes 1ª pessoa singular</b>	6,8	0,0	66,7	66,7	14,8	3,1	11,2
<b>Verbos / pronomes 3ª pessoa singular</b>	62,6	16,4	115,4	99,0	31,7	0,2	-1,2
<b>Voz passiva</b>	4,1	0,0	26,8	26,8	7,0	2,0	3,9

TABELA 25 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO REPORTAGEM (N = 655; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	52,8	0,0	130,0	130,0	17,8	0,3	0,7
<b>Advérbios</b>	129,3	23,3	254,9	231,6	31,7	0,2	0,7
<b>Aposições</b>	5,3	0,0	51,9	51,9	5,6	2,2	10,1
<b>Aspas</b>	9,5	0,0	50,5	50,5	9,1	1,3	1,7
<b>Conjunções coordenativas</b>	27,2	0,0	102,0	102,0	13,0	1,3	4,0
<b>Conjunções subordinativas</b>	12,4	0,0	58,5	58,5	9,5	1,0	1,6
<b>Futuro do pretérito</b>	2,5	0,0	27,2	27,2	3,8	2,3	7,9
<b>Nomes próprios</b>	74,4	0,0	238,1	238,1	29,6	0,7	1,6
<b>Números cardinais</b>	28,4	0,0	102,9	102,9	18,5	0,9	0,7
<b>Presente do indicativo</b>	50,3	0,0	169,2	169,2	23,6	0,5	0,8
<b>Pretérito imperfeito</b>	6,2	0,0	58,8	58,8	8,0	2,1	5,7
<b>Pretérito perfeito</b>	23,2	0,0	105,3	105,3	17,9	1,1	1,5
<b>Pronomes demonstrativos</b>	6,6	0,0	38,5	38,5	5,8	1,4	3,3
<b>Quantidade de palavras</b>	383,1	27,0	1951,0	1924,0	202,1	1,2	5,1
<b>Substantivos</b>	234,8	132,9	470,0	337,1	36,0	0,8	3,7
<b>Verbos públicos</b>	4,5	0,0	50,5	50,5	5,5	2,3	9,8
<b>Verbos / pronomes 1ª pessoa singular</b>	3,2	0,0	50,5	50,5	6,8	3,0	10,3
<b>Verbos / pronomes 3ª pessoa singular</b>	74,7	7,7	261,5	253,8	23,9	1,1	5,8
<b>Voz passiva</b>	3,9	0,0	25,9	25,9	4,2	1,7	3,9

TABELA 26 - ESTATÍSTICA DESCRITIVA DE FREQUÊNCIA DO GÊNERO RESENHA (N = 18; POR 1.000 PALAVRAS)

Variável	Média	Mínimo	Máximo	Alcance	Desvio padrão	Assimetria	Curtose
<b>Adjetivos</b>	62,9	21,3	94,3	73,1	18,0	-0,4	0,5
<b>Advérbios</b>	116,1	41,2	164,0	122,8	30,2	-0,7	1,5
<b>Aposições</b>	6,7	0,0	25,1	25,1	7,6	1,5	1,6
<b>Aspas</b>	16,3	0,0	57,7	57,7	14,1	1,9	3,9
<b>Conjunções coordenativas</b>	41,1	21,4	68,4	47,0	13,3	0,8	0,3
<b>Conjunções subordinativas</b>	9,3	0,0	22,6	22,6	6,2	0,2	0,1
<b>Futuro do pretérito</b>	1,3	0,0	11,9	11,9	2,9	3,3	12,3
<b>Nomes próprios</b>	76,2	32,1	154,8	122,7	38,3	1,0	0,1
<b>Números cardinais</b>	34,3	11,5	130,6	119,1	31,4	2,1	4,6
<b>Presente do indicativo</b>	51,0	20,6	111,8	91,2	23,3	1,6	2,5
<b>Pretérito imperfeito</b>	8,1	0,0	29,5	29,5	9,7	1,2	0,3
<b>Pretérito perfeito</b>	11,3	0,0	35,6	35,6	11,3	1,1	0,4
<b>Pronomes demonstrativos</b>	6,9	0,0	13,6	13,6	4,5	-0,4	-1,2
<b>Quantidade de palavras</b>	455,8	84,0	884,0	800,0	231,4	-0,2	-0,9
<b>Substantivos</b>	226,8	166,7	273,8	107,1	29,2	-0,2	-0,4
<b>Verbos públicos</b>	1,6	0,0	6,2	6,2	2,0	1,1	0,1
<b>Verbos / pronomes 1ª pessoa singular</b>	2,3	0,0	14,8	14,8	4,3	2,2	3,9
<b>Verbos / pronomes 3ª pessoa singular</b>	65,4	24,1	118,0	94,0	26,7	0,5	-0,7
<b>Voz passiva</b>	2,6	0,0	9,1	9,1	2,9	1,0	-0,1

ANEXO F - Versão Integral do Texto 6<sup>128</sup>

## A SOCIEDADE LÍQUIDA

Maria Lúcia Garcia Pallares-Burke

especial para a Folha

Um renomado periódico espanhol referiu-se recentemente a Zygmunt Bauman (1927) como um dos poucos sociólogos contemporâneos “nos quais ainda se encontram idéias”. Opinião semelhante é frequentemente exposta por críticos de várias partes do mundo quando refletem sobre o pensamento desse intelectual polonês radicado na Inglaterra desde 1971 e empenhado, há meio século, em “traduzir o mundo em textos”, como diz um deles. Indiferente às fronteiras disciplinares, Bauman é um dos líderes da chamada “sociologia humanística”.

De um lado, não se encontram em suas obras abstrações ou análises e levantamentos estatísticos, e, de outro, são ali aproveitadas quaisquer idéias e abordagens que possam ajudá-lo na tarefa de compreender a complexidade e diversidade da vida humana. Essa é uma das razões pelas quais Bauman tem muito a dizer para uma gama de leitores muito maior do que normalmente se espera de um trabalho de sociologia mais convencional, o que condiz com suas próprias ambições de atingir um público composto de pessoas comuns “se esforçando por ser humanas” num mundo mais e mais desumano. Como ele gosta de insistir, seu objetivo é mostrar a seus leitores que o mundo pode ser diferente e melhor do que é.

Autor prolífico e de renome internacional, pode-se dizer que sua fama e prolixidade aumentaram significativamente após sua aposentadoria, em 1990: 16 de seus 25 livros foram publicados após essa data e cinco obras dedicadas ao estudo de seu pensamento foram escritas nos últimos anos.

Descrito certa vez como “profeta da pós-modernidade” (com o que não concorda), por suas reflexões sobre as condições do mundo da “modernidade líquida”, os temas abordados por Bauman tendem a ser amplos, variados e especialmente focalizados na vida cotidiana dos homens e mulheres comuns. Holocausto, globalização, sociedade de consumo, amor, comunidade, individualidade são algumas das questões de que trata, sempre salientando a dimensão ética e humanitária que deve nortear tudo o que diz respeito à condição humana. Preocupado com a sina dos oprimidos, Bauman é uma das vozes a permanentemente questionar a ação dos governos neoliberais que dão amplo apoio às forças do mercado ao mesmo tempo em que abdicam da responsabilidade de promover a justiça social.

Nascido na Posnânia em 1925, Bauman escapou dos horrores do Holocausto que aguardavam os judeus poloneses na Segunda Guerra, ao fugir com sua família para a Rússia em 1939. De lá voltou após a guerra, quando se filiou ao Partido Comunista, estudou na Universidade de Varsóvia e conheceu Janina, com quem está casado há 55 anos e com quem teve três filhas: Anna (matemática), Lydia (pintora) e Irena (arquiteta).

Confiantes e animados pelo sonho de criar uma sociedade mais justa e igualitária, Zygmunt e Janina ali estiveram a construir suas carreiras (ele como professor da Universidade de Varsóvia e, ela, como editora de enredos cinematográficos) e criar sua família, até que uma nova onda de anti-semitismo e repressão esmagou os seus sonhos e os forçou ao exílio. Após três anos em Israel, o convite a Bauman para ser chefe do departamento de sociologia

---

<sup>128</sup> Cf. 4.5.2.

da Universidade de Leeds os trouxe à Inglaterra, onde permanecem até hoje.

Bauman recebeu o Mais! em Leeds, na confortável casa onde mora desde que ali chegou, há mais de 30 anos. “Naquela época achei a cidade horrível, imunda”, me disse Janina, comentando a mudança que ocorreu nos últimos tempos e que transformou Leeds, de um sujo centro industrial, em uma cidade bonita, verdejante e cheia de vida.

O senhor já foi descrito como um “profeta da pós-modernidade” e os termos “pós-moderno” e “pós-modernidade” aparecem em títulos de quatro de seus livros. Estaria sugerindo que ocorreu uma mudança cultural e social significativa na última geração suficientemente grande para que falemos de um novo período da história?

Uma das razões pelas quais passei a falar em “modernidade líquida” em vez de “pós-modernidade” (meus trabalhos mais recentes evitam esse termo) é que fiquei cansado de tentar esclarecer uma confusão semântica que não distingue sociologia pós-moderna de sociologia da pós-modernidade, entre “pós-modernismo” e “pós-modernidade”. No meu vocabulário, “pós-modernidade” significa uma sociedade (ou, se se preferir, um tipo de condição humana), enquanto que “pós-modernismo” se refere a uma visão de mundo que pode surgir, mas não necessariamente, da condição pós-moderna.

Procurei sempre enfatizar que, do mesmo modo que ser um ornitólogo não significa ser um pássaro, ser um sociólogo da pós-modernidade não significa ser um pós-modernista, o que definitivamente não sou. Ser um pós-modernista significa ter uma ideologia, uma percepção do mundo, uma determinada hierarquia de valores que, entre outras coisas, descarta a idéia de um tipo de regulamentação normativa da comunidade humana e assume que todos os tipos de vida humana se equivalem, que todas as sociedades são igualmente boas ou más; enfim, uma ideologia que se recusa a fazer julgamentos e a debater seriamente questões relativas a modos de vida viciosos e virtuosos, pois, no limite, acredita que não há nada a ser debatido. Isso é pós-modernismo.

Mas sempre estive interessado na sociologia da pós-modernidade, meu tema tem sempre sido compreender esse tipo curioso e em muitos sentidos misterioso de sociedade que vem surgindo ao nosso redor; e a vejo como uma condição que ainda se mantém eminentemente moderna nas suas ambições e no seu “modus operandi” (ou seja, no seu esforço de modernização compulsiva, obsessiva), mas que se acha desprovida das antigas ilusões de que o fim da jornada estava logo adiante.

É nesse sentido que pós-modernidade é, para mim, modernidade sem ilusões. Diferentemente da sociedade moderna anterior, a que eu chamo de modernidade sólida, que também estava sempre a desmontar a realidade herdada, a de agora não o faz com uma perspectiva de longa duração, com a intenção de torná-la melhor e novamente sólida. Tudo está agora sempre a ser permanentemente desmontado, mas sem perspectiva de nenhuma permanência. Tudo é temporário. É por isso que sugeri a metáfora da “liquidez” para caracterizar o estado da sociedade moderna, que, como os líquidos, se caracteriza por uma incapacidade de manter a forma. Nossas instituições, quadros de referência, estilos de vida, crenças e convicções mudam antes que tenham tempo de se solidificar em costumes, hábitos e verdades “auto-evidentes”. É verdade que a vida moderna foi desde o início “desenraizadora” e “derretia os sólidos e profanava os sagrados”, como os jovens Marx e Engels notaram. Mas, enquanto no passado isso se fazia para ser novamente “reenraizado”, agora as coisas todas –empregos, relacionamentos, know-hows etc.– tendem a permanecer em fluxo, voláteis, desreguladas, flexíveis.

Como um exemplo dessa perspectiva, li, num dia desses, que um famoso arquiteto de Los Angeles estava se propondo a construir casas que permanecessem lindas “para sempre”.

Ao ser questionado sobre o que queria dizer com isso, ele teria respondido: até daqui a 20 anos! Isso é hoje “para sempre”, grande duração. O que me interessa é, portanto, tentar compreender quais as consequências dessa situação para a lógica do indivíduo, para seu cotidiano. Virtualmente todos os aspectos da vida humana são afetados quando se vive a cada momento sem que a perspectiva de longo prazo tenha mais sentido.

Jean-Paul Sartre aconselhou seus discípulos em todo o mundo a terem um projeto de vida, a decidir o que queriam ser e, a partir daí, implementar esse programa consistentemente, passo a passo, hora a hora. Ora, ter uma identidade fixa, como Sartre aconselhava, é hoje, nesse mundo fluido, uma decisão de certo modo suicida. Se se pensa, por exemplo, nos dados levantado por Richard Sennett [sociólogo] –o tempo médio de emprego no vale do Silício [localizado na Califórnia, EUA, concentra um grande número de empresas de tecnologia e internet], por exemplo, é de oito meses–, quem pode pensar num projeto de vida nessas circunstâncias?

Na época da modernidade sólida, quem entrasse como aprendiz nas fábricas da Renault ou Ford iria com toda probabilidade ter ali um longa carreira e se aposentar após 40 ou 45 anos. Hoje em dia, quem trabalha para Bill Gates por um salário talvez cem vezes maior não tem idéia do que poderá lhe acontecer dali a meio ano! E isso faz uma diferença incrível em todos os aspectos da vida humana.

Em “Liquid Love” [Amor Líquido], eu exploro o impacto dessa situação nas relações humanas, quando o indivíduo se vê diante de um dilema terrível: de um lado, ele precisa dos outros como do ar que respira, mas, ao mesmo tempo, ele tem medo de desenvolver relacionamentos mais profundos, que o imobilizem num mundo em permanente movimento.

O sr. poderia discutir os riscos da pós-modernidade?

Uma das características do que eu chamo de “modernidade sólida” é a de que as maiores ameaças para a existência humana eram muito mais óbvias. Os perigos eram reais, palpáveis e não havia muito mistério sobre o que fazer para neutralizá-los ou, ao menos, aliviá-los. Era, por exemplo, óbvio que alimento –e só alimento– era o remédio para a fome.

Os riscos de hoje são de outra ordem, não se podendo sentir ou tocar em muitos deles, apesar de estarmos todos expostos, em algum grau, a suas consequências. Não podemos, por exemplo, cheirar, ouvir, ver ou tocar as condições climáticas que gradativamente, mas sem trégua, estão se deteriorando.

O mesmo acontece com os níveis de radiação e poluição, a diminuição das matérias-primas e fontes de energia não-renováveis e os processos de globalização sem controle político ou ético que solapam as bases de nossa existência e sobrecarregam a vida dos indivíduos com um grau de incerteza e ansiedade sem precedentes. É nesse ponto que a sociologia tem um papel importante a desempenhar.

Diferentemente dos perigos antigos, os riscos que envolvem a condição humana no mundo das dependências globais podem não só deixar de ser notados, mas também minimizados, mesmo quando notados. Do mesmo modo, as ações necessárias para exterminar ou limitar os riscos podem ser desviadas das verdadeiras fontes do perigo e canalizadas para alvos errados. Quando a complexidade da situação é descartada, fica fácil apontar para aquilo que está mais à mão como sendo causa das incertezas e ansiedades modernas.

Veja, por exemplo, o caso das manifestações contra imigrantes que ocorrem pela Europa. Vistos como “o inimigo” próximo, eles são apontados como os culpados pelas frustrações

da sociedade, como aqueles que põem obstáculo aos projetos de vida dos demais cidadãos. A noção de “solicitante de asilo” adquire, nesse quadro, uma conotação negativa, ao mesmo tempo em que as leis que regem a imigração e naturalização se tornam mais restritivas, e a promessa de construção de “centros de detenção” para estrangeiros confere vantagens eleitorais a plataformas políticas.

Para confrontar sua condição existencial e enfrentar seus desafios, a humanidade precisa se colocar acima dos dados da experiência a que tem acesso enquanto indivíduos. Ou seja, a percepção individual, para ser ampliada, necessita da assistência de intérpretes munidos com dados não amplamente disponíveis à experiência individual. E a sociologia, enquanto parte integrante desse processo interpretativo –um processo em andamento e permanentemente inconclusivo–, constitui um empenho constante para ampliar os horizontes cognitivos dos indivíduos e uma voz potencialmente poderosa nesse diálogo sem fim com a condição humana.

Em muitas partes de sua obra o senhor soa nostálgico, às vezes até mesmo do que chama de “modernidade sólida”, quando a humanidade aparentemente era menos ansiosa e tinha uma vida mais estável e segura. Concorda com essa interpretação?

Eu não diria isso. Não acredito que haja um progresso linear no que diz respeito à felicidade humana. Podemos dizer que, como um pêndulo, nos movemos de tempos mais felizes para tempos menos felizes e de menos felizes para mais felizes. Hoje temos medo e somos infelizes do mesmo modo como também tínhamos medo e éramos infelizes há cem anos, mas por razões diferentes. A modernidade sólida tinha um aspecto medonho: o espectro das botas dos soldados esmagando as faces humanas. Virtualmente todo mundo, quer na esquerda ou na direita, assumia que a democracia, quando existia, era para hoje ou amanhã, mas que uma ditadura estava sempre à vista; no limite, o totalitarismo poderia sempre chegar e sacrificar a liberdade em nome da segurança e da estabilidade.

De outro lado, como Sennett mostrou, a antiga condição de emprego poderia destruir a criatividade humana, as habilidades humanas, mas construía a vida humana, que podia ser planejada. Tanto os trabalhadores como os donos de fábrica sabiam muito bem que eles iriam se encontrar novamente amanhã, depois de amanhã, no ano seguinte, pois os dois lados dependiam um do outro.

Bem, nada disso existe hoje. Dificilmente um outro tipo de stalinismo voltará, e o pesadelo de hoje não é mais a bota dos soldados esmagando as faces humanas. Temos outros pesadelos. O chão onde piso pode, de repente, se abrir como num terremoto, sem que haja nada no que me segurar. A maioria das pessoas não pode planejar seu futuro por muito tempo adiante. Os acadêmicos são ainda umas das poucas pessoas que têm essa possibilidade. Na maioria dos empregos podemos ser demitidos sem uma palavra de alerta. Você chama isso de nostalgia? Não sei...

A questão é que, como já disse antes, aproximando-me dos meus 80 anos, não mais acredito que possa existir algo como uma sociedade perfeita. A vida é como um lençol muito curto: quando se cobre o nariz, os pés ficam frios, e, quando se cobrem os pés, o nariz fica gelado. Mas insisto em que a sociedade que obsessivamente se vê como não sendo suficientemente boa é a única definição que posso dar de uma boa sociedade.

Quando e como o senhor abandonou o marxismo? Considera-se ainda um socialista?

Nunca abandonei Marx, apesar de minha intoxicação pelo “marxismo realmente existente” ter sido, felizmente, breve; de fato, terminou bem cedo, no momento em que o vi como era: um imenso obstáculo para a recepção e manutenção da mensagem ética de Marx –de que a qualidade de uma sociedade deve ser testada pelo critério da justiça e “fair play” que

regulamenta a coletividade humana.

Eu espero ter o direito de dizer que nunca abandonei essa crença. O mesmo se aplica ao meu socialismo, que, em meu entender, se resume à convicção de que, assim como o poder de carga de uma ponte se mede não pela força média de todos os pilares, mas pela força de seu pilar mais fraco, a qualidade de uma sociedade também não se mede pelo PIB (Produto Interno Bruto), pela renda média de sua população, mas pela qualidade de vida de seus membros mais fracos.

O socialismo para mim não é o nome de um tipo particular de sociedade. É, sim, exatamente como o postulado de Marx de justiça social, uma dor aguda e constante de consciência que nos impulsiona a corrigir ou remover variedades sucessivas de injustiça. Não acredito mais na possibilidade (e até no desejo) de uma “sociedade perfeita”, mas acredito numa “boa sociedade”, definida como a sociedade que se recrimina sem cessar por não ser suficientemente boa e não estar fazendo o suficiente para se tornar melhor...

Quando se acompanha sua carreira, o sr. parece um filósofo que, devido às condições da Polônia do pós-guerra, foi temporariamente desviado de sua vocação, voltando-se para a sociologia. Concorda com essa descrição?

Essa seria uma reconstrução justa do que realmente aconteceu e de como eu encarava a situação, mas com uma ressalva. Eu não era um filósofo profissional antes de ter me desviado para a sociologia, como você sugere; nem desejava me tornar um.

Antes de me juntar ao Exército polonês e voltar para meu país natal por essa via, eu fiz dois anos de curso universitário de física por correspondência (na Rússia, os estrangeiros não tinham permissão de viver em cidades grandes, onde havia universidades).

Lembro-me de, como tantos adolescentes, me sentir um tanto apavorado e esmagado pelos mistérios e enigmas do universo e de desejar ardentemente dedicar minha vida a desvendar esses mistérios e a solucionar esses enigmas. Meus estudos foram, entretanto, interrompidos pelo apelo das armas quando eu tinha 18 anos, para jamais serem retomados. Deixando o Exército em 1945, eu me vi novamente numa Polônia arruinada pela ocupação nazista, que se somava a um anterior legado de miséria, de desemprego em massa, de conflitos étnicos e religiosos aparentemente insolúveis e de exploração de classe brutal. Os desafios que meu país confrontava eram, pois, muito maiores do que os do resto da Europa, pois, além de reconstruir fábricas e casas destruídas, semear campos abandonados e colocar a economia de pé novamente, a Polônia exigia uma batalha exaustiva contra uma pobreza sedimentada e contra profundas divisões de classe; a abertura das oportunidades educativas também era tarefa urgente, já que até então estas haviam estado fechadas à grande maioria da nação. Eu imagino que a crença de que a sociologia poderia melhorar a vida humana ao reformar o meio social no qual esta se conduzia era parte integral do “projeto de modernidade”. Eu até mesmo diria que o projeto consistia exatamente nisso. Assim, as pessoas que estavam seriamente empenhadas em levar a sociedade a desenvolver condições mais desejáveis a fim de ser “moderna” –ou seja, mais humana e melhor estruturada para promover a felicidade e dignidade humanas– não titubeavam um instante sobre que tipo de conhecimento deveria ser mais urgentemente adquirido, dominado e colocado em prática.

Certamente teria que ser a “ciência da sociedade”, a sociologia, a disciplina que surgira para servir ao “projeto de modernidade”. Tal convicção sobre a missão da sociologia e tal fé em seu poder de realizar sua missão deve, sem dúvida, intrigar um leitor contemporâneo, mas somente porque vivemos hoje numa era diferente, quando o mantra do dia não é mais “salvação pela sociedade”; infelizmente o que se ouve agora, como homílias insistentes, é

que devemos buscar soluções individuais para problemas produzidos socialmente e sofridos coletivamente.

Como foi a experiência de viver no que o senhor descreveu como a “idade áurea”, quando as “universidades polonesas tiraram o máximo de vantagem da liberdade ganha nas batalhas do ‘outubro polonês’ [relativa abertura do regime comunista, ocorrida em 1956]”? Foi algo fascinante, diferente de qualquer outra universidade que conheci; diferente, diria, de qualquer vida universitária existente. Há situações de liberdade acadêmica praticamente sem limites, quando todos os tipos de “Weltanschauungen” [visões de mundo], estratégias de pesquisa, hierarquias de relevância e prioridades, estilos de se contar histórias se encontram, conversam e argumentam.

E há também situações em que os sociólogos se movem pelo sentido de urgência, e não somente pela necessidade de completar dissertações a tempo e assegurar uma próxima promoção; urgência de dar sua própria contribuição para a batalha por uma sociedade melhor, mais hospitaleira aos seres humanos e à sua humanidade. E também por uma vocação, uma missão de só se dedicar a isso. O que foi peculiar na situação pós-outubro polonês foi que as duas situações emergiram ao mesmo tempo e continuaram durante algum tempo a coincidir e a se fertilizar reciprocamente.

Esse tipo de combinação entre sentimento de liberdade e de propósito é uma felicidade de que a maioria dos acadêmicos contemporâneos infelizmente carece, quer eles tenham ou não consciência do que estão perdendo. Na maioria dos lugares do mundo a liberdade de expressão acadêmica é completa ou quase completa, somente limitada pelos regulamentos e regras (muitas vezes penosas e até ridículas) da carreira e de outras invenções da burocracia universitária; mas, fora isso, as escolhas são deixadas inteiramente livres para cada um.

Há, no entanto, muito pouco sentido de propósito e particularmente pouco sentido da relevância de seu próprio trabalho para o mundo fora dos muros da academia, como se todos compartilhassem da sina da filosofia lamentada por Wittgenstein, de “deixar o mundo como é”. Como muitos sociólogos americanos e também alguns europeus se queixam, os estudos sociais acadêmicos perderam qualquer ligação com a agenda pública. Parece haver poucos, se é que há algum freguês para os modelos de “boa sociedade”, que costumava ser a preocupação central e o forte da sociologia com inclinações humanísticas. As classes educadas não estão mais interessadas na tarefa de ilustração e de elevação espiritual do povo. Os intelectuais pararam, em grande parte, de se definir pela responsabilidade que têm para com “o povo”, a nação e a humanidade.

O sr. se referiu aos “muros da academia” como um obstáculo para o pensamento livre. Há alguma esperança para as universidades?

O que quer que as universidades façam, elas não conseguirão jamais pôr um fim à curiosidade humana, que talvez tenha que sair da academia para se satisfazer. Se se pensar nas limitações que a organização universitária hoje impõe ao desenvolvimento do pensamento livre, basta olhar para o que acontece com a filosofia e a sociologia tal como são praticadas nos departamentos universitários e em outros “locais de autoridade”, ou seja, os lugares em que afirmações reconhecidas como pertencentes a uma dada disciplina podem ser feitas e de onde elas devem ser expressas para serem reconhecidas como tais. Nesse quadro, pois, a filosofia e a sociologia se ligam a interesses intelectuais, estilos de pensamento e modos de argumentação bastante diferentes.

Cada uma dessas duas disciplinas acadêmicas se pretende de posse de grupos distintos de “dados primários” e os processa, interpreta, verifica e refuta de maneiras diferentes. Dominar o cânon, tanto da sociologia quanto da filosofia, e adquirir credenciais oficialmente reconhecidas e confirmadas em cada uma delas toma todo o tempo dos estudantes universitários, e competência em uma dessas disciplinas acadêmicas é raramente exigida para se adquirir o grau na outra.

Posso entender a preocupação dos sociólogos acadêmicos com a circunscrição, as barreiras e a defesa de suas possessões contra os competidores que lutam pela obtenção do dinheiro das fundações e do governo; mas o que não podemos esquecer é que essa preocupação se origina na realidade da vida acadêmica, e não na lógica da experiência humana que a sociologia é chamada a servir.

Quão difícil foi para o senhor se ajustar à cultura da Grã-Bretanha, para onde veio com mais de 40 anos?

Ajustamento nunca ocupou um lugar prioritário no meu programa de vida. Nesse campo não fui além do básico, isto é, além de aprender o idioma local e me fazer compreensível, evitando os mais crassos “faux pas”. Tal como me recordo, ao chegar à Grã-Bretanha não estava particularmente preocupado em esconder, sufocar ou erradicar minha idiossincrasia, em abandonar o que no meu modo de agir e pensar poderia parecer estranho aos nativos. Tornar-me como os outros e me dissolver no plano de fundo não me parecia tarefa nem possível nem especialmente atraente e nunca foi minha intenção.

Como eu via na época, o desafio estava em outro lugar: como revelar para os meus colegas e alunos britânicos o sentido das minhas diferenças e talvez induzi-los a achar algum interesse e uso no que era inicialmente alheio a eles.

“Ajustamento” sugere uma via de mão única. Ao contrário, eu pensava em termos de troca igualitária: o único meio de retribuir a hospitalidade dos meus anfitriões britânicos era oferecer a eles algo que não tinham ainda e não poderiam adquirir a não ser num encontro face a face com um pensamento e modo de agir alternativos; algo novo e diferente, que pudesse, eventualmente, enriquecê-los do mesmo modo que eu tenho me enriquecido com o meu encontro com o cotidiano britânico. Eu, na verdade, desejava ser aceito, mas aceito precisamente pelo que eu era, por minha dessemelhança.

Minha sorte foi que, com essa atitude, eu aterrissei e me estabeleci na Grã-Bretanha. Posso pensar em muitos países em que viver com tal atitude teria sido muito mais difícil e social e espiritualmente custoso. Se alguém deve ser um exilado ou estrangeiro, a Grã-Bretanha é o lugar certo para estar. Pode-se aí esperar boa vontade, tolerância e bastante hospitalidade, com a condição de não querer fingir que é inglês.....

Em sua obra o senhor se refere frequentemente a romances. O que acha que a literatura pode ensinar sobre a sociedade e sobre a condição humana? Mais especificamente, o senhor confessa ser Borges uma de suas grandes fontes inspiradoras. Poderia nos explicar no que um escritor que parece não tratar especificamente de questões sociais lhe é importante?

Devo começar lembrando que meus professores na Polônia nunca se preocuparam com as diferenças entre “filosofia social” e “sociologia propriamente dita”; mas, acima de tudo, eles consideravam os romancistas e poetas como seus camaradas de armas, e não como competidores e, muito menos, como antagonistas. Eu aprendi a considerar a sociologia como uma daquelas numerosas narrativas, de muitos estilos e gêneros, que recontam, após

terem primeiramente processado e reinterpretado, a experiência humana de estar no mundo. A tarefa conjunta de tais narrativas era oferecer um insight mais profundo no modo como essa experiência foi construída e pensada e, desse modo, ajudar os seres humanos na sua luta pelo controle de seus destinos individuais e coletivos. Nessa tarefa, a narrativa sociológica não era “por direito” superior a outras narrativas, pois tinha que demonstrar e provar seu valor e utilidade pela qualidade de seu produto.

Eu, por exemplo, me lembro de ganhar de Tolstói, Balzac, Dickens, Dostoiévski, Kafka ou Thomas Morus muito mais insights sobre a substância das experiências humanas do que de centenas de relatórios de pesquisa sociológica. Acima de tudo aprendi a não perguntar de onde uma determinada idéia vem, mas somente como ela ajuda a iluminar as respostas humanas à sua condição, assunto tanto da sociologia quanto das “belle lettres”.

O que aprendi com Borges? Acima de tudo, aprendi sobre os limites de certas ilusões humanas: sobre a futilidade de sonhos de precisão total, de exatidão absoluta, de conhecimento completo, de informação exhaustiva sobre tudo; sobre as ambições humanas que, no final, se revelam ilusórias e nos mostram impotentes. Lembremos, por exemplo, do conto de Borges que fala sobre o mapa: o sonho do mapa exato que acaba ficando do mesmo tamanho da própria coisa mapeada e, portanto, sem nenhuma utilidade. Não me ocorre nenhum filósofo ou sociólogo que pôde tratar de tais questões tão persuasivamente, tão convincentemente, tão espetacularmente.

Em parte isso se deve à posição muito luxuosa e mesmo invejável de nunca ter sido um acadêmico e de nunca ter estado submetido a uma disciplina. Fora dos muros da academia os romancistas desfrutavam da liberdade que é negada, por exemplo, aos sociólogos profissionais, que têm seus trabalhos avaliados pela conformidade destes com os procedimentos que definem e distinguem a profissão, e não por sua relevância humana. Quando se envia um artigo a uma revista científica para ser avaliado por um “par”, isso só tem um impacto: reduzir a originalidade ao denominador comum!

Borges nunca teve que se submeter a esse tipo de coisa. Note que os dois cientistas sociais da modernidade realmente interessantes e ainda hoje extremamente tópicos foram Karl Marx [1818-1883] e George Simmel [1858-1918], e eles têm também essa característica comum: ambos eram free-lancers e nenhum deles ensinou nas universidades!

Mas, acima de tudo, a maior vantagem da narrativa dos romancistas é que ela se aproxima da experiência humana do que a maioria dos trabalhos das ciências sociais. Elas são capazes de reproduzir a não-determinação, a não-finalidade, a ambivalência obstinada e insidiosa da experiência humana e a ambiguidade de seu significado.

O senhor tem sempre enfatizado a necessidade de todos nós “questionarmos ostensivamente as premissas de nosso modo de vida”. Teria alguma sugestão a nos dar sobre as respostas a esses questionamentos?

Maurice Blanchot [escritor e crítico francês, 1907-2003] disse certa vez, em palavras que ficaram famosas, que as respostas são a má sorte das perguntas.

De fato, cada resposta implica fechamento, fim da estrada, fim da conversa. Também sugere nitidez, harmonia, elegância; enfim, qualidades que o mundo narrado não possui. Tenta forçar o mundo numa camisa-de-força na qual ele definitivamente não cabe. Corta as opções, a multidão de sentidos e possibilidades que toda condição humana implica a cada momento. Promete falsamente uma solução simples para um busca provocada e impelida pela complexidade. Também mente, pois declara que as contradições e incompatibilidades que provocam as questões são fantasmas –efeitos de erros linguísticos ou lógicos, em vez de qualidades endêmicas e irremovíveis da condição humana.

Creio que a experiência humana é mais rica do que qualquer de suas interpretações, pois nenhuma delas, por mais genial e “compreensiva” que seja, pode exauri-la. Aqueles que embarcam numa vida de conversação com a experiência humana deveriam abandonar todos os sonhos de um fim tranquilo de viagem. Essa viagem não tem um final feliz –toda sua felicidade se encontra na própria jornada.

O senhor descreveu modestamente um de seus livros mais recentes como um “discussion paper”. Diria que é por acaso ou propositadamente que tem se dedicado a escrever ensaios? No curso de meio século de estudos e de escrita nunca consegui adquirir a habilidade de terminar um livro... Com o passar do tempo, eu reconheço que todos os meus livros foram entregues ao editor inacabados. Em regra, antes mesmo que o manuscrito seja impresso, fica claro para mim que o que me parecia havia pouco como “o fim” era, de fato, um começo com uma sequência desconhecida, mas tremendamente necessária. Por trás de cada resposta percebo que novas questões estão piscando; que mais, muito mais, restou a ser explorado e compreendido e quão pouco, de fato, foi revelado pela “acabamento bem-sucedido” das explorações passadas. As perguntas mais intrigantes e provocantes emergem, via de regra, após as respostas.

No decurso dos anos aprendi a apreciar a queixa de Adorno [filósofo alemão, 1903-69] sobre a convenção linear da nossa escrita: por causa dessa convenção nós não conseguimos transmitir a lógica do pensamento que, diferentemente da escrita, se move em círculos e está invariavelmente forçada, pelo seu próprio progresso, a fazer perpétuos retornos.

Maria Lúcia Garcia Pallares-Burke é professora aposentada da USP e pesquisadora associada do Centro de Estudos Latino-Americanos da Universidade de Cambridge (Reino Unido). É autora de “As Muitas Faces da História” (ed. Unesp).