

O GLOBO X O DIA: UMA ANÁLISE LÉXICO-GRAMATICAL DA NOTÍCIA BASEADA NA LINGUÍSTICA DE CORPUS

Márcia Regina Alves Ribeiro Oliveira¹

Resumo: Esta pesquisa pretende discutir como as escolhas lexicais feitas por jornais populares e tradicionais deixam transparecer suas diferenças de conteúdo e focos. Para isso, apresenta uma análise contrastiva entre dois jornais cariocas: o tradicional "O Globo" e o popular "O Dia". A metodologia é baseada no ferramental da Linguística de Corpus. O corpus de análise consiste em um milhão de palavras e foi coletado no período de uma semana, no ano de 2008. Os resultados sugerem que as diferenças entre um jornal popular são marcadas em seus léxicos e que nem sempre as regras canônicas do jornalismo são seguidas por essas publicações.

Palavras-chave: Linguagem Jornalística; Linguística de Corpus; Análise de Corpus.

Abstract: This research seeks to discuss whether the lexical choices made by popular tabloids and quality newspapers mirror their differences in content and focus. To this end, it presents a contrastive analysis between two Brazilian newspaper ("O Globo" and "O Dia") based on Corpus Linguistics tools. The corpus of analysis consists of 1 million words, collected from the Internet, during a one-week period in 2008. The results suggest that the differences between popular and the traditional papers are clearly signaled in the lexical choices, highlighting what appears to be features which transcend the languages of each newspaper.

Keywords: Journalism Language; Corpus Linguistics, Corpus Analysis.

1. Jornal Popular X Jornal Tradicional

O presente trabalho tem como objetivo realizar uma análise contrastiva entre a linguagem de um jornal popular² e um jornal tradicional. Foram tomadas como objeto de análise notícias publicadas nos jornais *on-line* cariocas "O Dia" e "O Globo", que representam uma publicação popular e um periódico de referência, respectivamente.

Para realizar uma análise desses jornais, é relevante conhecer suas respectivas audiências. O público do jornal "O Dia" se concentra nas classes B, e, principalmente, na C. O periódico possui leitores residentes em áreas menos privilegiadas do Rio de Janeiro, tais como Baixada Fluminense e Leopoldina. Enquanto um jornal popular, sua missão é "*falar de perto com o carioca*", tratando primordialmente de temáticas relacionadas ao cotidiano e à

¹ Mestranda em Linguística na Universidade do Estado do Rio de Janeiro, bolsista Capes e orientanda da Professora Tania Shepherd. Bacharel em Comunicação Social (Jornalismo) pela Pontifícia Universidade Católica do Rio de Janeiro, atualmente se dedica a desenvolver pesquisas sobre temas relacionados à linguagem do jornalismo eletrônico.

² Neste artigo, entende-se por jornal popular aquele que se destina a setores populares da população e que se difere dos jornais tradicionais pela seleção de notícias, enquadramentos (*framing*) e fontes, que atendem aos interesses informativos desses setores.

prestação de serviços³. O jornal “O Globo” é atualmente o terceiro mais lido no país, concentrando 54% de seus leitores na cidade do Rio de Janeiro. Seu público se localiza nas classes A e B, possui nível de escolaridade superior e é, predominantemente, feminino. Sua missão é: “*esclarecer o que acontece de mais importante no Brasil e no mundo...*” e o jornal se descreve como “*... uma ferramenta de acesso ao melhor do entretenimento e cultura*”⁴.

O principal objetivo do estudo é identificar de que maneira as diferenças na abordagem dos fatos, no conteúdo e na qualidade transparecem no léxico. Além disso, pretende-se verificar se os periódicos analisados apresentam, também através de suas escolhas lexicais, coerência no que diz respeito às linhas editoriais que se propõem a seguir, indicadas através da missão e/ou o objetivo de cada empresa jornalística avaliada.

Em resumo, esta pesquisa busca investigar as diversas nuances da linguagem jornalística a partir do instrumental da Linguística de *Corpus*, que estuda a linguagem por meio de grandes quantidades de dados, reunidos em *corpora*⁵ (Cf. B. Sardinha, 2004).

1.1 Estudos anteriores

Embora relevantes para os estudos da linguagem em uso e para avaliação da evolução e produção do mercado jornalístico brasileiro, o estudo contrastivo da linguagem veiculada em jornais populares e de referência é um tema praticamente inexplorado. A maior parte dos trabalhos publicados se refere quase que exclusivamente à crítica ou à valorização da imprensa popular e se insere, assim como os demais trabalhos voltados para análise da linguagem jornalística, nas áreas de Semiologia (GILLESPIE, Maria; TOYNBEE, 2006) e Comunicação Social (GARCIA, 1992; DI FRANCO, 1996; PENA, 2008). Barthes (1964) foi um dos primeiros estudiosos a analisar a abordagem de fatos noticiosos realizada pela mídia, e, em especial, pelos jornais populares, que, na época, eram sinônimo de publicações sensacionalistas. O semiólogo francês introduziu a expressão *fait-divers*⁶ para

³ De acordo com informações publicadas no site do Jornal O Dia em <http://odia.terra.com.br/comercial/midiakit> (acesso em 15 de maio de 2008).

⁴ De acordo com informações publicadas no site do Jornal O Globo, disponíveis em <http://www.infoglobo.com.br> (acesso em 15 de maio de 2008).

⁵ Coletâneas de textos de linguagem natural, oral ou escrita, capazes de serem processadas por computador.

⁶ *Fait-divers* é uma expressão francesa que pode ser traduzida para o português por “fatos diversos”, conceito de teoria da comunicação jornalística que designa fatos inusitados e pitorescos.

designar notícias diversas, geralmente, sensacionalistas: notícias de catástrofes, acidentes, casos de polícia e de assuntos do cotidiano, que não apresentam grande relevância.

No Brasil, um dos estudos mais completos sobre jornalismo popular foi realizado há 12 anos por Serra (1986). O autor se apoia no conceito de *fait-divers* de Barthes para desenvolver uma crítica ao caráter sensacionalista dos jornais populares e também tradicionais. Uma das constatações mais relevantes de Serra é a de que os jornais “cometem crimes” pelo simples fato de noticiá-los em detrimento de outras questões mais relevantes e de interesse público.

Na área de estudos da linguagem, temas relacionados ao jornalismo normalmente recebem um enfoque mais voltado para a Análise do Discurso, como os estudos de textos de comunicação de Maingueneau (2002), investigação de neologismos, gêneros e variações linguísticas, com destaque para um estudo a respeito das marcas de oralidade do jornalismo popular realizado por Dias (2003).

Em consultas realizadas a periódicos acadêmicos de grande circulação no Brasil, não foram localizados trabalhos voltados para a análise contrastiva de *corpora* jornalísticos através da Linguística de *Corpus*, o que confere a esta pesquisa um caráter exploratório.

2. Enfoque metodológico

Este estudo apresenta um caráter empírico, probabilístico e predominantemente indutivo. Ele foi desenvolvido a partir do instrumental da Linguística de *Corpus*, que se baseia na observação de dados linguísticos autênticos, reunidos em *corpora* capazes de serem processados por computador. A escolha desta metodologia se deve à possibilidade de se trabalhar com uma grande quantidade de textos digitalizados e deles extrair, através de ferramentas computacionais específicas, frequências, padrões e fraseologias próprias da linguagem (neste caso, a linguagem jornalística) de cada publicação, que, posteriormente, foram interpretados com o objetivo de se chegar a resultados qualitativos. Além disso, estudos baseados em *corpora* permitem uma visão ampla do que está sendo estudado. Dessa forma, olhar os *corpora* construídos a partir de textos de jornais permite identificar suas principais características como um todo, não limitando as conclusões da pesquisa a meros estudos de caso.

A consistência dos dados linguísticos processados pelo ferramental da Linguística de *Corpus*, assegurada pela eficiência do processamento computacional, também é relevante para

a escolha do método e obtenção de resultados precisos, pois permite um afastamento do estudo linguístico intuitivo.

Neste trabalho, foi extraída uma coletânea de notícias dos jornais cariocas “O Dia” e “O Globo”, disponível em <http://www.odia.com.br> e <http://www.oglobo.com.br>, respectivamente. As notícias em questão foram publicadas nos *sites* dos jornais durante um período de sete dias (entre os dias 15 e 22 de maio de 2008).

Foi, então, composto um *corpus* de tamanho médio⁷, totalizando 1.001.300 palavras, divididas da seguinte maneira entre os dois jornais: 492.594 palavras de “O Globo” e 508.706 palavras de “O Dia”. É válido ressaltar que a diferença notada no número de palavras entre os dois periódicos não afeta os resultados desta pesquisa, pois se trata de um estudo baseado em probabilidade. As editorias escolhidas em ambos os jornais foram exatamente as mesmas nos dois jornais, a fim de se obter uma comparação mais coerente entre os dois periódicos, eliminando variáveis temáticas. São elas: “Cidade”, “País”, “Política”, “Cultura”, “Esportes”, “Ciências”, “Turismo”, “Tecnologia” e “Plantão”. As notícias foram coletadas a partir de um *software* escrito na linguagem *Perl*, desenvolvido especificamente para este trabalho, que funciona integrada com a tecnologia *RSS* (Really Simply Syndication), voltada para a captação de textos em massa via Internet. Após o período de coleta e armazenamento de dados, os textos foram processados pelo software *Wordsmith Tools*⁸, que permite a extração de dados linguísticos a partir de três ferramentas: um Listador de Palavras⁹, um Concordanciador¹⁰ e um Gerador de Palavras-chave¹¹.

É importante ressaltar que o software *Wordsmith Tools* não realiza uma análise definitiva do *corpus*. Os dados linguísticos precisam ser interpretados posteriormente pelo pesquisador. Os *corpora* também foram submetidos ao etiquetador morfossintático *on-line* chamado *Tree-Tagger*¹², que insere automaticamente determinados códigos de marcação

⁷ SARDINHA (2004) sugere que um *corpus* de tamanho médio possui entre 250 mil e 1 milhão de palavras.

⁸ Mais informações em <http://www.lexically.net/wordsmith>

⁹ Propicia a criação de listas de palavras de um determinado *corpus*

¹⁰ Gera concordâncias ou listagens das ocorrências de um item específico (palavra de busca ou nóculo) acompanhado do texto ao seu redor.

¹¹ Permite a seleção de itens de uma lista de palavras (ou mais) por meio da comparação de suas frequências com uma lista de referências. O resultado do contraste é uma lista de palavras-chave, ou palavras cujas frequências são estatisticamente diferentes no *corpus* de estudo e no *corpus* de referência.

¹² Mais informações em <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger>

gramatical de cada palavra no *corpus*. Essa ferramenta possibilita que as frequências de cada classe gramatical sejam quantificadas no programa *Wordsmith Tools*.

A aplicação das ferramentas da Linguística de *Corpus* resultou nas seguintes etapas da pesquisa:

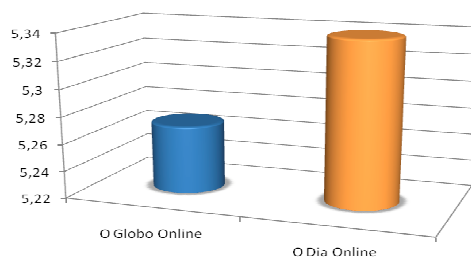
Etapa 1	Os <i>corpora</i> foram compilados: as coletâneas de textos foram reunidas e formatadas de maneira a serem interpretados por ferramentas computacionais.
Etapa 2	Foram extraídas as listas de palavras de cada <i>corpus</i> analisado através do programa <i>Wordsmith Tools</i> . Estas listas forneceram, para cada <i>corpus</i> , as palavras isoladas que os compõem, apresentadas em ordem de frequência. O processamento dos dados mostrou algumas estatísticas relacionadas a cada <i>corpus</i> , tais como números de itens (ocorrências), número de formas (vocábulos) e a densidade lexical (qualidade do conteúdo estabelecida em porcentagem) observada nos textos.
Etapa 3	De posse das listas de palavras, foi utilizada como <i>corpus</i> de referência uma coletânea de textos do jornal “Folha de São Paulo”, para que então fosse realizada a extração de palavras-chave por jornal. As listas de palavras-chave revelam as palavras que aparecem com significativa frequência no <i>corpus</i> de análise em questão. Também foram extraídas as palavras-chave positivas e negativas de “O Globo”, que são, respectivamente, as palavras mais características deste <i>corpus</i> e as palavras mais típicas do <i>corpus</i> de “O Dia”.
Etapa 4	Foram extraídos os trigramas e quadrigramas (agrupamentos multipalavras) de cada periódico.
Etapa 5	Foram detectadas diferenças semânticas no léxico de ambas as publicações, assim como foi analisado o uso dos verbos <i>dicendi</i> (declarativos ou de elocução), o que exigiu um trabalho manual dos dados já processados.
Etapa 6	Foram analisadas as classes de palavras e destacadas as mais recorrentes para estudo e aprofundamento.
Etapa 7	Passou-se para a etapa de descrição de resultados e das conclusões preliminares.

3. Resultados

A seguir, serão expostos alguns resultados obtidos em uma etapa inicial deste estudo, voltado para uma análise lexical e contrastiva dos dois jornais: o popular “O Dia” e o tradicional “O Globo”. Eles refletem a análise e interpretação de dados linguísticos quantitativos fornecidos pelo ferramental da Linguística de *Corpus*.

Ao constatar a densidade lexical de ambos os jornais, verificou-se que o jornal popular “O Dia” apresenta uma porcentagem ligeiramente maior que o jornal tradicional “O Globo”.

Gráfico 2 - Densidade lexical



Para verificar se esta diferença percentual é efetivamente significativa, os *corpora* foram lematizados e, em seguida, os lemas foram individualmente contabilizados. A quantidade destes foi também proporcionalmente analisada de acordo com o tamanho de cada *corpus*, e a diferença constatada entre ambos foi de apenas 0,02%. Isso significa, portanto, que as densidades lexicais dos dois periódicos são praticamente iguais, o que significa que este não é um quesito diferenciador entre o jornal popular e o jornal impresso analisados.

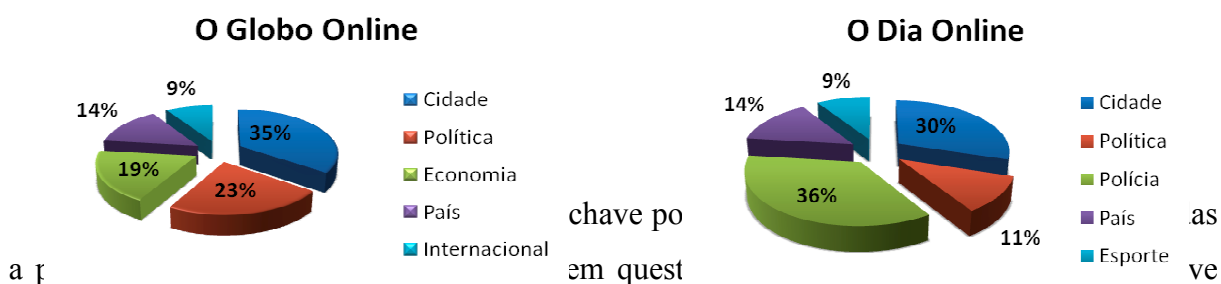
Partindo da missão ou objetivo divulgado por cada uma das empresas jornalísticas, foi realizado um agrupamento de palavras-chave de cada jornal por campos semânticos, a fim de se obter as temáticas mais frequentes de cada publicação, posteriormente separadas por editorias. Após a definição destas temáticas, elas foram comparadas com as propostas de linha editorial seguidas por cada uma das empresas, a fim de verificar se efetivamente eram coerentes com o conteúdo veiculado no cotidiano de seus diários.

O jornal “O Globo” prega, em sua missão, que “*além de esclarecer o que acontece de mais importante no Brasil e no mundo, também é uma ferramenta de acesso ao melhor do entretenimento e cultura*”. No entanto, o que se percebe após o agrupamento semântico de palavras-chave deste jornal é que se trata de uma publicação predominantemente focada em assuntos relacionados à editoria “Cidade”, sendo esta seguida pelo destaque aos conteúdos políticos e econômicos. Assuntos relacionados à cultura, que foram ressaltados em sua missão como um dos pontos fortes de “O Globo”, possuem uma presença bastante reduzida na publicação, assim como temas da editoria “Internacional”, que também são citados em sua missão.

O jornal “O Dia”, por sua vez, divulga que seu compromisso com o público é “*falar de perto com o carioca*”, o que permite supor que este seja um periódico com temática voltada para assuntos relacionados à editoria “Cidade”. Após o agrupamento de palavras-chave de “O Dia” em eixos temáticos, verificou-se a presença de notícias relacionadas à editoria “Polícia”, sendo esta seguida, então, por assuntos que dizem respeito à cidade do Rio de Janeiro.

A interpretação das palavras-chave de “O Dia” e seu agrupamento temático também permitem constatar que, apesar de este ter passado por uma reformulação editorial no início dos anos 90 com o objetivo de mudar seu enfoque à violência, o jornal continua dando ênfase a este tema em suas páginas.

Gráfico 3 - Palavras-chave agrupadas por campos semânticos



negativas são as mais frequentes na outra publicação em questão. Por exemplo: as palavras-chave negativas do jornal “O Globo” são as que aparecem com mais frequência em “O Dia”.

A partir das palavras-chave positivas, podem-se também identificar temáticas mais recorrentes em cada jornal, conforme listagem a seguir.

Tabela 3 - As palavras-chave positivas são as mais representativas em "O Globo", enquanto as negativas desempenham o mesmo papel em "O Dia".

O Globo Online	
+ POSITIVAS +	- NEGATIVAS -
Trânsito	Cativeiro
Presidente	Assassino
Bush	Jatobá
Hezbollah	Madrasta
Preços	Pagamento
Primeiro	Prisão
Alckmin	Sérgio
Valor	Polícia
Barril	Bandidos
Chaves	Favela
Trigo	Atriz
Gerda	Alerj
Lucro	Samba
Jogo	Novela
Atacante	Faetec
Crise	Isabella
Euro	Ronaldo
Embraer	Preso

Partindo para a análise de fraseologias (emprego típico de feixes lexicais) de cada diário, foram extraídos e analisados os trigramas e quadrigramas, que aparecem com frequência e são típicos de cada *corpus*.

Tabela 4 - Os números indicados entre parênteses mostram o número de ocorrências do feixe lexical em cada corpus

Feixes Lexicais	O Globo Online	O Dia Online
Trigramas	de # anos (301)	as informações são (701)
	de acordo com (276)	de acordo com (403)
	de r # (276)	de # anos (352)
	cerca de # (210)	de r # (344)
	de São Paulo (183)	informações são da (304)
Quadrigramas	mais de # (175)	Rio de Janeiro (288)
	Rio de Janeiro (167)	informações são do (278)
	em ### às #h#m (2.010)	as informações são da (304)
	por # a # (124)	do rio de janeiro (158)
	visite o site da (108)	de acordo com a (139)
	de acordo com o (105)	no dia # de (127)
	Inácio Lula da Silva (101)	das #h às #h (120)
	Luiz Inácio Lula da (100)	DE # A # (91)
	Presidente Luiz Inácio Lula (89)	Policiais do # BPM (76)
	Total	Trigramas: 6089 Quadrigramas: 2334

A partir da porcentagem contabilizada de trigramas e quadrigramas em ambos os jornais, percebe-se que “O Dia” repete uma quantidade maior de padrões lexicais que “O Globo”, embora a diferença entre eles não seja tão grande. Assim, pode-se dizer que o jornal “O Dia” aparenta ter um léxico mais padronizado que “O Globo”. Ao analisar as fraseologias, nota-se também a presença de marcadores de vagueza em “O Globo”. Tais traços não são desejáveis à linguagem jornalística de acordo com o Novo Manual da Redação (1992), que necessita ser clara e precisa para a transmissão correta da notícia. Dois exemplos de vagueza percebidos no jornal “O Globo” são as 210 ocorrências do trigrama “cerca de #¹³” e as 175 ocorrências de “mais de #”.

Entre os quadrigramas de “O Globo”, percebe-se uma referência constante ao Presidente Luiz Inácio Lula da Silva. Isso sugere que a temática política é uma das principais do periódico, conforme constatado em etapa anterior do desenvolvimento do trabalho, durante agrupamento das palavras-chave de acordo com eixos temáticos.

No jornal “O Dia”, é possível detectar através dos trigramas “informações são da” (304 ocorrências) e “informações são do” (278 ocorrências) que muitas das informações publicadas em suas páginas provêm de outras fontes. Também em “O Dia”, o quadrigrama “Policiais do # BPM” (76 ocorrências) reforça a sua temática policial recorrente, também identificada em etapa anterior do estudo em que foi realizado o agrupamento de palavras-chave em grupos temáticos.

Os verbos *dicendi* são também relevantes para análise neste trabalho, pois mostram como a “voz do jornal” transparece na notícia. Tanto em “O Globo” quanto em “O Dia” percebe-se a predominância dos verbos “dizer” e “afirmar” entre os verbos *dicendi*. Considerados verbos neutros, de acordo com o Novo Manual da Redação (1992), ambos, no entanto, carregam uma carga semântica diferenciada, pois “afirmar” é muito mais assertivo do que simplesmente “dizer”. Ao olhar o quadro a seguir, nota-se que o verbo “afirmar” é ainda mais frequente do que o verbo “dizer” nos dois jornais. Assim, é possível verificar que, pela carga semântica que o verbo “afirmar” carrega, os diários tentam passar uma maior responsabilidade das declarações para seus autores.

Gráfico 5 - Verbos *Dicendi* “Dizer” e “Afirmar” em relação ao total de verbos do corpus

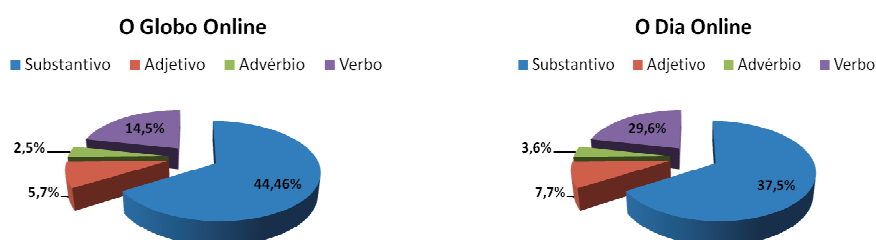


Outro aspecto relevante relacionado aos verbos *dicendi* está relacionado à utilização em larga escala do discurso indireto em ambos os diários. Esta constatação permite concluir que, na maioria das vezes, os jornais (em especial “O Globo”) reescrevem o que lhes foi dito por um entrevistado, o que pode dar margem a edições que alterem seu conteúdo, assim como a escolhas lexicais que provoquem a modificação de sentido da declaração concedida.

Por fim, foram analisadas as classes de palavras presentes nas notícias dos *corpora* em questão.

Gráfico 6 - Adjetivos e advérbios verificados em “O Dia” afetam a objetividade

¹³ O carácter “#” é utilizado pelo programa *Wordsmith Tools* para marcar a presença de números no corpus.



Seguindo uma tendência já verificada na Língua Portuguesa (SARDINHA, 2004), a maior parte das palavras encontra-se na classe dos substantivos, tanto em “O Dia” quanto em “O Globo”.

Normalmente, o que o leitor espera de uma notícia é que ela seja sustentada por fatos, sofrendo a menor interferência possível por parte do jornalista que a redigiu. Mas, surpreendentemente, os adjetivos, que são por natureza uma marca subjetiva impressa pelo autor ao texto, apareceram como a segunda classe gramatical mais frequente nos dois jornais.

Os advérbios (com exceção dos advérbios de tempo e lugar) vieram em terceira posição e, na redação jornalística, também convém economizar no uso desta classe, que, assim como a dos adjetivos, pode comprometer a neutralidade do texto.

Observou-se que especificamente “O Dia” possui uma quantidade 7% menor de substantivos e 2% maior de adjetivos em relação a “O Globo”. Isso aponta para uma possibilidade de que as notícias veiculadas neste jornal apresentem uma abordagem mais subjetiva dos fatos que “O Globo”.

4. Conclusão

O presente estudo mostra indícios de que as diferenças léxico-gramaticais entre um jornal popular e um jornal tradicional são perceptíveis através de análise quantitativa nos *corpora* em questão. Os resultados sugerem que as regras canônicas do jornalismo (objetividade, imparcialidade, uso restrito de adjetivos e advérbios) sugeridas pelo Novo Manual da Redação (1992) são frequentemente infringidas tanto pelo jornalismo de referência quanto pelo popular.

De forma geral, pode-se dizer que o olhar microscópico sobre os dados proporcionado pelos recursos da Linguística de *Corpus* permite uma observação mais apurada da linguagem jornalística, o que resulta em constatações mais precisas e objetivas. Pode-se afirmar que um trabalho como este seria uma tarefa impossível sem o ferramental da Linguística de *Corpus*,

pois a determinação de padrões e particularidades da linguagem revelados pelo processamento dos dados por computador torna-se um trabalho manualmente inviável em um *corpus* de tamanho médio, como o que foi trabalhado nesta pesquisa.

Referências

- AMARAL, Márcia Franz. **Jornalismo Popular**. São Paulo: Contexto, 2006.
- BARTHES, Roland. **Essais critiques**. Paris: Editions du Seuil, 1964.
- DI FRANCO, Carlos Alberto. **Jornalismo, ética e qualidade**. Petrópolis: Vozes, 1996.
- DIAS, Ana Rosa Ferreira. **As marcas da oralidade no jornalismo popular**. São Paulo: Cortez, 2003.
- GILLESPIE, Marie.; TOYNBEE, Jason. **Analysing Media Texts**. Berkshire: Open University Press, 2006.
- MAINGUENEAU, Dominique. **Análise de textos em comunicação**. São Paulo: Cortez, 2002
- _____. **Novo manual da redação**. 9ª ed. São Paulo: Folha de São Paulo, 1992.
- PENA, Felipe. **Teoria do Jornalismo**. 2 ed. São Paulo: Contexto, 2008.
- ROSA, Ana Rosa Ferreira. **O discurso da violência: as marcas da oralidade no jornalismo popular**. São Paulo: Cortez, 1996.
- SARDINHA, Tony Berber. **Linguística de Corpus**. Barueri: Manole, 2004
- SERRA, Antônio. **O Desvio Nosso de Cada Dia: A representação do cotidiano num jornal popular**. Rio de Janeiro: Dois Pontos, 1986.