

Jornal Popular X Jornal Tradicional: Análise léxico-gramatical da notícia a partir da Linguística de Corpus

Um estudo de casos dos jornais cariocas “O Globo” e “O Dia”

Márcia Regina Alves Ribeiro Oliveira

RESUMO: Esta pesquisa pretende discutir como as escolhas lexicais feitas por jornais populares e tradicionais transparecem 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 utilizada 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 claramente marcadas em seus léxicos e que nem sempre as regras canônicas do jornalismo são, na prática, seguidas por essas publicações.

Palavras-chave: jornalismo; linguagem jornalística; análise lexical; Linguística de Corpus

Introdução

1.1 Contextualizando o jornalismo popular e tradicional

Jornais populares¹ se tornaram um fenômeno nos últimos quinze anos no mercado editorial brasileiro, acompanhando uma tendência mundial de lançamento de jornais compactos. Com preços baixos, planejamento gráfico atraente, linguagem acessível e anúncios de produtos e serviços voltados o público de baixa renda, estes periódicos conquistaram novas audiências, que até então não tinham acesso nem o hábito de leitura de publicações diárias.

Para se ter idéia da expansão sofrida por este mercado, alguns dos jornais voltados ao consumidor de menor poder aquisitivo que surgiram após o ano de 1997 foram: Extra (RJ), Agora São Paulo (SP), Folha de Pernambuco (PE), Primeira Hora (MS), Notícia Agora (ES), Expresso Popular (SP) e Diário Gaúcho (RS). Outros, mais antigos, conseguiram sobreviver nesse mercado, como é o caso de O Dia (RJ), a Tribuna do Paraná (PR), o Jornal da Tarde (SP) e o Diário de S. Paulo (SP).

Atualmente, o mercado de periódicos populares se encontra em plena expansão: existem cerca de quinze publicações de grande circulação no país (em versão impressa e/ou eletrônica), sendo grande parte delas distribuída nos estados de Minas Gerais, Rio Grande do Sul e Rio de Janeiro, com foco nas classes C, D e E, segundo a Associação Nacional de Jornais². Este sucesso não se deve, no entanto, a uma aproximação qualitativa entre o jornalismo popular e o jornalismo de referência, mas principalmente à adoção de estratégias comerciais que aproveitaram o bom momento econômico e a maior distribuição de renda na população durante este período.

Diferente da linha editorial essencialmente sensacionalista, que era a “marca registrada” dessas publicações há cerca de vinte anos, os jornais populares sofreram um gradativo

¹ Neste artigo, entende-se por jornal popular aquele que se destina a setores populares da população e que se diferem do jornal tradicional pela seleção de notícias, enquadramentos (*framing*) e fontes, que atendem aos interesses informativos desses setores.

² Dados da Pesquisa “Quero Comprar” realizada pelo Instituto Ipsus Marplan em 2006.

reposicionamento e, atualmente, tratam prioritariamente de temas relacionados ao cotidiano de seu público (saúde, mercado de trabalho, transporte e educação), reservando parte de seu conteúdo também ao entretenimento, com destaque para notícias relacionadas a esportes e celebridades. Vale ressaltar que, apesar dessas mudanças, parte destes periódicos ainda mantém um espaço destinado aos casos policiais, histórias de interesse humano e feitos extraordinários, que pretendem despertar emoções em seus leitores da mesma maneira que seus antepassados de cunho sensacionalista o faziam. De acordo com Amaral (2006):

"A imprensa popular ligada a grandes empresas de comunicação existe pela necessidade de ampliar o mercado de consumidores de jornais para um público que vive numa situação social, cultural e econômica diferente da do público das classes A e B. Os jornais assumem formas específicas, porque o que move essa imprensa é, antes de qualquer coisa, a sedução do público e não a credibilidade ou o prestígio." (AMARAL, 2006, p.20)

Ao lado das publicações tipicamente populares, sobrevivem os jornais de referência (*quality papers*), normalmente ligados a grandes conglomerados de comunicação familiares e pioneiros no mercado editorial brasileiro. Eles continuam a servir como fonte de informação a um público predominantemente das classes A e B, sendo, portanto, destinados a parcelas minoritárias da população, que valorizam credibilidade, buscam profundidade na notícia, temas de maior abrangência, assim como a formação de opinião.

Abaixo, é possível visualizar algumas diferenças de linguagem e enfoque entre jornais populares e de referência através de suas manchetes (esta é uma das maneiras de identificar e analisar essas publicações de maneira contrastiva):

23/09/2005

| Jornais de referência | Jornais populares |
|--|--|
| Eleição de 2006 tumultua a sucessão de Severino – <i>O Globo/RJ</i> | Jeany entrega à PF agenda de festinhas – <i>O Dia/RJ</i> |
| Dantas financiou o mensalão – <i>Jornal do Brasil/RJ</i> | Bomba e pânico no condomínio – <i>Jornal da Tarde/SP</i> |
| Presidente do Conselho quer absolvição de 5 deputados – <i>Folha de S.Paulo/SP</i> | "Os filhos não sabem que o pai está preso", diz mulher de Edinho – <i>Diário de S.Paulo/SP</i> |
| Brasil quer atrair o capital externo para título público – <i>Valor Econômico/SP</i> | Leão faz blitz para cobrar dívidas antigas – <i>Agora São Paulo/SP</i> |
| 2 milhões fogem do furacão Rita – <i>Zero Hora/RS</i> | Polícia Civil assaltada pela miséria – <i>Diário Gaúcho/RS</i> |

1Manchetes de jornais de referência e jornais populares (AMARAL, 2006, p.53)

1.2. Os jornais “Globo Online” e “O Dia Online”

Visando a confrontar ambos os posicionamentos jornalísticos expostos acima, o presente trabalho tem como objetivo realizar uma análise contrastiva entre a linguagem de um jornal popular e um jornal tradicional. Assim, foram tomados 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 prestação de serviços. 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. A missão do jornal é: “*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*”.

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.

Outra questão a ser investigada diz respeito à riqueza lexical na imprensa popular e tradicional. Supõe-se, inicialmente, que o vocabulário utilizado no jornalismo popular não seja tão rico quantitativamente quanto em jornais tradicionais, tendo em vista o direcionamento de seu conteúdo a classes menos favorecidas, com leitores supostamente menos exigentes, interessados em uma leitura mais superficial.

Em resumo, esta pesquisa busca investigar as diversas nuances da linguagem jornalística a partir do instrumental da Lingüística de Corpus, que estuda a linguagem por meio de grandes quantidades de dados, reunidos em *corpora* (coletâneas de textos de linguagem natural, oral ou escrita, capazes de serem processadas por computador).

Estudos anteriores

Embora sejam de extrema importância 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* para designar notícias diversas, geralmente, sensacionalistas. São notícias de catástrofes, acidentes, casos de polícia e de assuntos do cotidiano, que despertam curiosidade mórbida, mas que não apresentam grande relevância. Segundo ele:

O *fait-divers* contém em si todo o seu saber; não é preciso conhecer nada no mundo para consumi-lo; ele não remete

formalmente a nada além dele próprio; evidentemente seu conteúdo não é estranho ao mundo; desastres, assassinatos, raptos, agressões, acidentes, roubos, esquisitices; tudo isso remete ao homem, à sua história, à sua alienação, a seus fantasmas, a seus sonhos, a seus medos. (BARTHES, 1964)

Assim, para Barthes (1964), a mídia, na busca por sua popularização, publica fatos desconectados de historicidade jornalística, que se referem apenas ao seu caráter interno e são apresentados como ocorrências inusitadas e pitorescas. Ao mesmo tempo em que os *fait-divers* atraem a atenção do leitor, eles acabam contribuindo para sua alienação em relação aos fatos históricos.

No Brasil, um dos estudos mais completos sobre jornalismo popular foi realizado por Serra (1986). O autor se apóia 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, investigação de neologismos, gêneros e variações lingüísticas, 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 Lingüística de Corpus, o que confere a esta pesquisa um caráter exploratório.

Enfoque metodológico

O presente estudo apresenta um caráter empírico, probabilístico e predominantemente indutivo. Ele foi desenvolvido a partir do instrumental da Lingüística de Corpus, que se baseia na observação de dados lingüísticos 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 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 lingüísticos processados pelo ferramental da Lingüística 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 estudo lingüístico intuitivo.

Neste trabalho, foi selecionada 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 15 e 22 de maio de 2008).

Foi, então, composto um *corpus* de tamanho médio³, 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. 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 Simple Syndication), voltada para a captação de textos em massa via Internet. Em seguida, os textos foram todos transformados para o padrão *Unicode* e salvos no formato texto simples, através de outro software também desenvolvido em *Perl*, que agilizou o processamento dos dados.

Após o período de coleta e armazenamento de dados, os textos foram processados pelo software *Wordsmith Tools*⁴, que permite a extração de dados lingüísticos a partir de três ferramentas: um Listador de Palavras⁵, um Concordanciador⁶ e um Gerador de palavras-chave⁷. É importante ressaltar que o software *Wordsmith Tools* não realiza uma análise definitiva do *corpus*. Os dados lingüísticos precisam ser interpretados posteriormente pelo pesquisador para que sejam obtidos resultados qualitativos.

Os *corpora* também foram submetidos ao etiquetador morfossintático on-line chamado *Tree-Tagger*⁸, que insere automaticamente determinados códigos de marcação gramatical de cada palavra no *corpus*.

Passou-se, então, para a etapa de interpretação dos *corpora*, cujos resultados preliminares serão apresentados na seção posterior deste trabalho.

Os conceitos fundamentais em que se baseou esta pesquisa foram identificados e revisados no Anexo 1. A seguir estão apresentadas de maneira resumida as etapas de desenvolvimento deste estudo:

Etapa 1

Os *corpora* foram compilados: as coletâneas de textos foram reunidas e formatadas de maneira a serem interpretados por ferramentas computacionais.

³ SARDINHA (2004) sugere que um corpus de tamanho médio possui entre 250 mil e 1 milhão de palavras.

⁴ Mais informações em <http://www.lexically.net/wordsmith>

⁵ Propicia a criação de listas de palavras de um determinado corpus

⁶ Gera concordâncias ou listagens das ocorrências de um item específico (palavra de busca ou nóculo) acompanhado do texto ao seu redor.

⁷ 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.

⁸ Mais informações em <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger>

| | |
|---------|--|
| 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 alfabética e por ordem de frequência. O processamento dos dados também mostrou algumas estatísticas relacionadas a cada <i>corpus</i> , tais como números de itens (<i>tokens</i> ou ocorrências), número de formas (<i>types</i> ou vocábulos) e a densidade lexical observada nos textos. |
| Etapa 3 | De posse das listas de palavras, foi utilizado 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. |
| Etapa 4 | Foram extraídos os trigramas e quadrigamas 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. |

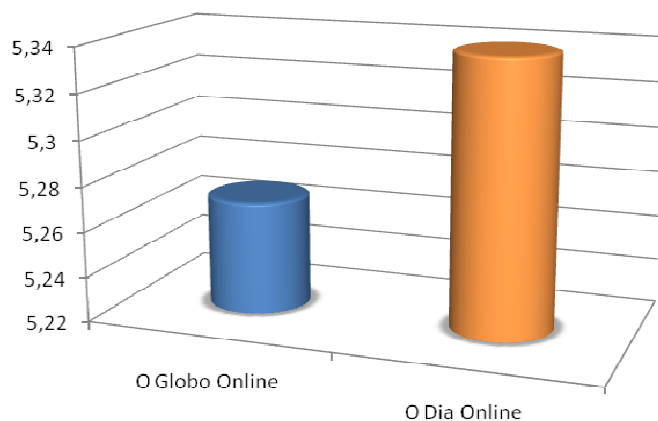
Estas etapas são, via de regra, seguidas em estudos baseados em corpora. A única exceção é a análise dos verbos *dicendi*, que se apresentam como categoria léxico-gramatical numerosa e merecem ser aprofundados em um estudo de linguagem jornalística.

Resultados

A seguir, serão expostos alguns resultados obtidos em uma etapa inicial deste estudo, voltado para uma análise lexical e contrastiva de dois jornais de grande circulação na cidade do Rio de Janeiro: o popular “O Dia” e o tradicional “O Globo”. Eles refletem a análise e interpretação de dados lingüísticos quantitativos fornecidos pelo ferramental da Lingüística de Corpus.

Uma das premissas deste estudo seria a de que um jornal popular apresentaria, em princípio, uma densidade lexical mais baixa que um jornal de referência. Tal suposição se sustenta baseando-se no fato de que a imprensa popular trata de temas majoritariamente ligados ao cotidiano das classes C, D e E, com uma linguagem mais acessível, contando com o auxílio de mais recursos gráficos para facilitar a apreensão de seu conteúdo.

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



2 Densidade Lexical

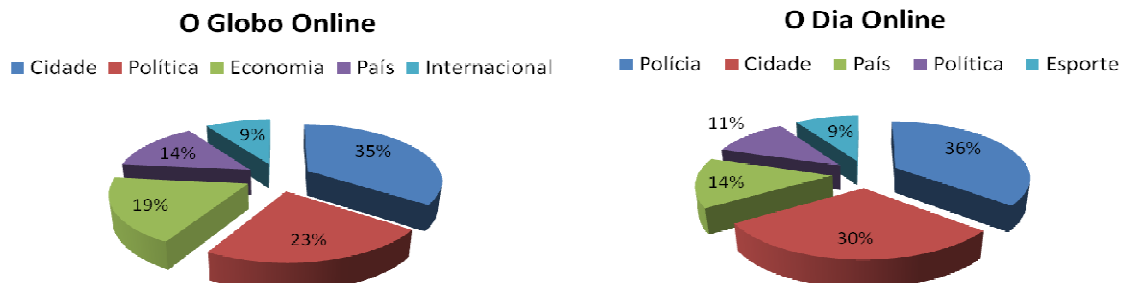
Para verificar se esta diferença percentual é efetivamente significativa, os *corpora* foram lematizados e, em seguida, os lemas foram individualmente contabilizados. A quantidade destes foi também proporcionalmente analisada de acordo com o tamanho de cada *corpus*, e a diferença constatada entre ambos 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 uma forte 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.



3 Palavras-chave agrupadas por campos semânticos

Também foram extraídas as palavras-chave positivas e negativas de cada jornal, obtidas a partir da comparação dos dois *corpora* em questão. As palavras-chave positivas são significativamente frequentes em uma publicação, ao passo que as negativas inexistem ou não aparecem com frequência relevante na mesma. Vale ressaltar que, as palavras-chave 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 pode-se também identificar temáticas mais recorrentes em cada jornal, conforme listagem a seguir.

| 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 |

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

Partindo para a análise de fraseologias (emprego típico de feixes lexicais) de cada diário, foram extraídos e analisados os trigramas e quadrigramas, que são agrupamentos de três e quatro palavras, que aparecem com frequência e são típicos de 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) |
| | MAIS DE # (175) | RIO DE JANEIRO (288) |
| | RIO DE JANEIRO (167) | INFORMAÇÕES SÃO DO (278) |
| Quadrigramas | 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 | Trigramas: 6965 |
| | Quadrigramas: 2334 | Quadrigramas: 2707 |

4 Os números indicados entre parênteses mostram o número de ocorrências do feixe lexical em cada corpus; O símbolo “#” significa que um determinado número se encontra nesta posição no textos analisados.

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, 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 trígama “cerca de #” e as 175 ocorrências de “mais de #”.

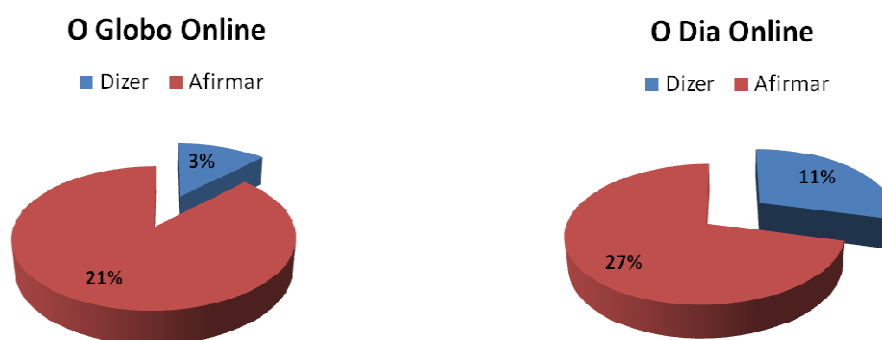
Entre os quadrigramas de “O Globo”, percebe-se uma referência constante ao Presidente Luiz Inácio Lula da Silva. Isso confirma a temática política como uma das principais do periódico, conforme constatado em etapa anterior do desenvolvimento do trabalho, durante agrupamento das palavras-chave por de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

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, ambos, no entanto, carregam uma carga semântica diferenciada, pois “afirmar” é muito mais assertivo do que simplesmente “dizer”.

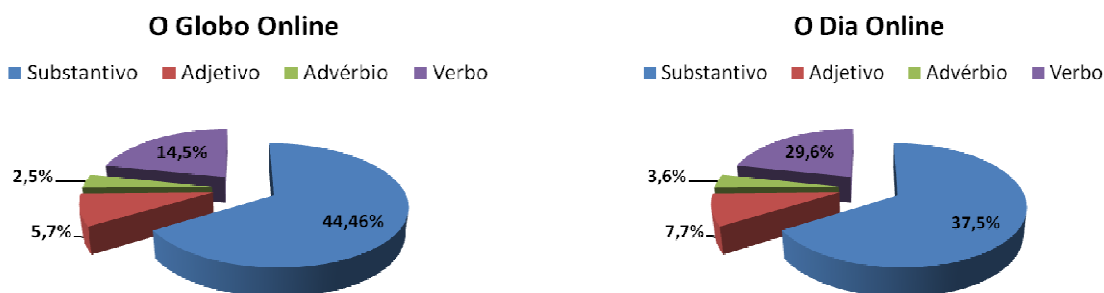
Ao olhar o quadro a seguir, nota-se que o verbo “afirmar” é ainda mais freqüente do que o verbo “dizer” nos dois jornais. Assim, é possível verificar que, pela carga semântica que o verbo “afirmar” carrega, os diários tentam passar uma maior responsabilidade das declarações para seus autores.



5 Verbos Dicendi “Dizer” e “Afirmar” em relação à quantidade total de verbos em “O Globo Online”

Outro aspecto relevante relacionado aos verbos *dicendi* diz respeito à utilização em larga escala do discurso indireto em ambos os diários. Esta constatação permite concluir que, na maior parte 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.



6 Presença de adjetivos e advérbios verificada em “O Dia” prejudica a objetividade da notícia.

Seguindo uma tendência já verificada na Língua Portuguesa, 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 gramatical classe mais freqüente.

Os advérbios 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”.

Conclusão

O presente estudo aponta indícios de que as diferenças léxico-gramaticais entre um jornal popular e um jornal tradicional são bastante significativas e que são sinalizadas de maneira de maneira bastante evidente nos *corpora* em questão.

Os resultados sugerem que as regras canônicas do jornalismo (objetividade, imparcialidade, uso restrito de adjetivos e advérbios) sugeridas por manuais de redação (são freqüentemente 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 utilizado nesta pesquisa 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. Afinal, 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 (em torno de um milhão de palavras), como o que foi trabalhado nesta primeira fase da pesquisa.

Referências

- AMARAL, M. F. *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.
- GARCIA, Luiz. *O Globo: manual de redação e estilo*. São Paulo: Editora Globo, 1992.
- GILLESPIE, Maria; TOYNBEE Jason. *Analysing Media Texts*. Berkshire: Open University Press, 2006.
- 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.

ANEXO – Conceitos da Linguística de Corpus utilizado neste estudo

| | |
|----------------------|---|
| Corpus | “Um conjunto de dados lingüísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em 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” (SARDINHA, 2004) |
| Corpora | Plural de corpus |
| Corpus de estudo | O corpus que se pretende descrever na pesquisa |
| Corpus de Referência | Também conhecido como corpus de controle. Fornece uma norma com a qual se fará a comparação das freqüências do corpus de estudo. |
| Tokens | Número de itens ou ocorrências |
| Types | Formas ou vocábulos |
| Densidade lexical | A razão forma/item (ou vocábulo/ocorrência). Corresponde à riqueza lexical de um texto ou corpus. É expressa em porcentagem. Quanto mais alta esta porcentagem, mais rico é o texto em questão em termos lexicais. |
| Feixes lexicais | São expressões recorrentes em um registro. |
| Trigrama | Feixe lexical formado por três palavras que aparecem regularmente em determinado contexto. Exemplo: “De acordo com” |
| Quadrigramas | Feixe lexical formado por quatro palavras que aparecem regularmente em determinado contexto. Exemplo: “De acordo com a” |
| Verbo dicendi | Verbos de elocução ou declarativos |

ABSTRACT

Popular X Traditional Newspapers: a lexical analysis based on Corpus Linguistics A Case study of two Brazilian newspapers

Márcia Regina Alves Ribeiro Oliveira

RESUMO: 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 newspapers (“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: Tabloid, Newspaper, Corpus Linguistics, corpus analysis, journalism