

## FREQUÊNCIAS DE VERBOS EM *CORPORA* DE JORNAIS POPULARES: DADOS PARA ATIVIDADES ENSINO COM OS JORNAIS “DIÁRIO GAÚCHO” E O “THE SUN”

### *VERB FREQUENCIES IN POPULAR NEWSPAPERS CORPORA: DATA FOR DIDACTIC ACTIVITIES WITH THE “DIÁRIO GAUCHO” AND “THE SUN”*

### *FRECUENCIAS VERBALES EN CORPORA DE LA PRENSA POPULAR: DATOS PARA ACTIVIDADES DIDÁCTICAS CON LOS DIARIOS “DIARIO GAUCHO” Y “THE SUN”*

Maria José Bocorny FINATTO<sup>1</sup>  
Aline Maciel PEREIRA<sup>2</sup>

**Resumo:** Frente à percepção de pouca repercussão das pesquisas com *corpora* nos cursos de Letras do Brasil, apresenta-se um exercício de pesquisa, de caráter didático, para observar frequências de uso de verbos em textos de jornais seguindo princípios da Linguística de Corpus. Relata-se uma observação de verbos mais empregados em amostras de textos do jornal popular *Diário Gaúcho* (DG) e do tabloide britânico *The Sun* em contraponto com usos em *corpora* gerais. Os dados levantados podem servir como ideias para atividades contrastivas inglês-português para estudantes brasileiros de Letras.

**Palavras-chave:** Linguística de Corpus; estatística lexical; uso de verbos; jornal popular brasileiro; tabloide britânico.

**Abstract:** Given the lack of repercussion of Corpus Linguistics research in Linguistics undergraduate programs in Brazilian universities, we present a didactic exercise involving the observation of the frequency of verbs in newspapers. For this study, we use text samples from the popular Brazilian newspaper *Diário Gaúcho* (DG) and the British tabloid *The Sun*, and we compare uses of verbs through Corpus Linguistics methodologies. We believe that the data collected and the methodology described may be reproduced in English and Portuguese contrastive and comparative activities developed with Linguistics undergraduate students.

---

<sup>1</sup> Docente do Programa de Pós-graduação em Letras da UFRGS. Pesquisa realizada entre 2013 e 2015 com apoios dos órgãos CNPq, CAPES, FAPERGS e SEAD-UFRGS. Pesquisadora PQ-CNPq. E-mail: mfinatto@terra.com.br.

<sup>2</sup> Acadêmica da UFRGS, Licenciatura em Língua Inglesa e Língua Portuguesa. Ex-bolsista do programa PIBIC-UFRGS e do Programa de Português para Estrangeiros (PPE) da UFRGS. Bolsista 2015-16 do Programa Inglês sem Fronteiras da UFRGS e UFCSPA. E-mail: aline\_mp\_ml@hotmail.com.

**Keywords:** Corpus linguistics; lexical statistics; uses of verbs; Brazilian newspapers; British tabloids.

**Resumen:** Dada la pequeña repercusión de la investigación lingüística basada en corpus en los programas de licenciatura en Lingüística y Letras en universidades brasileñas, presentamos un ejercicio didáctico con observación de frecuencia de verbos en periódicos populares. Para esto, utilizamos muestras de texto de un diario popular brasileño, el *Diário Gaúcho* (DG), y del tabloide británico *The Sun* y comparamos los usos de verbos a través de las metodologías de la lingüística de corpus. Creemos que los datos recogidos y la metodología descrita pueden ser provechosos en actividades contrastantes y comparativos portugués-inglés para estudiantes universitarios .

**Palabras-Clave:** Lingüística de Corpus; estadísticas léxicas; usos de verbos; periódicos brasileños; tabloides británicos.

## 1 Introdução

Quando levamos em conta o cenário pedagógico dos cursos de Letras no Brasil, seja da formação em língua materna ou em língua estrangeira, vemos que ainda pouco se exploram as potencialidades dos estudos lexicológicos, especialmente aqueles de base quantitativa advindos de qualquer viés teórico ou metodológico. No cenário da Linguística Aplicada brasileira, conforme percebemos ainda hoje, muitos dos trabalhos que lidam com *a face quantitativa da linguagem*<sup>3</sup>, especialmente os que aproveitam princípios teóricos ou mesmo apenas algumas das ferramentas da Linguística de Corpus (LC), ainda tendem a repercutir pouco entre nossos cursos de Letras, no dia a dia das salas de aula. Esse baixo impacto, que se transforma em quase nada no âmbito da escola pública em que atuarão a maioria dos nossos egressos de cursos de Letras, possivelmente relaciona-se à ainda hoje escassa presença da LC na maioria dos tópicos das atividades de ensino dos currículos (parecendo permanecer a situação já relatada por Berber Sardinha, 2010, p. 345-349).

Embora na pós-graduação em Letras haja, hoje, um impacto bastante significativo da LC em termos de número de pesquisas e de disciplinas, esses trabalhos ainda pouco conseguem chegar às salas de aula do ensino superior, e menos ainda às salas de aula do Ensino

---

<sup>3</sup> *A face quantitativa da linguagem* é justamente o título de um dos trabalhos sobre o tema no Brasil, de autoria da pioneira M.T.C. Biderman (BIDERMAN, 1998).

Médio ou Fundamental. Mas, a despeito de diferentes fatores condicionantes, a lacuna, na área da formação superior em Letras, é paradoxal frente à existência atual de obras como, por exemplo, *Corpora no Ensino de Línguas Estrangeiras* (VIANA, TAGNIN, 2010). Isso porque, nela, há várias experiências e atividades facilmente replicáveis em sala de aula – com atividades em diferentes idiomas.

Com tal percepção, trazemos aqui mais uma iniciativa de divulgação de trabalhos com *corpus*<sup>4</sup> para subsidiar o professor de línguas. Para tanto, recorreremos à apresentação de um singelo exercício inicial de pesquisa sobre verbos mais empregados em jornais populares do Brasil e da Inglaterra. A idéia é mostrar, de um modo relativamente simples, o quanto as metodologias de observação da linguagem em uso, com apoio de algumas poucas técnicas estatísticas e de ferramentas e de dados de acesso gratuito, podem ser úteis para diferentes atividades de ensino com línguas materna e estrangeiras. O foco foi a geração de dados ou de insumos para oportunidades de reflexão e de pesquisa sobre estruturas de língua e seu funcionamento. As atividades associadas à ideia o nosso exercício de estudo podem ser conjuntas, envolvendo, simultaneamente, alunos e professores de cursos de Letras e, com algumas adaptações, podem ser aproveitadas em salas de aula do ensino regular ou de cursos livres de língua inglesa. Essa dinâmica de ensino, que parte da coleta direta de dados de uso de língua, denominou-se, há bastante tempo, de *observação* em uma *abordagem lexical* do ensino de línguas (cf. LEWIS, 1997).

## **2 Alguns exemplos de trabalhos com corpora**

Antes do relato do nosso exercício, revisamos e indicamos, ao leitor interessado, dois trabalhos de maior fôlego que lidaram com *corpora* com algum contraste inglês-português. Um trabalho de Linguística e LC e um outro da área de Jornalismo. Com eles, desenhamos um cenário no qual nos inserimos como pesquisadoras.

---

<sup>4</sup> Aqui *corpus* significa uma coleção de textos, criteriosamente reunida, armazenada em formato digital, que visa representar um dado tipo de uso de língua. Essa coleção deve ser passível de exploração com apoio informatizado, conforme Berber Sardinha (2004).

## 2.1 Textos de Ciências da Saúde em português e em inglês

Conforme Rajagoplan (2008), a LC trouxe uma contribuição significativa para os Estudos da Linguagem no Brasil. Fora de Letras/Linguística, pesquisas baseadas em LC já foram muito bem acolhidas e reconhecidas. Esse é o caso, para apenas um exemplo recente, do ótimo trabalho de Lima (LIMA, 2013). Sua pesquisa reuniu e estudou um extenso *corpus* de textos em português e em inglês que tratavam sobre o tema da *anemia falciforme*<sup>5</sup>, um importante tópico de Saúde Pública no Brasil e nos Estados Unidos. Com apoio estatístico, computacional e aporte de um enfoque linguístico-textual específico, foram examinados, extensivamente, artigos científicos, manuais e cartilhas e *folders* de divulgação para leigos dos dois países. Um dos pontos de partida de Lima foram constatações de maior dificuldade de acesso e de compreensão da informação textual por parte de brasileiros usuários desses materiais, enquanto a população norte-americana inglesa parecia estar melhor atendida em termos de acessibilidade de textos a ela oferecidos.

De modo muito simplificado<sup>6</sup>, esse trabalho de Lima (2013): **a)** atestou a relevância de se levar em conta o que apontam os estudos estatísticos de texto com base em *corpora* e discriminou diferenças importantes desses materiais nos dois países; **b)** integrou Medicina e Letras em torno do objeto texto e da comunicação em ciências, seja entre especialistas ou entre especialistas e leigos – salientando o tema da acessibilidade da informação; e, **c)** trouxe propostas de redação para alguns tipos de texto serem mais eficientes no Brasil, sem contar o contraste com o perfil de materiais semelhantes em língua inglesa – cuja funcionalidade e eficiência comunicativas confirmaram-se ser maiores.

---

<sup>5</sup> Doença hereditária caracterizada pela alteração dos glóbulos vermelhos do sangue, tornando-os parecidos com uma foice, daí o nome falciforme. Têm sua membrana alterada e rompem-se mais facilmente, causando anemia. Condição mais comum em indivíduos negros. No Brasil, acomete cerca de 8% da população negra, mas, devido à intensa miscigenação histórica, observa-se também em brancos ou pardos. Fonte: Ministério da Saúde do Brasil, <<http://bvsmis.saude.gov.br/bvs/dicas/127anemiafalci.html>>.

<sup>6</sup> Uma notícia da repercussão do trabalho de Lima (2013) confere-se em <http://www.geledes.org.br>.

## 2.2 Jornais populares: Brasil e Inglaterra

O jornal popular, no Brasil, é um gênero jornalístico novo, um tipo intermediário ou híbrido, colocando-se entre o jornal sensacionalista e o jornal tradicional (AMARAL, 2006) o qual tem atraído um público de menor poder aquisitivo e menor hábito de leitura. Esse público de menor hábito de leitura tende a não ser muito mobilizado pela leitura de textos de jornais tradicionais à venda, cujo preço, em geral, pode parecer elevado. Conforme Oliveira (2009, p. 8), com preços baixos, planejamento gráfico atraente, linguagem acessível e anúncios de produtos e serviços voltados ao público de renda menor, esse tipo de jornal conquistou novas audiências, que até então não tinham o acesso nem o hábito da leitura de publicações diárias tradicionais. Colocou-se, assim, no cenário da comunicação jornalística do Brasil, um tipo de publicação de forte apelo popular e com um texto bastante moldado por temáticas mais familiares às camadas economicamente menos privilegiadas da sociedade urbana e suburbana brasileira.

Em Porto Alegre-RS, no ano 2000, o jornal *Diário Gaúcho* (doravante **DG**) serve como um marco muito importante desse novo gênero jornalístico no nosso país, até então inédito, pois que não tratava de mera publicação sensacionalista e, sim, de algo popular. Já há 15 anos, segue buscando o diálogo com seu leitor, o cidadão e o consumidor das classes econômicas C, D e E. O **DG** circula sem oferecer assinatura; é vendido apenas em bancas e por jornaleiros ambulantes, com preço em torno de R\$ 1. Esse **jornal** também oferece brindes pela compra sucessiva de exemplares e incentiva ações comunitárias.

De 2000 até 2015, o **DG** mostra circulação muito expressiva, centrada na região metropolitana de Porto Alegre. Chega a vender quase 150 mil exemplares/dia, mesmo com a chegada dos jornais gratuitos distribuídos em estações de trem e metrô. Enquanto isso, um jornal tradicional da mesma região, tal como o ZH – editado na mesma cidade e pela mesma empresa do **DG**, voltado para um público bem mais escolarizado, fica em torno de 60 mil/dia, mas com venda em todo um Estado.

No cenário britânico, o jornal popular, do tipo tablóide, não é nada recente. O *The Sun* (doravante **TS**) existe desde a década de 60 e

é considerado um dos jornais mais lidos do mundo. Segundo o *National Readership Survey*, é um dos jornais da Inglaterra que tem mais leitores. Em agosto de 2013, segundo o *site The Guardian*, baseado em dados da ABC, o **TS** teve uma circulação de 2.258.259 jornais/ano. Considerando 365 dias/ano, tem-se uma média de circulação 6 mil jornais/dia, imaginando-se que circule todos os dias da semana. Esse número parece pouco diante dos 150 mil/dia do **DG**, que não circula aos domingos.

Com esses dados em mente e tendo-se em vista os diferentes perfis de leitores e de jornais, inferimos que a condição de “jornal popular” não deve ser a mesma nessas duas culturas. Justamente por isso, o **TS** e o **DG** tornam-se boa escolha para atividades didáticas com jornais brasileiros e ingleses. Afinal, já foram tratados tanto no âmbito dos estudos de Comunicação Social quanto em pesquisas linguísticas. Conhecer esses estudos pode ser bem útil para o professor interessado em explorar esse material com seus alunos.

Além do acesso livre desses jornais em suas versões *on-line* (ver o **TS** em <http://www.thesun.co.uk> e o **DG** em <http://diariogaucha.clicrbs.com.br>), torna-se valioso saber que o **TS** já foi pesquisado e comparado, justamente com o **DG** (BERTHIER, SILVA, 2012). Esse trabalho salientou possíveis correspondências temáticas e de estilo entre ambos, identificando temas, *layout*, conteúdos, formatos, abordagens, texto e linguagem de cada uma das publicações em uma amostra de edições do ano de 2011.

Nesse estudo de Jornalismo, conforme também se pensa em LC, acolheu-se o pensamento de que

é importante uma amostragem considerável de textos, capaz de ser representativa do funcionamento de um tipo de discurso em um período determinado, pois, a tendência hoje é compreender de forma mais exaustiva os mecanismos de construção do discurso jornalístico (BENETTI; LAGO, 2007, p. 121).

Segundo confirmam Berthier e Silva (2012), o **TS**, como o **DG**, seria um jornal “do povo”, barato, custando apenas 60p (o que corresponde a 60 centavos de *pound*). Além disso, conforme as autoras, o **TS** tende a aproximar-se do leitor, com narrativas construídas com dramaticidade, linguagem simples, popular e de fácil

compreensão. Forte em temas de entretenimento, o *TS* também expressa engajamento e revolta com os problemas políticos da atualidade e as desigualdades sociais. Seu diferencial, em relação aos jornais tradicionais ingleses, assim como vemos no *DG*, é o de ilustrar suas notícias com fotos bastante apelativas.

Na comparação, as autoras verificaram especificidades não apenas por serem fruto de culturas e de países diferentes, nem por manterem tipos distintos de público, mas *porque se expressam de um modo diferente* (grifo nosso) em relação aos jornais tradicionais. Percorrer e reconhecer os caminhos dessas diferenças entre jornais tradicionais e populares pode, justamente, ser uma boa ideia para aulas de língua e cultura em inglês e português num curso de Letras.

### **3 Frequências de verbos em jornais populares brasileiros e britânicos**

No nosso exercício, buscamos depreender, partindo do léxico, uma expressão diferenciada do *DG* e do *TS*. Quisemos, assim: **a)** verificar em que medida os textos do *DG* e do *TS* se diferenciariam ou se aproximariam no que se refere ao uso de variedades de verbos; **b)** identificar os verbos mais utilizados em cada jornal; **c)** observar padrões de emprego de verbos mais frequente desses jornais em relação ao uso em outros tipos de texto, com apoio de registros de frequência de grandes *corpora* do inglês e do português do Brasil.

Para nos subsidiar, recorreremos a trabalhos com dados sobre verbos em *corpora* e que tratassem de vocabulário. A seguir, destacamos alguns.

#### **3.1 Trabalhos de Biderman sobre verbos em corpora**

Em 1998, Biderman (BIDERMAN, 1998) já destacava os 20 verbos mais frequentes do português brasileiro, que foram pesquisados a partir de um *corpus* de 5 milhões de palavras compilado por ela. Naquele *corpus*, estavam presentes textos de literatura romanesca, literatura dramática, literatura técnico-científica, literatura jornalística, e literatura oratória (discursos e palestras transcritos). Todos esses textos foram publicados entre 1950 a 1990. Naquele *corpus*, havia uma

tendência de os textos, independentemente dos gêneros envolvidos, apresentarem um mesmo *ranking* de 20 verbos mais frequentes:

[...] Além disso, esses 20 verbos mais frequentes situam-se na escala decrescente de frequência em posições quase idênticas; isso confirma também que distribucionalmente eles operam de maneira muito similar na língua, não importando o tipo de variáveis linguísticas consideradas, a saber: língua falada ou escrita, linguagem literária, técnico-científica, jornalística etc. [...] (BIDERMAN, 1998, p. 171).

Outra importante conclusão da autora foi que “Por enorme que seja o léxico de uma língua, é reduzido o repertório desse acervo [de verbos] efetivamente utilizado pelos falantes do idioma”. (BIDERMAN, 1998, p. 178). Isso acontecia mesmo na língua escrita, na qual se poderia esperar um vocabulário mais diferenciado do que o da fala espontânea. Além disso, a autora também apontava (*op. cit.* p. 1) que, naquele seu grande *corpus*, o vocabulário jornalístico era o mais neutro e o menos temático, constituindo uma espécie de média entre os outros gêneros de linguagem.

### 3.2 Vocabulário em jornais populares e tradicionais brasileiros

Oliveira (2009) examinou, com apoio de recursos e de técnicas estatísticas da LC, o vocabulário dos jornais brasileiros *O Dia* (um jornal popular) e *O Globo* (jornal tradicional), ambos da cidade do Rio de Janeiro. Essa pesquisa visou a identificar, nos seus textos, de que maneira as diferenças de abordagem dos fatos e de conteúdo transpareceriam no seu léxico. Pretendeu encontrar, via léxico, elementos caracterizadores do jornal popular frente ao tradicional. Foram utilizadas apenas as edições *on-line* desses jornais.

Um tópico do estudo foi a observação da variedade do vocabulário ou riqueza lexical, uma medida em estatística lexical denominada *types/token ratio*. Essa medida envolve uma proporção entre o número total de palavras de um texto (*tokens*) e o número de palavras diferentes nele empregadas (*types*). Quanto maior a proporção (sendo o número de *types* dividido pelo de *tokens*), mais variado ou menos repetitivo é o vocabulário. Assim, por exemplo, um texto que tenha 200 palavras sendo essas 200 formadas pela repetição das mesmas 25, a proporção será de 12,5%. Conforme Oliveira, o jornal



popular apresentava uma porcentagem ligeiramente maior de variedade de vocabulário do que a do jornal tradicional. Esse traço, conforme concluiu, não seria um bom diferenciador do vocabulário de ambos. Por outro lado, a autora constatou que *O Dia* aparentava um léxico mais padronizado do que o de *O Globo*, o que verificou em termos do número de sequências de palavras recorrentemente associadas, formando feixes fixos de palavras em agrupamentos (*clusters*) que tendem a se repetir. Ainda assim (grifo nosso) **o grande diferencial do vocabulário entre ambos ficava por conta do número de verbos empregados**: no jornal popular, 29,6% das palavras eram verbos, enquanto que no jornal tradicional havia apenas 14,5% (OLIVEIRA, 2009, p. 16).

Considerando essas indicações, iniciamos nosso exercício exploratório. É que relatamos a seguir.

#### 4 Materiais e procedimentos do exercício

Como amostra do **DG**, recorreremos às edições completas desse jornal do mês de janeiro de 2008. São textos de temáticas variadas – Futebol, Notícias, Celebidades, Horóscopo etc., que perfizeram um conjunto de 150.977 palavras (*tokens*). Um extenso *corpus* do **DG** – com mais de 1 milhão de palavras – acessa-se, gratuitamente, no *site* do projeto PorPopular <<http://www.ufrgs.br/textecc/porlexbras/porpopular/>>. Nesse mesmo *site*, também é oferecido material do jornal popular baiano *Massa!*. Nosso *corpus* amostra do **TS**, por sua vez, é bastante pequeno, com apenas 105 textos, de edições de meses variados de 1999 a 2000. As temáticas envolvem *courts* (tribunais) e *showbiz* (celebidades). Nosso conjunto de textos ficou composto por 24.642 palavras (*tokens*), e integra o *THE METER Corpus*, disponível gratuitamente para *download* em <<http://nlp.shef.ac.uk/meter/>>.

De cada amostra de textos, **DG** e **TS**, foram produzidos, com apoio de ferramentas específicas e bem conhecidas em LC – tais como as ferramentas geradores de contextos ou *concordanciadores* e os geradores de listas de palavras ou *wordlists*, dados sobre as palavras mais frequentes em cada conjunto de textos. No *site* PorPopular, por exemplo, é possível gerar automaticamente essas listagens, sem necessidade de se baixar nada (veja o *site* antes citado). Para o *corpus*

**TS**, uma vez escolhidos e salvos os textos a examinar, pode-se também usar as mesmas ferramentas na opção *upload*.

Gerada a *wordlist* de cada *corpus* – o que não dispensou leitura de **muitos** contextos de ocorrência, chegamos a um levantamento preliminar dos 20 verbos mais frequentemente empregados em cada amostra de jornais e organizamos esses dados em listas. Em seguida, buscamos cotejar nossos dados com os de trabalhos ou levantamentos de dados pré-existentes – em português e em inglês, tentando obter informações sobre quais verbos seriam os mais utilizados em cada língua/jornal, independentemente dos tipos e temáticas de texto.

#### 4.1 Comparando listas de frequências de palavras mais usadas

Tendo explorado esses *corpora* e recorrido a algumas fontes de apoio que trataram do assunto, obtivemos quatro pacotes de dados diferentes: **a)** uma lista dos 20 verbos mais frequentes na amostra do **DG**. Nela, descontamos os verbos de ligação ou auxiliares; **b)** uma lista de 20 verbos mais frequentes no português do Brasil realizada por Biderman (1998) na construção do seu *Dicionário de Frequências* do léxico do português brasileiro contemporâneo; **c)** uma lista dos 20 verbos mais frequentes do *corpus* amostra **TS**; **d)** uma lista de palavras mais frequentes do inglês em geral – independentemente de tipos de textos – produzida pela editora Oxford na pesquisa *Facts about language* (<http://www.oxforddictionaries.com/words/the-oec-facts-about-the-language>).

A lista feita por Biderman (1998) foi baseada no seu *corpus* de 5 milhões de palavras antes citado, o que foi referência para o seu *Dicionário de Frequências do léxico do português brasileiro contemporâneo*. Não foram excluídos dela verbos auxiliares ou verbos de ligação, de modo que essa diferença teria que ser administrada no nosso exercício.

A lista de 20 verbos mais frequentes do inglês segundo o estudo da Editora Oxford está disponível *on-line* no *site*: <http://www.oxforddictionaries.com/words/the-oec-facts-about-the-language>. Para uma comparação inicial, utilizamos apenas os 20 primeiros verbos da lista. Essa lista Oxford visa subsidiar materiais didáticos para o ensino de inglês como língua estrangeira, especialmente dicionários para aprendizes. Não há, entretanto,

referências detalhadas sobre como o levantamento foi produzido. Portanto, cabe alguma cautela no seu uso.

## 4.2 Resultados iniciais, problemas e indícios

Apesar de o recorte temático do *corpus* amostra do *TS* ser específico para os assuntos *courts* e *showbiz*, seu vocabulário aproxima-se muito do inglês num âmbito geral, representado esse âmbito aqui apenas pela lista Oxford. Por outro lado, na amostra do *DG*, com textos de temas bem mais variados do que os do *TS*, há uma diferença grande dos verbos mais utilizados em relação ao levantamento de Biderman – sem deixarmos de ter em mente a diferença entre ter-se ou não verbos auxiliares nas contagens – visto que são muito frequentes no português. Esse conjunto de dados iniciais é que tentamos representar no Quadro 1 a seguir.

| Ranking | Amostra DG | Biderman  | Amostra TS | Lista Oxford |
|---------|------------|-----------|------------|--------------|
| 1       | ACOMPANHAR | SER       | BE (1)     | BE (1)       |
| 2       | FALAR (1)  | TER (5)   | HAVE (2)   | HAVE (2)     |
| 3       | ESCREVER   | IR (7)    | SAY (3)    | DO (4)       |
| 4       | GOSTAR     | ESTAR     | DO (4)     | SAY (3)      |
| 5       | CONHECER   | PODER     | TELL (5)   | GET (7)      |
| 6       | ACONTECER  | DIZER     | GO (6)     | MAKE (8)     |
| 7       | QUERER (2) | HAYER     | GET (7)    | GO (6)       |
| 8       | SABER (3)  | FAZER (6) | MAKE (8)   | KNOW         |
| 9       | VER (4)    | DAR       | CLAIM      | TAKE (10)    |
| 10      | OCORRER    | VER (4)   | FIND (9)   | SEE (15)     |

|    |                  |                   |                  |                  |
|----|------------------|-------------------|------------------|------------------|
| 11 | COBRAR           | <b>SABER (3)</b>  | <b>TAKE (10)</b> | <b>COME (11)</b> |
| 12 | <b>TER (5)</b>   | <b>QUERER (2)</b> | <b>COME (11)</b> | THINK            |
| 13 | <b>FAZER (6)</b> | FICAR             | <b>WANT (12)</b> | LOOK             |
| 14 | ASSISTIR         | ACHAR             | <b>USE (13)</b>  | <b>WANT (12)</b> |
| 15 | <b>IR (7)</b>    | DEVER             | <b>GIVE (14)</b> | <b>GIVE (14)</b> |
| 16 | ENVIAR           | <b>FALAR (1)</b>  | <b>SEE (15)</b>  | <b>USE (13)</b>  |
| 17 | ESTREAR          | CHEGAR            | ACCUSE           | <b>FIND (9)</b>  |
| 18 | AJUDAR           | PRECISAR          | DENY             | <b>TELL (5)</b>  |
| 19 | ESCOLHER         | COMEÇAR           | LEAVE            | <b>ASK (16)</b>  |
| 20 | ENTRAR           | OLHAR             | <b>ASK (16)</b>  | WORK             |

**Quadro 1-** Comparação de listas de 20 verbos mais frequentes em inglês e português

Nesse Quadro 1, estão os 20 verbos em quatro colunas. O verbo FALAR (1), por exemplo, está negrito indicando item comum entre o **DG** e o levantamento de Biderman. FALAR está em 2º lugar no *ranking* dos mais usados no *corpus* **DG** e traz um número (1) entre parênteses ao seu lado. Percorrendo a coluna ao lado, é possível encontrar FALAR na 16ª posição em Biderman. Vale o mesmo para as colunas do inglês, que têm mais itens negritos, que são também os verbos em comum entre o **TS** e a Lista Oxford.

Neste ponto, uma sugestão de atividade com nossos estudantes seria propor a eles um contraste entre padrões de FALAR e de TO SAY – sendo justamente muito importante observar também situações de uso com TO TELL, cuja frequência de usos parece ser bem diferente entre a *Lista Oxford* e o jornal. Nessa atividade, considerando-se a tarefa de examinar muitos contextos nesses *corpora*, teríamos, por exemplo, as seguintes situações de uso entre o **DG** e o **TS** apresentadas no *Quadro 2*.

|   |   |
|---|---|
| A sua meta é <b>falar</b> sobre a necessidade de promover a paz entre os povos. | Tagg <b>said</b> the man in front insulted and tried to provoke her by accusing her of being drunk. |
| Encerrado o show, ela foi embora sem <b>falar</b> com ninguém!                  | I have the right to <b>say</b> anything on the death of my son.                                     |

**Quadro 2-** Exemplos de FALAR/TO SAY nos *corpora*

Preliminarmente, apenas **sete** verbos se repetiram nas duas primeiras listas em português, enquanto, nas listas do inglês, **dezesesseis** verbos são comuns. Este indicativo, muito *grosso modo*, contrariaria Biderman (1998), pois os seus 20 verbos mais frequentes tenderiam a se confirmar como os mais empregados **em qualquer tipo de texto** (grifo nosso). Mas, como em qualquer viés quantitativo de descrição de dados, é preciso muita cautela e contextualização dos elementos ou tendências que se revelam, fugindo-se do perigo da generalização apressada. Aqui, muito possivelmente, é necessário ponderar sobre os verbos auxiliares e de ligação presentes na lista de Biderman, excluídos da nossa do **DG**. Sem a exclusão desses verbos da lista do **DG**, os resultados entre ambas seriam mais próximos, tal como se vê no inglês. Um dado interessante, no inglês, é que os verbos mais frequentes do **TS**, nos quais as seções abordam apenas os temas *courts* e *showbiz*, são igualmente tão utilizadas no vocabulário geral da língua, representado pela *Lista Oxford*.

Em síntese, o português mostraria mais diferenças entre verbos mais usados no jornal e num português, em tese, mais geral. Por outro lado, o material em inglês traz maiores coincidências entre o espectro de verbos do jornal e de um *corpus* geral, embora a temática seja mais restrita que a dos textos em português. Esses dados nos levam a uma nova pergunta: se fossem resgatados os verbos auxiliares e de ligação em português, haveria uma homogeneidade geral nas duas línguas? Aqui mais uma ideia para explorar em sala de aula.

Ainda que o nosso exercício seja metodologicamente irregular, pela exclusão *a priori* de um dado tipo de verbo, e também pela dimensão heterogênea dos corpora **TS** e **DG** (respectivamente, com cerca de 50 mil e 150 mil palavras), colocam-se, partindo-se do Quadro 1, interessantes temas de estudo contrastivo com verbos entre o português e o inglês. Nesse caso, justamente, poder-se-iam explorar

dados, com os estudantes de Letras Inglês-Português, sobre o papel e a distribuição frequencial de verbos auxiliares e de ligação em cada língua. Além de dados de *corpora*, naturalmente, é importante contar com o apoio dados de dicionários e de gramáticas das duas línguas. Valeria, especialmente, conhecerem-se os *sites* que nos oferecem os *corpora* de referência de cada língua, como o *BNC (British Nacional Corpus)* e o *Corpus Brasileiro*.

## 5 Perspectivas e considerações finais

Para uma investigação cuidadosa sobre prováveis diferenças ou semelhanças entre o repertório verbal de jornais populares brasileiros e britânicos, é preciso, frisamos, incorporar vários ajustes ao nosso método inicial. Entre eles, destacamos o que se poderia: **a)** buscar suporte metodológico em trabalhos de Linguística Contrastiva, que partem de uma validação de elementos comparáveis entre línguas tão diferentes como o inglês e o português; **b)** buscar mais suporte de investigações sobre características da linguagem nesse tipo de jornal nessas duas culturas; **c)** estabelecer um *corpus* comparável, com textos de temática semelhante entre o **DG** e **TS** – apenas com textos sobre esportes, por exemplo, visto que são bem abundantes na internet; **d)** realizar contagens de frequência com e sem categorização dos verbos por tipos nas duas línguas (auxiliares, verbos-suporte, verbos de ligação, verbos equivalentes, modais, etc.); **e)** centrar a observação em apenas um tipo de verbo, mais comum ou mais semelhantemente funcional aos dois idiomas ou mais significativo para o tipo de jornal em foco, e que tenha mostrado índices de frequências mais ou menos semelhantes; **f)** observar associações recorrentes com os verbos, aproveitando trabalhos pré-existentes sobre *subcategorization frames*, papéis semânticos, ou registros de dicionários específicos sobre usos de verbos e seus sujeitos e complementos mais usuais; **g)** escolher um par de verbos semântica e sintaticamente semelhantes entre as duas línguas, como, por exemplo, poderiam ser os verbos **TO KILL** e **MATAR**, e observar o seus comportamentos de uso e de distribuição nos dois jornais.

Como é fácil perceber, um simples exercício inicial de observação sobre o uso de verbos em textos de jornais demandaria toda uma série de procedimentos e de cuidados metodológicos. O rigor

metodológico é uma condição da pesquisa científica, mas poderia ser relativizado e até usado como tema de estudo em uma atividade didática pré-científica, que visasse justamente atrair o estudante para a pesquisa e reflexão sobre os funcionamentos das línguas. Não basta ter um *corpus* e apenas contar palavras sem ponderação. Até porque as palavras aparecem como seres quase vivos, inquietos, que sempre demandam que levemos em conta as suas especificidades, tal como vimos com a consideração ou não das categorizações ou tipologias de diferentes verbos que se queiram ou prefiram observar. Ainda assim, o nosso exercício de construção de um “método” inicial, o contato do estudante com *corpora*, a percepção de falhas e de dificuldades do quantitativo frente ao qualitativo da língua – ainda que sejam as duas faces de uma mesma moeda - têm um papel fundamental para a aprendizagem da pesquisa linguística em geral. Afinal, muito se pode fazer, em termos do estudo de e com textos, com o apoio de técnicas e/ou de princípios da LC. Os textos de Literatura de Língua Inglesa e de Língua Portuguesa mais populares entre os alunos de Letras, nesse sentido, por exemplo, seriam ótimos territórios para diferentes exercícios contrastivos.

### **Agradecimentos**

Ao programa PIBIC-CNPq da UFRGs e ao CNPq pelos constantes apoios à pesquisa linguística no âmbito universitário.

### **Referências**

AMARAL, Márcia Franz. *Jornalismo Popular*. São Paulo: Contexto, 2006.

BENETTI; Márcia. LAGO, Cláudia. *Metodologia de pesquisa em jornalismo*. Petrópolis: Vozes, 2007.

BERBER SARDINHA, Tony. *Linguística de Corpus*. Barueri, SP: Manole, 2004.

\_\_\_\_\_. Como usar a Linguística de Corpus no ensino de Língua Estrangeira. Por uma Linguística de Corpus educacional Brasileira. In: VIANA, V.; TAGNIN, E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 301-356.

BERTHIER, Camilla Afonso; SILVA, Paola. Jornalismo popular: não necessariamente sensacionalista. *Revista Científica do ITPAC*. Araguaína, n. 2, abr. 2012.

BIDERMAN, Maria Tereza Camargo. A face quantitativa da linguagem: um dicionário de frequências do português. *Alfa*, São Paulo, v. 42, p. 161-181, 1998.

LIMA, Kelen Cristina Sant'Anna de. *Caracterização de registros orientada para a produção textual no ambiente multilíngue: um estudo baseado em corpora comparáveis*. 2013. 251 f. Tese (Doutorado em Linguística Aplicada). Programa de Pós-graduação em Estudos Linguísticos da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

LEWIS, Michael. *Implementing the lexical approach*. Rove- England: Language Teaching Publications, 1997.

OLIVEIRA, Márcia Regina Alves Ribeiro. 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”. *Revista de estudos linguísticos Veredas*, Juiz de Fora, v.13, n.2, p. 07-19, 2009. Disponível em <<http://www.ufjf.br/revistaveredas/files/2009/11/ARTIGO-M%C3%A1rcia-Regina-Alves-Ribeiro-Oliveira1.pdf>>. Acesso em: 11 jun. 2014.

RAJAGOPALAN, Kanavalil. A Linguística de Corpus no tempo e no espaço: visão reflexiva. In: GERBER, Marcia. R.; VASILÉVSKI, Vera (Org). *Um percurso para pesquisas com base em corpus*. Florianópolis-SC: Editora da UFSC, Florianópolis, 2008. p.23-44.



VIANA, V.; TAGNIN, E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 301-356.

*Recebido em: 08/07/2014*

*Aceito em: 30/10/2014*