

A WEB COMO O BANCO DE DADOS MAIS RICO EM AMOSTRAS DA NORMA COLOQUIAL: AS EXPRESSÕES IDIOMÁTICAS

LA WEB COMO EL BANCO DE DATOS MÁS RICO EN MUESTRAS DE LA NORMA
COLOQUIAL: LAS EXPRESIONES IDIOMÁTICAS

THE WEB AS A RICHEST DATABASE IN SAMPLES OF THE COLLOQUIAL NORM: THE
IDIOMATIC EXPRESSIONS

Camila Maria Corrêa Rocha*

Instituto Federal Catarinense (IFC) | Campus Brusque

RESUMO: As expressões idiomáticas têm recebido a atenção dos linguistas, especialmente daqueles que têm o léxico como objeto de estudo (ORTIZ ÁLVAREZ; XATARA, 2001). Entretanto, apesar da sua recorrência na variante coloquial e do crescente interesse nelas por estudiosos, estudos anteriores mostraram que os dicionários gerais não as apresentam e, quando o fazem, a inclusão é assistemática, não condizente com suas especificidades léxico-semânticas. Nesta linha, tem-se como objetivo, neste estudo, verificar, no buscador *Google*, o número de ocorrências de um *corpus* de expressões idiomáticas, com vistas a comprovar a necessidade de que recebam um tratamento lexicográfico que condiga com suas especificidades. A utilização da *web* como instrumento de busca justifica-se posto que ela é, na atualidade, o banco de dados mais rico em amostras da norma coloquial – nível de linguagem em que se inserem as EIs.

PALAVRAS-CHAVE: Dicionário geral. Expressão idiomática. Web.

RESUMEN: Las expresiones idiomáticas están recibiendo la atención de los lingüistas, especialmente de aquellos que tienen el léxico por objeto de investigación (ORTIZ ÁLVAREZ; XATARA, 2001). Sin embargo, a pesar de su recurrencia en la variante coloquial y del creciente interés en ellas por estudiosos, investigaciones anteriores han mostrado que los diccionarios generales no las presentan y, si lo hacen, la inclusión se hace de manera asistemática, que no condice con sus especificidades léxico-semánticas. En este sentido, en esta investigación se tiene por objetivo, verificar, en el buscador *Google*, el número de ocurrencias de un *corpus* de expresiones idiomáticas, con miras a comprobar la necesidad de que reciban un tratamiento lexicográfico que condiga con sus especificidades.

* Professora do Instituto Federal Catarinense - campus Brusque. E-mail: camila.rocha@ifc.edu.br.

La utilización de la *web* como instrumento de búsqueda se justifica puesto que ella es, en la actualidad, el banco de datos más rico en muestras de la norma coloquial, nivel de lenguaje en el que se introducen las expresiones idiomáticas.

PALABRAS CLAVE: Diccionario general. Expresiones idiomáticas. Web.

ABSTRACT: Idiomatic expressions have received attention from linguists, especially those who have the lexicon as an object of study (ORTIZ ÁLVAREZ; XATARA, 2001). However, despite their recurrence in the colloquial variant of the language and the growing interest by scholars, previous studies have shown that general dictionaries usually do not present them and, when they do, inclusion is unsystematic and inconsistent with their lexical-semantic specificities. Therefore, the objective of this study is to verify the number of occurrences of a corpus of idiomatic expressions, to prove the need for receiving a lexicographic treatment that matches their specificities. The use of the web as a search tool is justified because it is, currently, the richest database in samples of the colloquial norm — language level in which the IEs are inserted.

KEYWORDS: General dictionary. Idiomatic expression. Web.

1 INTRODUÇÃO

As expressões idiomáticas (EIs) são unidades lexicais recorrentes na linguagem cotidiana, motivo pelo qual têm recebido cada vez mais a atenção dos linguistas, especialmente daqueles que têm o léxico como objeto de estudo (ORTIZ ÁLVAREZ; XATARA, 2001).

Nosso interesse por elas surgiu na graduação em Letras com habilitação em Espanhol, quando participamos de um projeto de iniciação científica que resultou em um glossário contrastivo de EIs do espanhol peninsular e do português. Posteriormente, na monografia elaborada ao final do curso de especialização em Ensino de Línguas Estrangeiras, fizemos a análise de alguns materiais didáticos de espanhol direcionados a aprendizes brasileiros, para observar como as EIs eram abordadas neles. A continuidade destas investigações ocorreu no mestrado; utilizamo-nos de parte das EIs pertencentes ao glossário referido para comprovar sua natureza metafórica. Na tese de doutoramento, elaboramos um repertório semibílingue de expressões idiomáticas do português com seus equivalentes no espanhol da variante argentina, público no qual focamos nossas atenções. Nosso corpus base foi aquele que desenvolvemos na graduação; porém, selecionamos dele somente as EIs pertencentes ao campo semântico dos corpos humano e animal.

Concebemos as EIs, neste estudo, como unidades lexicais de base figurativa, criadas e difundidas pelo imaginário popular. Contudo, embora sejam recorrentes na variedade coloquial do português, estudos anteriores mostraram que elas nem sempre são incluídas nos dicionários gerais de língua e, quando o são, a inclusão não é uniforme, ou seja, não condiz com suas especificidades léxico-semânticas, ou seja, não é levado em consideração o fato delas serem lexias complexas, pluriverbais, conotativas e indecomponíveis em seu sentido.

Assim, diante da problemática apresentada, tem-se como objetivo principal, neste estudo, verificar na *web*, no buscador *Google*, o número de ocorrências de um pequeno *corpus*¹ de EIs da língua portuguesa. Pretendemos comprovar sua recorrência neste espaço digital e destacar a necessidade de que elas sejam dicionarizadas e recebam um tratamento lexicográfico que condiga com suas especificidades léxico- semânticas.

Para tal, também pretendemos apresentar os resultados da observação do *corpus* em três dicionários monolíngues gerais e impressos do português, objetivando comprovar a assintematicidade com a qual elas figuram neles.

A utilização da *web* como instrumento de busca justifica-se, pois ela é, na atualidade, o banco de dados mais rico em amostras da norma coloquial, nível de linguagem em que se inserem as EIs.

¹ O *corpus* de que nos utilizaremos aqui constitui um pequeno recorte daquele que foi trabalhado em tese de doutoramento concluída em 2014, o qual era composto por aproximadamente 80 EIs pertencentes ao campo semântico dos corpos humano e animal.

Os pressupostos teóricos norteadores do presente estudo são: a) delimitação e definição de EI; b) considerações sobre sua inserção nos dicionários gerais; c) conceituação do dicionário nos estudos lexicográficos; d) contribuições da *web* aos estudos lexicográficos.

2 PRESSUPOSTOS TEÓRICOS

2.1. EXPRESSÕES IDIOMÁTICAS: DEFINIÇÕES

As expressões idiomáticas (EIs) são unidades lexicais recorrentes na linguagem coloquial e, portanto, reveladoras da cultura daqueles que as utilizam. Xatara (1998, p. 149) define expressão idiomática como uma “[...] lexia complexa indecomponível, conotativa e cristalizada em um idioma pela tradição cultural” e Noble (2002), nesta mesma linha, concebe-a como uma criação linguística de origem popular que se vulgariza e cristaliza em um idioma.

Roncolato (2004, p. 47) enfatiza, em sua definição, o aspecto estrutural da EI, ao concebê-la como uma “[...] construção pluriverbal, estável, fruto de um processo metafórico de formação, que pode funcionar como uma parte da oração ou como uma oração completa”, assim como Rivas González (2005), para quem ela é uma combinação de, no mínimo, dois elementos, cujo significado reside no bloco linguístico.

Xatara (1995) define-as, do ponto de vista pragmático, como criações dos falantes, quando eles não encontram no repertório lexical de que dispõem as unidades lexicais capazes de significar, com realce e originalidade, a mensagem que desejam transmitir. Porém, a autora destaca que, para que tais criações sejam consideradas EIs, elas devem cristalizar-se e expandir-se na linguagem cotidiana como um registro informal, oral ou escrito. Sob outra perspectiva, Lapa (1998) atribui sua criação à tendência que o homem tem de economizar esforços ao expressar-se, ou seja, é vantagem, a seu ver, que as lexias ocorram por grupos.

Ortiz Álvarez (2000) concebe-as sob as perspectivas estrutural e semântica. Estruturalmente, são unidades lexicais indecomponíveis, constituídas por mais de uma palavra, ao passo que, do ponto de vista semântico, consistem em sintagmas cujo sentido global não corresponde ao significado isolado dos elementos que as compõem.

Com base nas definições acima explicitadas, concebemos as EIs, neste estudo, como unidades lexicais: a) complexas, pluriverbais e indecomponíveis em seu significado; b) construídas com base figurativa e, portanto, conotativas; c) criadas pelo imaginário popular, o qual pode cristalizá-las ou excluí-las; d) construídas da necessidade que o falante tem de comunicar uma ideia com mais expressividade que a que a linguagem denotativa oferece.

2.2. AS EXPRESSÕES IDIOMÁTICAS EM DICIONÁRIOS GERAIS

Nos estudos linguísticos, as disciplinas da fala têm recebido atenção de maneira crescente. As áreas de estudo relacionadas ao léxico estão, cada vez mais, sendo foco de pesquisa. Essa mudança decorre do fato de que a oralidade, antes vista como inferior à escrita, hoje nivela-se a ela.

Entretanto, no que concerne ao conteúdo lexical expressões idiomáticas, quando pensadas no âmbito da teoria lexicográfica, observa-se que não lhes é dado um tratamento, nos dicionários gerais, que condiga com suas especificidades léxico-semânticas; vê-se que sua inclusão na microestrutura de dicionários gerais de língua é, em geral, feita de maneira aleatória e sem um embasamento teórico constituído. Como consequência dessa assistemática, os lexicógrafos evitam incluí-las em suas obras de referência e, quando o fazem, apresentam-nas a partir de um verbete base, eleito aleatoriamente.

Ortiz Álvarez (2001, p. 94) atribui esta dificuldade à complexidade que seu estudo impõe:

Primeiramente devemos ressaltar que tanto as expressões idiomáticas como outros tipos de fraseologismos, em geral, não se encontram repertoriados em obras de referência, particularmente nos dicionários de língua. Desse modo, a busca de seus equivalentes requer a pesquisa em várias fontes, sem que, muitas vezes, se obtenha resultados satisfatórios. Provavelmente, tal fato seja decorrente da complexidade do tratamento do tema, principalmente em relação ao seu reconhecimento. Assim, se é difícil identificá-las, não há como incluí-las em dicionários.

Assim, é anterior à dificuldade de inclusão nos dicionários gerais o próprio reconhecimento delas, em virtude da existência de uma gama de fraseologismos.

Também é problemática, na lematização das EIs, de acordo com Krieger (2006a), a inclusão de informações relativas à sua estrutura mórfica. Isso porque, ao contrário das lexias simples, para as quais há um modelo canônico a ser seguido, ainda não há uma estrutura mórfica canônica delimitada para as estruturas complexas. Isto quer dizer que, enquanto está estabelecido na teoria lexicográfica que os substantivos sejam lematizados no masculino e no singular e que os verbos apresentem-se no infinitivo, não há consenso sobre como lematizar as EIs.

A tradução delas também é motivo de debates entre os lexicógrafos. Noble (2002) constatou, em seus estudos, que ela é, em geral, deficiente, posto que as transferências semânticas que comumente são feitas não abarcam o significado conceitual da EI em sua totalidade. Para Xatara, Riva e Rios (2001), não basta que elas sejam definidas por uma paráfrase; é necessária a proposição de outra EI que lhes corresponda conotativamente, para que não se percam seus valores conotativos. Dessa forma, o tradutor deixa de simplesmente transferir significados e passa a construí-los, sob a influência do contexto cultural no qual está inserido.

De fato, advogamos que as EIs sejam definidas por uma paráfrase, que deve estar acompanhada da sua correspondente na outra língua; exemplificamos com a EI “nas costas dos outros”, a qual, além de ter seu significado explicitado na paráfrase “sem trabalhar, às custas de outra pessoa”, também deve ser definida por um sinônimo (“na jugular”), de forma que seja preservada a sua essência conotativa.

2.3. O DICIONÁRIO

Concebemos o dicionário, neste estudo, como a mais importante referência cultural da civilização moderna, o lugar onde devem estar registradas as normas sociais da época, assim como as marcas de uso dos falantes, dentre elas, as EIs, nosso objeto do estudo.

Rey-Debove (1984, p. 65) concebe-o como o objeto sociocultural que demarca o léxico, e Alcaraz (2006, p. 123) como “[...] a arma que deve estar sempre ao alcance de quem está permanentemente às voltas com as palavras”.

Biderman (1998) define o dicionário como o porta voz da sociedade, aquele que deve registrar o uso da língua consagrada pelos escritores, bem como os usos linguísticos correntes. Segundo a autora, ainda que o dicionário privilegie a língua escrita socialmente valorizada, “[...] ele deve descrever também os diferentes níveis de linguagem, os registros sociais e, assim, não só identificar o vocabulário e os usos marcados como típicos da linguagem coloquial, apontando os itens lexicais característicos de um uso popular, vulgar, chulo, as gírias e palavras e expressões obscenas.” (BIDERMAN, 1998, p. 166).

Lara (2004, p. 143) destaca algumas de suas funções, podendo servir “[...] como instrumento de tradução e entendimento entre os falantes de duas ou mais línguas, quando se trata dos dicionários bi- e multilíngues; como horizonte normativo dos falantes de uma língua e de seus diferentes dialetos; como discurso culto, referido ao estado em que se encontra uma comunidade linguística particular, e situado em seu caráter político e cultural”.

Para Krieger (2006), o dicionário é o único lugar onde o léxico de um idioma aparece delimitado e organizado. No entanto, a autora ressalta que este instrumento não abarca todo o vocabulário, que está em constante expansão, recebe criações neológicas e perde

lexias, que caem em desuso tornando-se arcaísmos. Assim, diante da impossibilidade de enquadrar todo o vocabulário de uma língua, Correia (2008) defende que os bons dicionários são aqueles que o delimitam, utilizando como critérios sua frequência de uso e seu público-alvo.

Sob outra perspectiva, Alcaraz (2006) chama a atenção para a necessidade de ele seja, para o falante, um instrumento norteador, e não um guia a ser cegamente seguido, dadas as limitações que lhe são inerentes. Sabe-se que, em alguns casos, uma consulta a ele não será suficiente para se resolver problemas relativos à significação de uma unidade lexical, principalmente quando o verbete está incompleto.

Na linguística moderna, o dicionário foi concebido sob perspectivas diversas. Inicialmente, como o foco era a oralidade das línguas e a naturalidade da fala, todos os instrumentos veiculadores da língua escrita foram negligenciados, o que levou à “[...] quase ausência de uma consideração/análise dos dicionários na semântica contemporânea, na pragmática, qualquer que seja sua definição, e na teoria da linguagem” (LARA, 2004, p. 134).

Foi somente na década de 1970 que o dicionário começou a ser objeto de estudo científico. Reconheceu-se, nessa década, segundo Azorín Fernandez (2003), sua importância como instrumento do bem falar e passaram a ser priorizados os procedimentos metodológicos utilizados em sua constituição. Nos anos 90, a importância dos estudos metalexigráficos foi incontestável entre os linguistas, os quais abordaram os dicionários em seus diversos aspectos: constituição histórica, contexto de produção e papel na didática das línguas, como explica Azorín Fernandez (2003).

Diante da importância que atribuímos a ele como instrumento portador dos usos correntes e consagrados da língua, pretendemos, neste estudo, averiguar na *web*, o banco de dados mais rico em amostras dos mais diversos registros da linguagem, o *status* de um pequeno *corpus* de EIs. Nosso intuito é verificar seu número de ocorrências no *Google*, o qual supomos ser alto, e, portanto, confirmar a necessidade de que as EIs sejam dicionarizadas, dada a sua representatividade no registro oral do português do Brasil e na memória coletiva dos falantes de português.

A utilização da *web* como ferramenta de busca para a detecção do número de ocorrências das EIs justifica-se pelo fato de que, na atualidade, a tecnologia tornou-se uma ferramenta importante para a Linguística, ao auxiliá-la em tratamentos empíricos da língua e por possibilitar a manipulação de um número maior de informações. Desse modo, propomos aqui um trabalho lexicográfico que combina o uso de *corpora* eletrônicos e de dicionários, uma vez que os primeiros viabilizam o trabalho com grandes quantidades de dados provindos de ambientes naturais, facilitando o trabalho do lexicógrafo; os segundos, por sua vez, garantem fiabilidade ao fornecerem dados que privilegiam a língua escrita.

Ainda que seja consenso, na lexicografia, que o trabalho lexicográfico deve ser baseado em um *corpus*, Riva (2008) explica que o reconhecimento da *web* como o maior e mais abrangente dos *corpora* implica em divergências entre os estudiosos. Alega-se, entre outras razões, que os textos disponíveis na internet não são previamente organizados, além de possuírem uma grande quantidade de erros. Para Xatara (2008, p. 771), ao contrário, “[...] a dimensão da *Web* corresponde mais adequadamente tanto para a demonstração do idiomatismo em um contexto real, quanto para acrescentar importantes informações concernentes à significação e uso da cada EI”. Riva (2008, p. 88), na mesma linha, ressalta que “A *web* é hoje o maior banco de dados disponível no mundo e é largamente difundida e utilizada pela facilidade de acesso e pela amplitude de campos do conhecimento que abrange. Trata-se, pois, de um retrato da língua em uso, com possibilidade de análises semântica e pragmática por conta da ampla tipologia textual que apresenta.”.

Acreditamos que recorrer à *web* no trabalho com nosso *corpus* é imprescindível, visto que ela aloja a linguagem coloquial – lugar em que se encontram as EIs. Isto nos possibilitará atestar a quantidade de ocorrências delas, bem como a necessidade de que sua inclusão nos dicionários seja feita mediante um tratamento lexicográfico adequado.

3 METODOLOGIA

Selecionamos, para este estudo, nove EIs do português, todas pertencentes ao campo semântico dos corpos humano e animal. Elas compõem o *corpus* trabalhado em tese de doutoramento, composto aproximadamente por oitenta EIs relativas campo semântico dos corpos humano e animal. Apresentamos, abaixo, cada uma delas, a partir da sua palavra-chave:

Palavra-chave da EI para a sua lematização	Expressão idiomática
Bico	No bico do urubu
Boca	Na boca do povo
Cabeça	Esquentar a cabeça
Cara	Encher a cara
Olho	De olhos bem abertos
Pé	Sem pé nem cabeça
Rabo	De cabo a rabo
Saco	Puxar o saco
Unha	Com unhas e dentes

Tabela 1: O *corpus* de Eis

Fonte: a autora

3.1. O *CORPUS* NOS DICIONÁRIOS

O primeiro passo metodológico foi averiguar como o *corpus* foi incluído em três dicionários monolíngues gerais e impressos do português: o Ferreira (2009), o Houaiss e Villar (2004) e o Michaelis (2002). Nosso objetivo foi comprovar a assiduidade com a qual eles, de maneira geral, trataram as EIs ao inseri-las. A escolha de tais dicionários justifica-se pelo fato de serem considerados os mais representativos da cultura lexicográfica brasileira.

A análise permitiu inferir se houve homogeneidade no tratamento lexicográfico dado às EIs pelos três, se os critérios que nortearam sua inserção foram coerentes e consensuais e se foram levadas em consideração suas particularidades léxico-semânticas. Analisamos três elementos da teoria lexicográfica, enumerados abaixo em 1, 2 e 3.

1. A localização da EI (na macroestrutura ou na microestrutura) e a palavra-chave que a introduziu (no caso de estar localizada na microestrutura);
2. A definição lexicográfica utilizada, com base na tipologia de definições proposta por Porto Dapena (2002). O autor classifica-a em: definição enciclopédica (DE), definição linguística conceitual perifrástica (DLCP), definição linguística conceitual sinonímica (DLCS) e definição explicativa (DLE);

3. A presença de abonações

A tabela abaixo reproduz a observação feita:

Expressão idiomática	Dicionário Aurélio (2009)	Dicionário Houaiss (2004)	Dicionário Michaelis (2002)
No bico do urubu	-	-	-
Na boca do povo	<p>1. Na microestrutura de <u>boca</u>.</p> <p>2. Tipo de definição: DLCP Cair na boca do povo: <i>Ser alvo de maledicência; tornar-se falado</i> (p. 307).</p> <p>3. Não há exemplo.</p>	-	-
Esquentar a cabeça	<p>1. Na microestrutura de <u>cabeça</u>.</p> <p>2. Tipo de definição: DLCP <i>Preocupar-se, inquietar-se, afligir-se [Tb. se diz apenas esquentar]</i> (p. 345).</p> <p>3. Não há exemplo.</p>	<p>1. Na microestrutura de <u>cabeça</u>.</p> <p>2. Tipo de definição: DLCP <i>Ficar preocupado; afligir-se</i> (p. 541).</p> <p>3. Não há exemplo.</p>	-
Encher a cara	<p>1. Na microestrutura de <u>cara</u>.</p> <p>2. Tipo de definição: DLCP <i>Ingerir em grande quantidade (bebida alcoólica): “o secreta enchera a cara de cachaça” (Jorge Amado, Dona Flor e seus dois maridos, p. 394)”</i> (p. 398).</p> <p>3. Exemplo retirado de fonte.</p>	<p>1. Na microestrutura de <u>cara</u>.</p> <p>2. Tipo de definição: DLCP <i>Beber muito, embriagar-se</i> (p. 615).</p> <p>3. Não há exemplo.</p>	<p>1. Na microestrutura de <u>cara</u>.</p> <p>2. Tipo de definição: DLCP <i>Abusar da ingestão de bebida alcoólica</i> (p. 426).</p> <p>3. Não há exemplo.</p>
De olhos bem abertos	-	-	-
Sem pé nem cabeça	-	-	<p>1. Na microestrutura de <u>pé</u>.</p> <p>2. Tipo de definição: DLCP Sem pés nem cabeça: <i>Despropositado, disparatado</i> (p. 1575).</p> <p>3. Não há exemplo.</p>
De cabo a rabo	<p>1. Na microestrutura de <u>cabo</u>.</p> <p>2. Tipo de definição: DLCP <i>De uma extremidade a outra</i> (p. 348).</p>	<p>1. Na microestrutura de <u>cabo</u>.</p> <p>2. Tipo de definição: DLCP <i>De um extremo ao outro, do início ao fim; em toda extensão ou duração;</i></p>	<p>1. Na microestrutura de <u>rabo</u>.</p> <p>2. Tipo de definição: DLCP <i>Do princípio ao fim</i> (p. 1760).</p>

	3. Não há exemplo.	<i>inteiramente, por completo</i> (p. 544).	3. Não há exemplo.
Puxar o saco	1. Na microestrutura de <u>saco</u> . 2. Tipo de definição: DLCP <i>Puxar o saco de: V. bajular</i> (p. 1787). 3. Não há exemplo.	1. Na microestrutura de <u>saco</u> . 2. Tipo de definição: DLCP Puxar o saco de: <i>Adular, bajular (alguém) excessivamente</i> (p. 2493). 3. Não há exemplo.	-
Com unhas e dentes	1. Na microestrutura de <u>unha</u> . 2. Tipo de definição: DLCP <i>De todas as maneiras possíveis; com todas as forças</i> (p. 2019). 3. Não há exemplo.	-	1. Na microestrutura de <u>unha</u> . 2. Tipo de definição: DLCP A unhas e dentes ou com unhas e dentes: <i>Ferozmente; de todas as formas possíveis, com todos os recursos, com todas as forças</i> (p. 2156). 3. Não há exemplo.

Tabela 2: O corpus nos dicionários

Fonte: a autora

O segundo passo metodológico, explicitado a seguir, consistiu na observação do *corpus* na *web*.

3.2. O CORPUS NA WEB

Verificamos, no buscador *Google* (do Brasil), o número de páginas nas quais cada EI apareceu.

Utilizamo-nos do limiar de frequência mínimo de 56 ocorrências estabelecido por Xatara (2008), a partir dos trabalhos de Colson (2003). De acordo com Colson (*apud* XATARA, 2008), para que uma EI seja considerada frequente, ela deve ocorrer uma vez a cada milhão de palavras (1 PMW). O PMW é uma unidade de medida utilizada para medir a frequência das unidades lexicais na *web*.

Assim, levando-se em consideração a estimativa de que haja, na *web*, cerca de 56 milhões de páginas do português do Brasil, ao aplicar o coeficiente de 1 PMW a este número de páginas, tem-se como limiar de frequência 56 ocorrências para o português. Assim, somente foram consideradas recorrentes as EIs cuja ocorrência no *Google* (nas páginas do Brasil) foi igual ou superior a 56.

A verificação do *corpus* na *web* ocorreu no dia 13 de dezembro de 2019. As EIs foram pesquisadas no buscador *Google* da seguinte forma: “EI” site: br (ex. “no bico do urubu” site: br). O número de ocorrências de cada uma delas foi o seguinte:

Expressão idiomática	Número de ocorrências na <i>web</i>
No bico do urubu	4.010
Na boca do povo	886.000
Esquentar a cabeça	102.000
Encher a cara	338.000
De olhos bem abertos	152.000
Sem pé nem cabeça	191.000
De cabo a rabo	873.000
Puxar o saco	134.000
Com unhas e dentes	426.000

Tabela 3: As EIs no *Google*

Fonte: a autora

4 RESULTADOS E DISCUSSÕES

4.1. O CORPUS NOS DICIONÁRIOS

Quanto à análise dos três dicionários, o quadro da subseção 3.1 mostrou que, em geral, houve uma coincidência entre o Ferreira (2009) e o Houaiss e Villar (2004) na proposição das EIs e no modo de fazê-lo. O Michaelis (2002), por sua vez, divergiu deles na quantidade apresentada e na maneira de dispô-las. Duas EIs não figuraram em nenhum dicionário: “no bico do urubu” e “de olhos bem abertos”.

Quanto à sua localização (na macroestrutura ou na microestrutura) e à palavra-chave que as introduziu, parece ter havido consenso entre os autores dos dicionários, visto que elas foram incluídas como subentradas de um de seus itens lexicais. Neste caso, Castillo Carballo (2003, p. 94) sugere sua inserção no verbete a partir de uma das lexias que as formam; assim, as que contêm, em sua estrutura, um substantivo, devem elegê-lo como palavra-chave e, portanto, como entrada. Na ausência dele, devem ser usados os verbos (quando forem principais, não auxiliares), os adjetivos, os pronomes e, finalmente, os advérbios.

Observamos, na análise, que o critério para a eleição da palavra-chave parece ter sido a escolha do primeiro substantivo presente em sua estrutura. Por este motivo, nem todas elas foram incluídas pela lexia que julgamos ser a sua palavra-chave (os substantivos pertencentes ao campo semântico dos corpos humano e animal). Esse foi o caso da EI “de cabo a rabo”, incluída, em Ferreira (2009) e em Houaiss e Villar (2004), na microestrutura do substantivo “cabo”.

No tocante à definição lexicográfica, todas elas foram conceituadas a partir da definição linguística conceitual perifrástica (DLCP), que é a priorizada na tradição lexicográfica monolíngue. Seu uso justifica-se porque ela é analítica, define a unidade lexical minuciosamente. Como exemplo, citamos a DLCP empregada na EI “encher a cara”: beber muito, embriagar-se. Reconhecemos sua relevância na elucidação do sentido da EI, porém, advogamos pela combinação dela (DLCP) com a sinonímica (DLCS). Isso porque seu sentido não pode ser apreendido mediante a observação do significado isolado dos vocábulos que as constituem; disso resulta uma significação complexa, cuja dificuldade na compreensão poder ser minimizada pela proposição de uma definição conceitual, que seja complementada por uma ou mais equivalências semânticas, no caso de havê-las. Das EIs em análise, receberam as duas definições: “na boca do povo” e “esquentar a cabeça”, ambas em Ferreira (2009).

Por fim, quanto à proposição de abonações, detectamos a ausência de preocupação dos dicionaristas na elucidação das EIs com exemplos, posto que somente uma, das nove analisadas, foi exemplificada: “encher a cara”, pelo Aurélio (2009).

4.2. O CORPUS NA WEB

Quanto à busca do número de ocorrências do *corpus* na *web* (subseção 3.2), observamos que todas ocorreram mais de 56 vezes – limiar de frequência estabelecido por Xatara (2008), a partir dos estudos de Colson (2003), e adotado por nós. Ressaltamos que, ainda que algumas possam apresentar-se, também, em sentido literal e, portanto, figurarem nas páginas do *Google* e nos resultados, a maior parte das ocorrências foi em seu formato idiomático, conforme averiguamos.

5 CONSIDERAÇÕES FINAIS

Procuramos mostrar, com o presente estudo, que as expressões idiomáticas devem constar nos dicionários gerais monolíngues e receber um tratamento lexicográfico adequado às suas particularidades estruturais e semânticas, uma vez que, conforme defende Biderman (1998), o dicionário deve ser o porta voz da sociedade, registrar o uso da língua consagrada pelos escritores, bem como os usos linguísticos correntes. Segundo a autora, ainda que privilegie a língua escrita socialmente valorizada, “[...] ele deve descrever também os diferentes níveis de linguagem, os registros sociais e, assim, não só identificar o vocabulário e os usos marcados como típicos da linguagem coloquial, apontando os itens lexicais característicos de um uso popular, vulgar, chulo, as gírias e palavras e expressões obscenas.” (BIDERMAN, 1998, p. 166).

A análise dos dicionários demonstrou que houve uma coincidência entre os dicionários Ferreira (2009) e Houaiss e Villar (2004) na proposição das EIs e no modo de fazê-lo. O Michaelis (2002), por sua vez, divergiu deles na quantidade apresentada e na maneira de dispô-las. Porém, ao observarmos a definição lexicográfica que lhes foi atribuída, todas elas foram definidas da maneira que é priorizada na tradição lexicográfica monolíngue, pela perífrase. Acreditamos que, para as unidades lexicais simples, a definição perifrástica seja suficiente e a proposição de um sinônimo seja um complemento dela, o que não acontece no caso das EIs, as quais exigem, a nosso ver, ambas as definições.

Quanto à pesquisa na *web*, o principal objetivo deste estudo, comprovamos sua recorrência nas linguagens popular e coloquial, o que mostra que estão cristalizadas na memória coletiva dos falantes de português e são consagradas pelo uso no português do Brasil. Cabe à Lexicografia, portanto, aprofundá-las e torná-las um tópico dos estudos lexicográficos e metalexográficos.

REFERÊNCIAS

- ALCARAZ, R. C. Dicionário: alcance e limites. *Fragmentos*, Florianópolis, n. 30, p. 121-127, 2006.
- AZORÍN FERNANDEZ, D. La lexicografía como disciplina lingüística. In: MEDINA GUERRA, A. M. (coord.). *Lexicografía española*. Espanha: Ariel Lingüística, 2003. p. 33-52.
- BIDERMAN, M. T. C.. O dicionário como norma na sociedade. In: CARVALHO, N. M. de; SILVA, M. E. B.. *Lexicologia, lexicografía e terminología: questões gerais do GT de lexicologia, lexicografía e terminología da ANPOLL*. Rio de Janeiro: UFRJ, 1998. p. 161-180.
- CASTILLO CARBALLO, M. A. La macroestructura del diccionario. In: MEDINA GUERRA, A. M. (coord.). *Lexicografía española*. Espanha: Ariel, 2003. p. 79-100.
- CORREIA, M. Lexicografía no início do século XXI: novas perspectivas, novos recursos e suas consequências. In: ALEXANDRE

- JÚNIOR, M. (coord.). *Actas do colóquio*. Lisboa: Centro de Estudos Clássicos, 2008. p. 1-13.
- FERREIRA, A. B. de H. *Novo dicionário Aurélio da Língua Portuguesa*. 4. ed. Curitiba: Positivo, 2009.
- HOUAISS, A.; VILLAR, M. de S. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2004.
- KRIEGER, M. da G. *et al.* O século XX, cenário dos dicionários fundadores da lexicografia brasileira: relações com a identidade do português do Brasil. *Alfa*, São José do Rio Preto, v. 50, n. 2, p. 173-187, 2006.
- KRIEGER, M. da G. Lexicografia: o léxico no dicionário. In: SEABRA, M. C. T. C. (org.). *O léxico em estudo*. Belo Horizonte: Faculdade de Letras da UFMG, 2006a. p. 157-172.
- LAPA, M. R. *Estilística da língua portuguesa*. 4. ed. São Paulo: Martins Fontes, 1998.
- LARA, L. F. O dicionário e suas disciplinas. In: ISQUIERDO, A. N.; KRIEGER, M. da G. (org.). *As ciências do léxico: lexicologia, lexicografia, terminologia*. Campo Grande: UFMS, 2004. p. 133-152.
- MICHAELIS. *Moderno dicionário da língua portuguesa*. São Paulo: Companhia Melhoramentos, 2002.
- NOBLE, G. D. Comparación de las expresiones idiomáticas del portugués al español. *Cadernos do Instituto de Letras*, Porto Alegre, n. 21/22, 2002.
- ORTÍZ ALVAREZ, M. L. *Expressões idiomáticas do português do Brasil e do espanhol de Cuba: estudo contrastivo e implicações para o ensino de português como língua estrangeira*. 2000. Tese (Doutorado em Linguística Aplicada) - Instituto de Estudos da Linguagem, Campinas, 2000.
- ORTÍZ ALVAREZ, M. L. Dicionário de expressões idiomáticas ou dicionário fraseológico? *Revista Línguas & Letras*, Cascavel, v. 2, n. 2, p. 83-96, 2001.
- PORTO DAPENA, J. A. *Manual de técnica lexicográfica*. Madrid: Arco Libros, 2002.
- REY-DEBOVE, J. Léxico e dicionário. Tradução de Clóvis Barleta de Moraes. *Alfa*, São Paulo, v. 28, p. 45-69, 1984.
- RIVA, H. C. *Dicionário onomasiológico de expressões idiomáticas usuais na língua portuguesa do Brasil*. 2008. Tese (Doutorado em Estudos Linguísticos) - Universidade Estadual Paulista, São José do Rio Preto, 2008.
- RIVAS GONZÁLEZ, M. Tratamiento de las expresiones fraseológicas en los principales diccionarios de español: propuesta para el aprendizaje de español como lengua extranjera. In: CONGRESO INTERNACIONAL DE ASELE, 16., 2005, Sevilla. *Anales [...]*. Sevilla: Secretariado de Publicaciones de la Universidad de Sevilla, 2005.
- RONCOLATTO, E. Critérios para a organização de dicionários fraseológicos. *Cadernos de Estudos Linguísticos*, Campinas, v. 46, n. 1, p. 43-52, 2004.

XATARA, C. M. O resgate das expressões idiomáticas. *Alfa*, São Paulo, v. 39, p. 195-210, 1995.

XATARA, C. M. O campo minado das expressões idiomáticas. *Alfa*, São Paulo, v. 42, p. 147-159, 1998.

XATARA, C. M. A web para um levantamento de frequência. In: MAGALHÃES, J. S.; TRAVAGLIA, L. C. (org.). *Múltiplas perspectivas em linguística*. Uberlândia: EDUFU, 2008, p. 770-777.

XATARA, C. M.; RIVA, H. C.; RIOS, T. H. C. As dificuldades na tradução de idiomatismos. *Cadernos de Tradução, Florianópolis*, v. 2, n. 8, p. 183-194, 2001. Disponível em: <https://periodicos.ufsc.br/index.php/traducao/article/view/5892/5572>. Acesso em: 25 ago. 2013.



Recebido em 13/12/2019. Aceito em 05/04/2020.