

# **Léxico em foco: dicionários com que sonhamos**

**Cristina Martins Fargetti  
Clotilde de Almeida Azevedo Murakawa  
Odair Luiz Nadin  
(Org.)**

**CULTURA  
ACADÊMICA**  
*Editora*

**LÉXICO EM FOCO:  
DICIONÁRIOS COM QUE  
SONHAMOS**

*SÉRIE*  
**TRILHAS LINGUÍSTICAS**  
n° 32 – 2019

**Faculdade de Ciências e Letras, UNESP – Univ Estadual Paulista,  
Campus Araraquara**

Reitor: Sandro Roberto Valentini

Vice-Reitor: Sergio Roberto Nobre

Diretor: Cláudio Cesar de Paiva

Vice-Diretora: Rosa Fátima de Souza Chaloba

**Programa de Pós-Graduação em Linguística e Língua Portuguesa**

Coordenador: Prof. Dr. Jean Cristtus Portela

Vice-Coordenadora: Profa. Dra. Alessandra Del Ré

***SÉRIE TRILHAS LINGUÍSTICAS Nº 32***

**Comissão Editorial da Pós-graduação em Linguística e Língua Portuguesa**

Alessandra Del Ré

Angélica Terezinha Carmo Rodrigues

Cristina Martins Fargetti

Jean Cristtus Portela

Marina Célia Mendonça

Nildicéia Aparecida Rocha

Odair Luiz Nadin da Silva

Rosane de Andrade Berlinck

**Diagramação:** Eron Pedroso Januskevictz

**Normalização:** Biblioteca da Faculdade de Ciências e Letras

**LÉXICO EM FOCO:  
DICIONÁRIOS COM QUE  
SONHAMOS**

Cristina Martins Fargetti  
Clotilde de Almeida Azevedo Murakawa  
Odair Luiz Nadin  
(Org.)

**CULTURA**  
**ACADÊMICA**   

---

*Editora*

Copyright © 2019 by FCL-UNESP Laboratório Editorial  
Direitos de publicação reservados a:  
Laboratório Editorial da FCL

Rod. Araraquara-Jaú, km 1  
14800-901 – Araraquara – SP  
Tel.: (16) 3334-6275

E-mail: [laboratorioeditorial.fclar@unesp.br](mailto:laboratorioeditorial.fclar@unesp.br)  
Site: <http://www.fclar.unesp.br/laboratorioeditorial>

Obra disponível em formato impresso e eletrônico  
(consultar endereço acima).

L591 Léxico em foco: dicionários com que sonhamos / Organizado por:  
Cristina Martins Fargetti, Clotilde de Almeida Azevedo  
Murakawa e Odair Luiz Nadin. –  
São Paulo, SP : Cultura Acadêmica, 2019.  
272 p. ; 14x21 cm. – (Série Trilhas Linguísticas; 32)

ISBN 978-85-7249-018-4

1. Lexicografia. 2. Enciclopédias e Dicionários. 3. Lexicologia.  
I. Fargetti, Cristina Martins. II. Murakawa, Clotilde de Almeida Azevedo.  
III. Nadin, Odair Luiz. VI. Série.

CDD 410

# SUMÁRIO

|   |     |
|---|-----|
| Apresentação  |     |
| <i>Cristina Martins Fargetti</i> .....  | 7   |
| A pesquisa reversa: dicionários codificadores e dicionários onomasiológicos   |     |
| <i>Álvaro Iriarte Sanromán</i> .....  | 15  |
| Verbetes que sonhamos... Verbetes que fazemos: a equivalência em recentes trabalhos lexicográficos                                |     |
| <i>Claudia Zavaglia</i> .....   | 31  |
| Dicionários que fizemos: algumas reflexões sobre o Dicionário Histórico do Português do Brasil – séculos XVI, XVII e XVIII (DHPB) |     |
| <i>Clotilde de Almeida Azevedo Murakawa</i> .....   | 49  |
| Léxico histórico do português brasileiro: caminhos percorridos, metas a atingir   |     |
| <i>Vanderci de Andrade Aguilera e Fabiane Cristina Altino</i> .....   | 67  |
| O desafio da retrodatação para os estudos etimológicos de língua portuguesa   |     |
| <i>Mário Eduardo Viaro, Aldo Luiz Bizzocchi, Mariana Giacomini Botta, Marco Dimas Gubitoso e Gustavo Luiz Vieira</i> .....        | 83  |
| Dicionários para línguas indígenas do Brasil  |     |
| <i>Jorge Domingues Lopes e Ana Suelly Arruda Câmara Cabral</i> .....  | 101 |

|   |     |
|---|-----|
| Dicionário do Nordeste<br><i>Nelly Carvalho</i> .....   | 115 |
| Quando a elaboração de dicionários pode auxiliar na reafirmação de uma identidade étnica: o caso dos professores Mura de Autazes<br><i>Cristina Borella e Eneida Alice Gonzaga dos Santos</i> .....       | 127 |
| A elaboração do dicionário crítico da obra de Domingos Vieira Filho<br><i>Conceição de Maria de Araujo Ramos, José de Ribamar Mendes Bezerra, Luís Henrique Serra e Maria de Fátima Sopas Rocha</i> ..... | 143 |
| Brincadeiras infantis e o léxico regional: um estudo de dados geolinguísticos<br><i>Láisa Bauermeister Stelo e Aparecida Negri Isquerdo</i> .....   | 155 |
| O vocabulário rural da Serra da Canastra/MG: um estudo linguístico na nascente do rio São Francisco<br><i>Gisele Aparecida Ribeiro</i> .....  | 167 |
| O regional e o global nos nomes próprios de pessoas no Brasil<br><i>Jéssica Paula Vescovi e Márcia Sipavicius Seide</i> .....   | 183 |
| A regionalidade na seleção lexical de informantes do Atlas Linguístico do Estado de Alagoas (ALEAL)<br><i>Maranúbia Pereira Barbosa Doiron</i> .....  | 199 |
| Criações lexicais literárias: a irreverência expressa pelas criações lexicais na poética de João Cabral de Melo Neto<br><i>Rosana Maria Sant'Ana Cotrim</i> .....   | 215 |
| Informatização e trabalho humano em Terminologia: extrair terminologias em <i>corpus</i> é o mesmo que identificá-las?<br><i>Maria José Bocorny Finatto</i> .....   | 231 |
| Anotação de <i>corpus</i> do grego antigo: projeto de tradução alinhada, <i>treebank</i> e léxico grego-português em serviços <i>web</i><br><i>Anise D'Orange Ferreira</i> .....                          | 247 |

# APRESENTAÇÃO

**Cristina Martins Fargetti**

Com grande satisfação, apresentamos este volume da Série Trilhas, com textos que resultam de apresentações durante o I CINELI – Congresso Internacional Estudos do Léxico e suas Interfaces - realizado na UNESP, Faculdade de Ciências e Letras, Araraquara, de 7 a 9 de maio de 2014. Tal congresso internacional teve expressiva participação de pesquisadores brasileiros e estrangeiros, tendo tido a estimativa de aproximadamente 400 participantes. Contou com apoios financeiros da CAPES e FAPESP, e com apoios dos Grupos LINBRA e GPEL, bem como do Programa de Pós-Graduação em Linguística e Língua Portuguesa e do Departamento de Linguística.

Os trabalhos aqui reunidos abrangem as áreas de Lexicografia, Regionalismos e Linguística de Corpus. Com autores nacionais e estrangeiros de reconhecida produção na área, estes trabalhos propiciam bons debates, com possibilidades de revisões bibliográficas e novos delineamentos teórico-metodológicos. É para nós, portanto, motivo de grande alegria esta publicação, muito importante para os Estudos do Léxico, área com uma história consolidada na FCL-Ar, por seus estudos como teses, dissertações e monografias defendidas, bem como por suas publicações de dicionários reconhecidos amplamente, como o DUP, o dicionário de grego, o DHPB, entre outros.

## **Lexicografia**

Como primeira obra deste volume, temos o texto de Álvaro Iriarte Sanromán, que se constitui de sua conferência de abertura

do referido evento. Professor da UMINHO, em Braga, Portugal, responsável pela edição recente do Dicionário Porto de Português e Espanhol, delinea possibilidades de aplicação de ferramentas de pesquisa reversa em dicionários eletrônicos, para poder transformá-los em dicionários codificadores (que permitem observar as combinações entre palavras para se expressar ideias), além de onomasiológicos (que permitem apenas saber a que palavra corresponde uma ideia). Exemplificando as dificuldades com conteúdo semântico para um consulente, o autor nos mostra que os dicionários que temos raramente nos permitem obter informações, por exemplo, sobre fraseologias como “tirar foto” – a associação entre esse verbo, e não outro, não fica explícita no verbete do verbo “tirar”, nem no verbete do nome “foto”. Nos mostra então considerações sobre o Dicionário Aberto, disponível on-line, com possibilidades promissoras para o reconhecimento e extração automáticos de relações léxico-conceituais, e com funções muito produtivas para variados tipos de estudos linguísticos.

Claudia Zavaglia, docente da Unesp, em São José do Rio Preto, autora de obras lexicográficas e importantes trabalhos na área, em texto de sua participação em mesa redonda do evento, reflete sobre a equivalência em obras lexicográficas. Aborda a princípio a dificuldade que pode sentir um consulente de dicionário monolíngue, ao se deparar com remissivas que, longe de esclarecer o significado de um item lexical, em português, lançam mais dúvidas por sua falta de detalhes. Questiona-se então sobre as dificuldades com equivalentes em dicionários bilíngues e multilíngues. Passa a descrever um dicionário multilíngue de regência verbal, de que é co-autora, em publicação recente, com tratamento de seis línguas diferentes. A dificuldade com as equivalências “exatas” e mesmo a sua inexistência é mencionada e é feita uma crítica a remissivas excessivas, em detrimento de melhores verbetes, mais completos. Muito trabalho a ser feito, com certeza.

Nossa pesquisadora homenageada, Clotilde de Almeida Azevedo Murakawa, docente do programa de pós graduação da Unesp de Araraquara, coordenadora do DHPB, discute a elaboração desta obra de fôlego, importante para estudos histórico comparativos,

entre outros. Com um banco de dados de sete milhões e meio de ocorrências, reunindo documentos/textos dos séculos XVI, XVII, XVIII e começo do XIX até o ano de 1808, este dicionário histórico do português do Brasil trouxe questionamentos para sua elaboração, requerendo discussões teórico-metodológicas sobre como estruturar seus verbetes; o que a autora passa a relatar. Apresenta verbetes e as decisões para sua construção, referentes a variação de grafia, homonímia, etimologia, datação, entre outros, levando em conta a proposta da obra, o público a que se destina, e a abrangência do período delimitado. Suas considerações nos motivam a uma consulta a esta obra, construída com a participação e inúmeros lexicógrafos, inclusive estrangeiros, constituindo-se uma referência.

Vanderci de Andrade Aguilera e Fabiane Cristina Altino, professoras da UEL, discutem o andamento de projeto também relacionado a documentos/textos do português do Brasil, procurando levantar o léxico dos séculos XVIII e XIX. Projeto sediado na UEL, vincula-se ao PHPB. Assim, o *Léxico Histórico do Português Brasileiro* pode contribuir para uma compreensão melhor do léxico do Brasil Colonial e do Brasil Império, e pode trazer um importante banco de dados para estudos sincrônicos e histórico-comparativos. As autoras apontam as fontes do banco de dados, constituídas de documentos oficiais colhidos em vários estados, podendo inclusive contar com bancos de dados já construídos previamente. As dificuldades de leitura e compreensão são apontadas e observa-se o trabalho de arqueologia das palavras, também enfrentado pelo projeto do DHPB e, certamente, enfrentado por todo o que se dedica ao estudo de manuscritos. Projeto concluído, com a participação de muitos pesquisadores, são apresentadas suas etapas de trabalho, exemplos de verbetes construídos e as próximas etapas de trabalho, com seus desdobramentos em pesquisas futuras. Seguramente, uma obra de fôlego e de inegável importância.

Na sequência, Mário Eduardo Viaro, Aldo Luiz Bizzocchi, Mariana Giacomini Botta, Marco Dimas Gubitoso e Gustavo Luiz Vieira refletem sobre a Etimologia, suas bases históricas e seus estudos atuais, com menção ao Núcleo de Apoio à Pesquisa em Etimologia e História da Língua Portuguesa (NEHiLP), sediado na

USP, que pretende constituir o *Dicionário Etimológico da Língua Portuguesa* (DELPo). Pretende-se discutir possibilidades de sincronia e diacronia do léxico português, tematizando substratos germânico, árabe, tupi, quimbundo e outros. Fixou-se para início a primeira edição do *Dictionarium ex Lusitanico in latinum sermonem*, de Jerônimo Cardoso de Lamego (1508-1569), com o ano 1562, portanto, como *terminus a quo* das primeiras palavras do banco de dados do NEHiLP, que constituem base de comparação de todas as obras submetidas pelo programa *Moedor*, denominada TAQ-zero. Este programa, de acesso livre, é discutido, com suas etapas de aplicação, e é descrito o banco de dados já constituído. Com a possibilidade de datação e retrodatação do maior número possível das primeiras ocorrências de palavras do português, busca-se uma obra lexicográfica de Etimologia, baseada em fundamentação científica, o que não se encontra em obras anteriores.

Jorge Domingues Lopes e Ana Suelly Arruda Câmara Cabral, pesquisadores do LALLI – Laboratório de Línguas e Literaturas Indígenas – da UNB, refletem sobre a necessidade de mais dicionários de línguas indígenas de nosso país, tendo em vista o processo de enfraquecimento do uso de tais línguas e mesmo observando que muitas já possuem escrita consolidada. Uma breve análise de dicionários já existentes é feita, focalizando desde dicionários do século XVII a dicionários mais recentes. Notam-se diferenças grandes em sua constituição, o que deve ocorrer em boa parte da quantidade de obras mencionada pelos autores, que se questionam para quem e por quem são feitas estas obras. Buscando contribuir com a área, Jorge Lopes construiu o Programa Línguas, o qual se constitui uma ferramenta muito bem-vinda para o estudo do léxico de línguas indígenas, por sua versatilidade e inclusive pela sua origem, totalmente brasileira.

## **Regionalismos**

Nelly Carvalho, professora emérita da UFPE, realiza uma análise metalexográfica do *Dicionário do Nordeste*, de Fred Navarro, na versão de 2013, com 10500 verbetes, com 55 neologismos con-

tabilizados. A autora reproduz diversos verbetes da obra, de lexe-  
mas definidos para vários estados nordestinos, e afirma sua sin-  
gularidade e mesmo seu arcaísmo, em relação a outras variedades  
brasileiras.

Cristina Borella e Eneida Alice dos Santos, professoras da  
UFAM, apresentam uma experiência de elaboração de dicionários  
temáticos por professores mura, povo amazônico. Apesar de não  
mais falantes de sua língua, e também do nheengatu, língua geral  
amazônica, solicitaram, em curso de formação de professores, o  
trabalho com o nheengatu, para seu povo poder falar novamente  
uma língua indígena. Esta solicitação acabou resultando no traba-  
lho lexicográfico, em português local, em que TCCs trataram de  
campos semânticos como vegetação, remédios e peixes. São então  
apresentados e analisados por elas alguns verbetes construídos pelos  
professores indígenas, explicando suas escolhas e decisões. Os nove  
dicionários mencionados constituem importante material didático  
para as comunidades mura e se apresentam como uma possibilida-  
de de obras lexicográficas, realizadas pelos próprios indígenas.

Conceição de Maria de Araujo Ramos, José de Ribamar  
Mendes Bezerra, Luís Henrique Serra e Maria de Fátima Sopas  
Rocha, ligados à UFMA e à USP, discutem o *Dicionário Crítico  
da Obra de Domingos Vieira Filho*, que surgiu da necessidade de  
resgatar parte importante da produção científica sobre a linguagem  
e a cultura popular do Maranhão. Este autor, portanto, constitui  
referência nos estudos do patrimônio cultural do estado, realizados  
no século XX e de difícil acesso ao público por tratar-se de tipo de  
publicação esparsa, motivo pelo qual a elaboração do dicionário se  
justifica plenamente. Os autores então apresentam seus critérios e  
métodos para o trabalho, o acesso às fontes, e o banco de dados.  
Mostram um exemplo dos cento e trinta verbetes da obra, que,  
seguramente, é ponto de partida para o conhecimento do patrimô-  
nio cultural maranhense.

Laísa Bauermeister Stelo e Aparecida Negri Isquerdo, da  
UFMS, analisam vários atlas do Mato Grosso e do Mato Grosso  
do Sul, com relação à ocorrência de “bola de gude”, para a região  
Centro-Oeste. O estudo provém do projeto *Tesouro do Léxico*

*Patrimonial Galego e Português: Brasil*, de natureza internacional, sediado na Universidade de Santiago de Compostela que abrange o galego, o português europeu e o português brasileiro. Os dados encontrados no corpus mostram evidências de influência do espanhol na variedade da região.

Gisele Aparecida Ribeiro, da UEMG, tratou da análise de 81 lexias da variedade do português da Serra da Canastra, em MG. Realizou consultas a dois dicionários gerais do português, fazendo considerações a respeito. Alerta para o estágio inicial de sua pesquisa, podendo esta ser ampliada com novas análises, contudo, aponta que a variedade do português brasileiro estudada é encontrada em região de terras férteis com presença de riqueza material, o que poderia levar a uma certa homogeneidade linguística da região, com extensas redes semânticas, e vocabulário específico.

Jéssica Paula Vescovi e Márcia Sipavicius Seide, da Unioeste, apresentam um estudo antroponímico, abordando dados do basco, na Espanha, e de regiões do Paraná, no Brasil. Tais dados vêm de levantamentos realizados anteriormente, no caso do país basco, de pesquisa pelo governo, e no Brasil, de levantamentos da revista *Superinteressante* e das próprias autoras. Elas procuram observar nomes próprios que seriam globais, encontrados em todos os locais. Posteriormente, observam os nomes próprios regionais e se questionam sobre suas origens. Há, na região brasileira estudada, uma influência da colonização estrangeira do local. Como as comunidades estudadas possuem nomes regionais e globais, concluem que isso se deve ao fato de serem sociedades híbridas, com homogeneização de um lado e resistência à globalização, de outro, o que, certamente, pode ser um fenômeno de âmbito geral.

Maranúbia Pereira Barbosa Doiron aborda o léxico regional, buscando a constituição do ALAL – Atlas Linguístico de Alagoas, que contempla 21 localidades do estado brasileiro, em projeto coordenado por Vanderci de Andrade Aguilera. A autora focaliza seu presente estudo em duas unidades léxicas, encontrando variantes diferentes que ela conclui poderem estar também presentes em outras regiões do país, pois as palavras não podem ser pensadas

como estáticas. O que dialoga com os textos anteriores, que abordam o regional, mas que pode ser também global.

Rosana Maria Sant'Ana Cotrim, professora da UFMT, apresenta análise de nove criações lexicais na obra poética de João Cabral de Melo Neto. Ela as descreve e apresenta abonações na poética do autor. Baseada em quatro dicionários como *corpora* de exclusão, ela trata da neologia estilística. Aponta hipóteses para sua constituição, revelando intenções e mostrando que a crítica histórico-social vem acompanhada de aspectos formais que a constituem enquanto crítica.

### **Linguística de corpus**

Maria José Bocorny Finatto, professora da UFRGS, aborda as metodologias para extração de terminologias em *corpora* digitais através de ferramentas computacionais. Tratando de Linguística de Corpus (LC) e Processamento de Linguagem Natural (PLN), a autora reflete sobre Ciência da Computação e sua relação com a Linguística, e a superação da relação complexa homem x máquina. Ela discute programas e suas possibilidades de uso, seus produtos para a Terminologia e a Terminografia, que, no futuro, serão aliadas das pesquisas em PLN.

Anise D'Orange Ferreira, professora da Unesp, Araraquara, fecha com seu texto este volume, discutindo os projetos atuais na área de línguas clássicas, envolvendo recursos digitais. Dentro da linguística de corpus, o tratamento de extensos *corpora* de grego e de latim tem sido possível através de tais ferramentas. Por isso, longe de ser obsoleta, a área é extremamente ativa em ensino e pesquisa, preocupando-se também com a divulgação de suas descobertas a um público maior, através de sites de acesso irrestrito. As pesquisas, realizadas por meios computacionais, se encontram diretamente ligadas ao ensino-aprendizagem das línguas, com a participação de estudantes como pesquisadores, desde o início de seus estudos, o que se revela em consonância com as tendências teórico-metodológicas mais atuais. Assim, os classicistas da era digital, embora continuadores de uma antiga tradição de estudos, têm hoje ferra-

mentas que os tornam modernos e interligados, em nível internacional, em seus trabalhos, com maior chance de eficácia em seus esforços para o tratamento de extensos corpora.

Concluimos este último volume das publicações do I CINELI, contentes de trazer a público produções de renomados pesquisadores, de clara inserção internacional. Os estudos do Léxico foram contemplados nestes livros, em suas várias áreas, com a participação de pesquisadores representantes de todo Brasil e do exterior. Sem dúvida alguma, esta é uma honra para nossa universidade, anfitriã e criadora do evento, que esperamos que possa promover outros encontros.

# A PESQUISA REVERSA: DICIONÁRIOS CODIFICADORES E DICIONÁRIOS ONOMASIOLÓGICOS

Álvaro IRIARTE SANROMÁN

## Dicionários codificadores e dicionários onomasiológicos

Apesar do princípio de arbitrariedade do signo linguístico, estabelecido por Saussure (1999), há um século, no seu *Curso de Linguística Geral*, muito frequentemente esquecemos que o significado de uma palavra (isto é, o seu sentido) é, de facto, o significado dessa palavra quando combinada com outras palavras. Com efeito, Saussure entende este carácter arbitrário do signo não só como oposto à motivação onomatopeica, no sentido óbvio de que não existe uma relação direta entre o significante e o significado mas apresenta também uma muito mais importante concepção de arbitrariedade que o leva a conceber a língua como uma estrutura em que se estabelecem relações. Saussure fala em arbitrariedade porque, em português, “**passeio**” combina com “**dar**” (dar um passeio), em francês, “*promenade*” combina com “*faire*” (*faire une promenade*), em inglês, “*walk*” combina com “*go for*” ou com “*take*” (*go for a walk / take a walk*).

É essa concepção discreta, ou atomista, da linguagem (uma palavra ligada a um significado; um conjunto de significantes ligados, um a um, a outros tantos significados) que leva, muito frequentemente, os dicionários a apresentarem como aceção de uma

palavra o que, em rigor, é o significado dessa palavra combinada com outras. Veja-se, por exemplo, na entrada **ouvido**<sup>1</sup> do dicionário Aurélio, como a aceção **3** tem o mesmo valor que a combinação “**ter bom ouvido**”, também registrada na parte sintagmática do dicionário:

**ouvido.** [Part. de *ouvir*] S. m. **1.** Faculdade de ouvir, [...] **2.** Anat. Cada um dos dois [...]. **3.** Aptidão para captar com relativa precisão sons musicais ou não, e de reproduzir aqueles sem o auxílio de partitura: *ter bom ouvido* [...]

• [...] **Ter bom ouvido.** Ter fácil percepção de sons, especialmente musicais. (Aurélio. FERREIRA, 1999).

Neste caso (que não é único nos dicionários portugueses) regista-se uma contradição, pois o mesmo significado (‘aptidão para captar com relativa precisão sons musicais’) é atribuído tanto à palavra **ouvido** como à expressão **ter bom ouvido**, que aparece como locução no fim do artigo.

Tal valor corresponde, de facto, à combinação lexical **ter bom ouvido**, devendo ser registado, portanto, na parte sintagmática ou combinatória do dicionário. É o que acontece, por exemplo, no *Caldas Aulete*, um grande dicionário quanto ao tratamento da combinatória lexical, como já indicava Rodrigues Lapa (1984)<sup>2</sup>.

Se o adjetivo **cego** só adquire determinado valor quando combinado com o substantivo **amor** e o adjetivo **mortal** só adquire esse mesmo valor quando combinado com o substantivo **ódio**, não deveríamos explicitar isso nas entradas correspondentes? Contudo, nos dicionários, continuamos a deparar-nos com entradas que recolhem um suposto sentido literal e outro sentido figurado das palavras, consideradas isoladamente:

<sup>1</sup> O exemplo é adaptado de Calderón Campos (1994, p. 58).

<sup>2</sup> “É precisamente neste capítulo da fraseologia, muito importante, que os dicionários correntes deixam mais a desejar. O mais celebrado de entre eles e o mais moderno dos grandes dicionários, o de Cândido de Figueiredo, é muito pobre em grupos fraseológicos, o que constitui um grave defeito, porque é nessas locuções que se imprime o chamado génio da língua. Como repositório de fraseologia, nada há que possa substituir entre nós o *Dicionário Contemporâneo* de Caldas Aulete.” (LAPA, 1984, p. 83).

**cego** [...] **1.** Que não pode ver: ... **2.** FIG. em alto grau, intenso, descontrolado, ... [...];

**mortal** [...] **1.** letal: ... **2.** FIG. em alto grau, intenso, descontrolado, ... [...].

Repare-se que o adjetivo **cego** só tem o significado recolhido na aceção **2** quando combinado com a palavra **amor** (e, ocasionalmente, duas ou três palavras mais). Trata-se de um fenómeno lexical, não semântico ou sintáctico, que devemos assumir como claramente irregular. Por isso, não é possível recorrer aqui ao uso de um “etc”: “**cego** [...] **2.** FIG. em alto grau, intenso, descontrolado [amor, **etc.**]” ou ao uso de um eventual hiperónimo ou nome genérico: “**cego** [...] **2.** FIG. em alto grau, intenso, descontrolado [aplicado a **sentimentos**]”.

Não podemos generalizar dizendo que o adjetivo *cego* tem este significado quando aplicado a “sentimentos”. Não é possível usar estas estratégias simplesmente porque outros substantivos, que também expressam sentimentos, seleccionam adjetivos diferentes para significar “em alto grau, intenso...”, como é o caso de:

*admiração* ⇒ extrema;

*afeto* ⇒ enorme;

*agonia* ⇒ profunda;

*alegria* ⇒ imensa;

*amargura* ⇒ profunda;

*apreço* ⇒ especial;

*beleza* ⇒ infinita;

*consolo* ⇒ enorme;

*culpa* ⇒ grande;

*deceção* ⇒ forte;

*desejo* ⇒ ardente;

*desgosto* ⇒ grande;

*desinteresse* ⇒ total;

*desprezo* ⇒ absoluto;

*dor* ⇒ viva;

*dor* ⇒ forte;

*dor* ⇒ intensa;

*entusiasmo* ⇒ caloroso;

*felicidade* ⇒ indescritível;

*frustração* ⇒ enorme;

*infelicidade* ⇒ grande;

*interesse* ⇒ vivo;

*ódio* ⇒ figadal;

*ódio* ⇒ mortal;

*pânico* ⇒ absoluto;

*pesar* ⇒ grande;

*prazer* ⇒ sumo;

*preocupação* ⇒ enorme;

*saudade* ⇒ imensa;

*saudade* ⇒ infinita;

*tristeza* ⇒ profunda;

*vontade* ⇒ louca;

*etc.*

Da mesma maneira, não é possível encontrar um hiperónimo que abranja todos os nomes que selecionam, por exemplo, a adjetivo **forte** para significar “em alto grau, intenso ...”:

alto +

*preço*

*qualidade*

*conceito*

*voz*

*traição*

*matemática*

*inteligência*

*estilo*

*façanha*

**hiperónimo ?**

Falámos acima na necessidade de explicitar, na entrada correspondente, esta informação (que um adjetivo como **cego** só adquire determinado valor quando combinado com o substantivo **amor**). Mas em que entrada é que deveríamos fornecer esta informação? Sob o lema **cego** ou sob o lema **amor**? Que palavra é que devo procurar (ou deveria poder procurar) quando o que desconheço é que, em português, **passeio** combina com **dar** ou que é **cego** o adjetivo que devo usar para exprimir a ideia de “amor muito intenso, ou não controlado pela razão”? O utilizador não poderá procurar informação no artigo **dar** ou no artigo **cego** porque desconhece que são estas as palavras que em português se combinam com **passeio** e com **amor**, respetivamente, para exprimir os sentidos desejados. Mas é isso o que acontece na maior parte dos dicionários.

O problema coloca-se quando nos situamos na direção de síntese, ou codificação ou produção textual, devido a que as possibilidades colocacionais (combinatórias) variam de uma língua para outra (e, em geral, os usos léxico-sintático e pragmáticos).

Um falante espanhol não terá problemas para entender os significados das combinações “**tirar uma fotografia**” e “**tirar conclusões**”, mas poderá ter dificuldades para produzir tais combinações em português (“tirar una fotografia”, em espanhol, significa “deitar

fora uma fotografia” e as “*conclusiones*”, em espanhol, “*se sacan*”). Neste caso, este utilizador não poderá procurar a informação que necessita sob o lema **tirar** porque desconhece que é esta a palavra que em português se combina com **fotografia** e com **conclusão**. Deveria poder consultar as entradas **fotografia** e **conclusão** para saber qual o verbo que combina com estes substantivos para exprimir o sentido pretendido. Mas, tal como estão elaborados a maior parte dos nossos dicionários, o utilizador só poderá consultar essa informação na entrada **tirar** (justamente a palavra que ele desconhece!):

**tirar** [...] v. [...] **II**. Como verbo suporte de predicação, combina-se com nomes ... **1**. Actos físicos. ≈ FAZER. ... *Tirar uma fotografia* ... **2**. Actos morais ≈ OBTER. *Tirar conclusões* (Academia. CASTELEIRO, 2001).

**tirar** [...] v. **31**. captar (imagem), ... fotografando, ...: *t. uma foto de uma pessoa*. (HOUAISS, 2001).

**tirar** V. t. d. [...] **15**. Fazer (uma fotografia [2]); *Fique aí quieto, vou tirar a fotografia*. **16**. Fazer tirar, parar para tirar (uma fotografia [2]); *Aprontou-se toda para tirar o retrato*. [...] (Aurélio. FERREIRA, 1999).

**TIRAR** V. tr. [...] *Tirar (alguém) o retrato, fazer-se retratar: Fui tirar o retrato para a carteira de identidade.* } || *Tirar o retrato a alguém, fazer-lhe o retrato: [...]* (Caldas Aulete. AULETE, 1987).

**tirar** V. t. [...] Derivar: *tirar conclusões*. [...] (Cândido de Figueiredo. FIGUEIREDO, 1982).

Só no dicionário *Houaiss*<sup>3</sup> e no *iDicionário Aulete* (AULETE DIGITAL, 2008) é que temos também a informação na entra-

---

<sup>3</sup> Aliás, no dicionário Houaiss (2001), sob a informação gramatical da entrada **dar** podemos ler: “**a. 3**) por sua importância, diversas acepções de *dar*, usado como verbo-suporte, estão registradas no corpo deste verbete; diversas outras devem ser procuradas pelo substantivo que faz parte do objeto direto, como de hábito no restante dicionário”.

da fotografia (mas nada na entrada “conclusão”): “**fotografia** [...] **tirar f.**: acionar uma câmara para obter uma imagem [...] (HOUAISS, 2001).”, “**fotografia** [...] **Tirar fotografia**: 1. Fotografar [...] (AULETE DIGITAL, 2008).”

Outra exceção é o pequeno *Dicionário Básico da Língua Portuguesa* (VILELA, 1991), com um magnífico tratamento da questão:

**tirar** [...]

**S. 8** *tirar + nome, sentido VIII*, equivale a um verbo simples: CONCLUIR (FRASE 1), [...] (VILELA, 1991).

**conclusão** [...]

[...] // (pessoa) **tirar conclusões**: (5) – Que conclusões podemos tirar da sua atitude? • (6) - Não quero tirar conclusões erradas do caso. [...]

**S. •** *Tirar conclusões* (frases 5, 6) tem como sins.: CONCLUIR, TIRAR ILAÇÕES OU DEDUÇÕES ... (VILELA, 1991).

Afirmar que o significado de uma palavra é, de facto, o significado dessa palavra quando combinada com outras não implica que estes conjuntos de palavras devam ser considerados como lemas nos nossos dicionários. Há várias razões para que as entradas dos dicionários (lemas) devam continuar a ser palavras individuais (no sentido em que as pessoas instintivamente entendem por palavra: um conjunto de letras delimitadas por dois espaços em branco, um espaço em branco e um signo de pontuação, etc.), nomeadamente, as expectativas conservadoras dos utilizadores de dicionários (COWIE, 1983) e o bem-fazer da tradição lexicográfica, com excelentes amostras de dicionários “*user-friendly*” (BERGENHOLTZ; TARP, 1995).

Num dicionário tradicional (em papel ou digitalizado), podemos procurar informação sobre uma palavra percorrendo os lemas incluídos na nomenclatura, normalmente ordenados por ordem alfabética. Mas o que acontece quando não sei que palavra é que devo procurar? Por exemplo, como se chama a caixa em que se

transportam e comercializam as garrafas de cerveja? Que eufemismo ou termo (politicamente) correto posso usar em vez de um eufemismo? Neste caso usamos os chamados dicionários onomasiológicos, que são dicionários orientados ao conceito, não à palavra. Quando usamos um dicionário onomasiológico, o que procuramos é uma ideia, um conceito e, normalmente, os resultados são as maneiras de exprimir essa ideia em palavras. Para isso, navegamos por lista estruturada de conceitos, ordenada por assuntos, com a ajuda de listas de hiperónimos ou termos mais genéricos, que nos conduzem à palavra que procuramos.

### **Implementação de ferramentas de pesquisa reversa nos dicionários eletrônicos**

A implementação de ferramentas de pesquisa reversa nos dicionários eletrônicos transforma estas obras em dicionários codificadores e dicionários onomasiológicos, permitindo-nos encontrar não apenas as palavras que correspondem a uma ideia (como nos dicionários onomasiológicos) como também que palavra pode ser combinada com outra para exprimir uma ideia (como nos dicionários codificadores).

É comum encontrarmos ferramentas de pesquisa avançada nos dicionários eletrônicos (embora, lamentavelmente, não tanto nas versões *on-line* dos dicionários). Essas ferramentas permitem, por exemplo, pesquisar por formas de palavra, e não apenas por lema; pesquisar palavras ortograficamente semelhantes; pesquisar na microestrutura, e não apenas navegar pela macroestrutura, utilizando mais do que um termo de pesquisa<sup>4</sup>; pesquisar partes do lema, permitindo estudar, por exemplo, a produtividade de determinados afixos; etc.

---

<sup>4</sup> Fazendo, deste modo, um uso onomasiológico e codificador do dicionário:

– “Como se diz endurecer o aço ?” ⇒ (metal + endurecer ⇒ *temperar*)

– “O que acontece à água com o frio?” ⇒ (água + frio ⇒ *desnevado, fresco, gelo, neve*)

– “Quem é o médico dos olhos?” ⇒ (médico + olhos ⇒ *oculista, oftalmiatra, oftalmologista*).

Como qualquer dicionário eletrônico, o *Dicionário Aberto* (DA)<sup>5</sup> possui ferramentas de pesquisa como as que acabamos de apresentar, com a particularidade de que é possível descarregar os resultados das pesquisas.

**Figura 1** – Adjetivos em **-vel** e de advérbios em **-velmente** descarregados do *Dicionário Aberto*

|                     |  |                            |
|---------------------|--|----------------------------|
| ...                 |  |                            |
| <b>Agitável</b>     |  | <b>Abominavelmente</b>     |
| <b>Aglutinável</b>  |  | <b>Admiravelmente</b>      |
| <b>Agradável</b>    |  | <b>Adoravelmente</b>       |
| <b>Agradecível</b>  |  | <b>Afavelmente</b>         |
| <b>Agricultável</b> |  | <b>Affavelmente</b>        |
| <b>Ajuntável</b>    |  | <b>Agradavelmente</b>      |
| <b>Alcançável</b>   |  | <b>Amavelmente</b>         |
| <b>Alcoolizável</b> |  | <b>Amigavelmente</b>       |
| <b>Alheável</b>     |  | <b>Amoravelmente</b>       |
| <b>Aliável</b>      |  | <b>Aprazivelmente</b>      |
| <b>Alienável</b>    |  | <b>Civilmente</b>          |
| <b>Alliável</b>     |  | <b>Comendavelmente</b>     |
| <b>Alterável</b>    |  | <b>Commendavelmente</b>    |
| <b>Amável</b>       |  | <b>Compativelmente</b>     |
| <b>Amigável</b>     |  | <b>Compreensivelmente</b>  |
| <b>Amisível</b>     |  | <b>Compreensivelmente</b>  |
| <b>Amoedável</b>    |  | <b>Consideravelmente</b>   |
| <b>Amoldável</b>    |  | <b>Defensavelmente</b>     |
| <b>Amolgável</b>    |  | <b>Deleitavelmente</b>     |
| <b>Amorável</b>     |  | <b>Deploravelmente</b>     |
| <b>Amortizável</b>  |  | <b>Desagradavelmente</b>   |
| <b>Amotinável</b>   |  | <b>Desamoravelmente</b>    |
| <b>Amovível</b>     |  | <b>Desconversavelmente</b> |
| <b>Amparável</b>    |  | <b>Desculpavelmente</b>    |
| ...                 |  | ...                        |

**Fonte:** Elaboração própria.

Esta característica faz do DA um recurso importante para a investigação linguística, bem como um excelente auxiliar para a elaboração de gramáticas e de outros dicionários. Utilizamos esta capacidade, por exemplo, num exercício em que perguntamos aos

<sup>5</sup> O Dicionário Aberto está disponível na rede, para consulta e para extração automática de informação, em <<http://www.dicionario-aberto.net>>, mas também para uso local, de modo aberto e gratuito. O projeto iniciou-se em junho de 2005, com a transcrição da edição de 1913 dos dois volumes do *Novo Dicionário da Língua Portuguesa*, de Cândido de Figueiredo. Sobre o Dicionário Aberto ver Simões e Farinha (2011), Simões, Iriarte e Almeida (2012, 2014).

alunos se todos os adjetivos formados com o sufixo **-vel** podem, em português, formar advérbios em **-mente**. Após a extração de todos os adjetivos acabados em **-vel** e de todos os advérbios acabados em **-velmente** (vd. Figura 1) e após descarregar todos resultados, estes podem ser facilmente alinhados (vd. Figura 2), respondendo à pergunta formulada de se todos os adjetivos em **-vel** podem formar advérbios em **-mente**, e ajudando a verificar hipóteses que expliquem quais podem e quais não.

**Figura 2** – Adjetivos em **-vel** e advérbios em **-velmente** do *Dicionário Aberto*, alinhados

|                     |                       |  |
|---------------------|-----------------------|--|
| ...                 | ...                   |  |
| <b>Agitável</b>     | -                     |  |
| <b>Aglutinável</b>  | -                     |  |
| <b>Agradável</b>    | <b>Agradavelmente</b> |  |
| <b>Agradecível</b>  | -                     |  |
| <b>Agricultável</b> | -                     |  |
| <b>Ajuntável</b>    | -                     |  |
| <b>Alcançável</b>   | -                     |  |
| <b>Alcoolizável</b> | -                     |  |
| <b>Alheável</b>     | -                     |  |
| <b>Aliável</b>      | -                     |  |
| <b>Alienável</b>    | -                     |  |
| <b>Alliável</b>     | -                     |  |
| <b>Alterável</b>    | -                     |  |
| <b>Amável</b>       | <b>Amavelmente</b>    |  |
| <b>Amigável</b>     | <b>Amigavelmente</b>  |  |
| <b>Amissível</b>    | -                     |  |
| <b>Amoedável</b>    | -                     |  |
| <b>Amoldável</b>    | -                     |  |
| <b>Amolgável</b>    | -                     |  |
| <b>Amorável</b>     | <b>Amoravelmente</b>  |  |
| <b>Amortizável</b>  | -                     |  |
| <b>Amotinável</b>   | -                     |  |
| <b>Amovível</b>     | -                     |  |
| <b>Amparável</b>    | -                     |  |
| ...                 | ...                   |  |

**Fonte:** Elaboração própria.

Mas muito mais interessantes são as funcionalidades avançadas de pesquisa, que poderão transformar o DA num sofisticado dicionário onomasiológico e codificador ao mesmo tempo. Utilizando estas funcionalidades o utilizador poderá procurar, por exemplo, unidades lexicais relacionadas (sinónimos, quase-sinónimos, hiperónimos, hipónimos, merónimos, holónimos, co-ocorrentes, etc.) a partir de uma ou de várias palavras. Estas funcionalidades avançadas, disponíveis para utilizadores registados, podem ser, como dissemos, ferramentas muito úteis para linguistas e investigadores em Processamento da Linguagem Natural.

O caso da “pesquisa ontológica” utiliza a experiência na extração de relações léxico-semânticas mediante técnicas de Processamento de Linguagem Natural (PLN) a partir de padrões léxico-semânticos e léxico-sintáticos, ou padrões de conhecimento (*knowledge patterns*) (HEARST, 1992), com aplicações na análise automática de dados textuais em áreas como a medicina, etc. No nosso caso, e graças à regularidade da estrutura das definições lexicográficas, e com o recurso aos dicionários em formato eletrónico (*Machine-Readable Dictionaries*), podem ser estabelecidos um conjunto de regras ou padrões (HEARST, 1992) utilizando as sequências de palavras que se pretendem encontrar nas definições e que indicam a grande probabilidade de a palavra que segue o padrão estar relacionada com o lema da respetiva definição.

Para o reconhecimento e a extração de relações léxico-conceptuais foram usadas estruturas como (SIMÕES; IRIARTE; ALMEIDA, 2012, p. 297, grifo do autor):

- *o mesmo (ou melhor) que ...* ⇒ Sinonímia (SYN);
- *que não é ...* ⇒ Antonímia (ANT);
- *espécie de ...* ⇒ Hiponímia (HIPO);
- *que tem por tipo...* ⇒ Hiperonímia (HIPER);
- *cada uma das partes que formam ...* ⇒ Meronímia (MERO);
- *composto por...* ⇒ Holonímia (HOLO).

Também se usaram relações calculadas (por exemplo, utilizando a transitividade da relação de hiperonímia), inferindo novos relacionamentos a partir das relações iniciais:

– A propriedade de simetria entre sinónimos: se **a** é sinónimo de **b**, então **b** também é sinónimo de **a** (a SYN b  $\Rightarrow$  b SYN a). Esta relação é bastante produtiva, permitindo recuperar sinónimos que apenas foram recolhidos pelo lexicógrafo numa das entradas das palavras envolvidas.

– A relação de co-hiponímia pode ser calculada a partir das relações de hiponímia: se duas palavras **a** e **c** são hipónimos da mesma palavra **b**, então **a** e **c** são co-hipónimos (a HIPO b  $\wedge$  c HIPO b  $\Rightarrow$  a COHIPO c).

– A transitividade das relações hierárquicas, como é o caso da hiperonímia/hiponímia, permite que se possa procurar termos genéricos com base num termo mais específico ou termos específicos usando termos mais genéricos (a HIPO b  $\wedge$  b HIPO c  $\Rightarrow$  a HIPO c). Assim, podemos encontrar entradas referentes a “animais” a partir da pesquisa de termos mais específicos como “mamífero” ou “peixe” (hipónimos de “animal”).

Evidentemente, o reconhecimento e a extração automáticos de relações léxico-conceptuais não é fácil. Somos conscientes de que algumas regras podem ser problemáticas (por exemplo, no caso dos sinónimos e quase-sinónimos). Em todo o caso, parece-nos preferível apresentar um conjunto de possíveis falsos resultados do que garantirmos a correção e diminuir drasticamente o número de relações da ontologia resultante.

Na Figura 3 temos uma amostra de pesquisa reversa normal, em que todos os resultados contêm o termo introduzido na janela de pesquisa (“**mamífero**”).

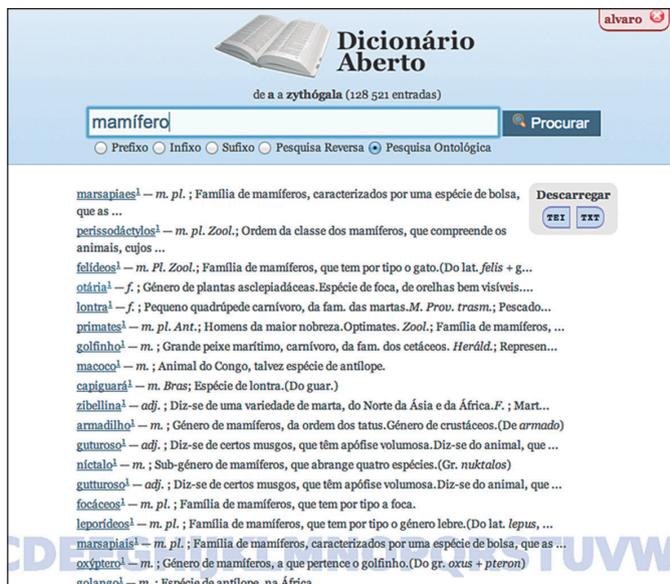
Figura 3 – Pesquisa reversa a partir do termo “mamífero”

The screenshot shows the 'Dicionário Aberto' interface. At the top, there is a search bar containing the word 'mamífero'. Below the search bar, there are radio buttons for different search types: Prefixo, Infixo, Sufixo, Pesquisa Reversa (which is selected), and Pesquisa Ontológica. To the right of the search bar is a 'Procurar' button. Below the search bar, there is a list of search results, each with a small icon and a definition. The results include terms like 'aguti', 'almiqui', 'anomalu', 'antamba', 'aperca', 'ariranha', 'arminho', 'arvícola', 'baleia', 'cabra', 'caititu', 'callocéfalo', 'calocéfalo', 'canguru', 'capivara', 'carneiro-almiscarado', 'castor', 'caxinguelé', and 'cazu'. A 'Descarregar' button is also visible on the right side of the results list.

Fonte: Dicionário Aberto.

Ao contrário do que acontece com a pesquisa reversa normal, graças à extração automática de relações lexicais, na pesquisa ontológica (Figura 4), alguns dos resultados são entradas que não contêm nenhum dos termos introduzidos pelo utilizador na janela de pesquisa, neste caso, “**mamífero**”:

Figura 4 – Pesquisa ontológica a partir do termo “mamífero”



The screenshot shows the 'Dicionário Aberto' interface. At the top, there is a search bar containing the word 'mamífero'. Below the search bar, there are radio buttons for different search types: Prefixo, Infixo, Sufixo, Pesquisa Reversa, and Pesquisa Ontológica (which is selected). To the right of the search bar is a 'Procurar' button. Below the search bar, there is a list of search results for 'mamífero'. The results include terms like 'marsupiais', 'perissodactylos', 'felídeos', 'otária', 'lontra', 'primates', 'golfinho', 'macaco', 'capiguará', 'zibellina', 'armadillo', 'gutuoso', 'nictalo', 'gutuoso', 'focáceos', 'leporídeos', 'marsupiais', 'oxypetero', and 'zolanero', each followed by a brief definition. On the right side of the results, there is a 'Descarregar' button with 'TXT' and 'PDF' options.

Fonte: Dicionário Aberto.

## Conclusões

Estamos convictos de que o DA é uma excelente ferramenta que pode ser usada como um dicionário tradicional, mas também como um recurso para tarefas de processamento de linguagem natural, como um auxílio para a elaboração de gramáticas e de outros dicionários bem como uma ferramenta para ajudar na verificação de hipóteses colocadas na investigação linguística.

O facto de ser um recurso aberto poderá também significar que as suas funcionalidades poderão vir a aumentar em quantidade e em qualidade.

Por outro lado, e embora a evolução do conteúdo linguístico do dicionário seja lenta, a implementação de algoritmos e a execução de experiências sobre o DA têm sido bastante proveitosa, demonstrando que é possível criar funcionalidades úteis a partir de dicionários convencionais.

## REFERÊNCIAS

AULETE, F. J. C. **Dicionário Contemporâneo da Língua Portuguesa Caldas Aulete**. 5. ed. Rio de Janeiro: Delta, 1987.

AULETE DIGITAL. 2008. Disponível em: <<http://www.aulete.com.br/>>. Acesso em: 25 ago. 2014.

BERGENHOLTZ, H.; TARP, S. (Ed.). **Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries**. Amsterdam; Philadelphia: John Benjamins Publishing Company, 1995.

CALDERÓN CAMPOS, M. **Sobre la elaboración de diccionarios monolingües de producción: Las definiciones, los ejemplos y las colocaciones léxicas**. Granada: Ed. de la Universidad de Granada, 1994.

CASTELEIRO, J. M. (Coord.). **Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa**. Lisboa: Academia das Ciências de Lisboa: Editorial Verbo, 2001.

COWIE, A. P. On Specifying Grammar: On Specifying Grammatical Form and Function. In: HARTMANN, R. R. K. (Ed.). **Lexicography: Principles and Practice**. London: Academic Press, 1983. p. 99-107.

FERREIRA, A. B. de H. **Novo Aurélio Século XXI: O Dicionário da Língua Portuguesa**. Rio de Janeiro: Nova Fronteira, 1999.

FIGUEIREDO, C. de. **Dicionário da Língua Portuguesa**. 16. ed. Lisboa: Livraria Bertrand, 1982.

HEARST, M. Automatic acquisition of hyponyms from large text corpora. **Proceedings of the Fourteenth International Conference on Computational Linguistics**, Nantes, France, v.2, p.539-545, 1992.

HOUAISS, A. (Coord.). **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objectiva, 2001.

LAPA, M. R. **Estilística da Língua Portuguesa**. Coimbra: Coimbra Editora, 1984.

SAUSSURE, F. **Curso de linguística geral**. Lisboa: Publicações Dom Quixote, 1999.

SIMÕES, A.; FARINHA, R. Dicionário Aberto: Um novo recurso para PLN. **Vice-versa**, Vigo, Espanha, v. 16, p.159-171, 2011.

SIMÕES, A.; IRIARTE, Á.; ALMEIDA, J. J. Dicionário-Aberto: Construção semiautomática de uma funcionalidade codificadora. In: CONGRES INTERNATIONAL DE LINGUISTIQUE ET DE PHILOLOGIE ROMANES, 27., Nancy. **Actes...** Nancy : ATILF, 2014. Section 16: Projets en cours; ressources et outils nouveaux. Disponible en: <<http://www.atilf.fr/cilpr2013/actes.php>>. Acesso em: 25 ago. 2014.

SIMÕES, A.; IRIARTE, Á.; ALMEIDA, J. J. Dicionário aberto: a source of resources for the portuguese language processing. In: CASELI, H. et al. (Ed.). **Computational Processing of the Portuguese Language, Lecture Notes for Artificial Intelligence**. Berlim: Springer, 2012. p.121-127.

VILELA, M. **Dicionário do Português Básico**. Porto: Edições Asa, 1991.



# VERBETES QUE SONHAMOS... VERBETES QUE FAZEMOS: A EQUIVALÊNCIA EM RECENTES TRABALHOS LEXICOGRÁFICOS

Claudia ZAVAGLIA

Quando recebi o título desta mesa, ou seja, *Dicionários que queremos... Dicionários que fizemos: reflexões sobre Lexicografia*, foi-me quase instantâneo pensar nos **verbetes que queremos para esses dicionários e os verbetes que fizemos**. E continuei pensando nos verbetes que fazemos, hoje, e naqueles que sonhamos, ou com aqueles que pelo menos eu sonho!

Comecei então a indagar por que muitas vezes quando vamos consultar os dicionários Houaiss e Aurélio começamos a reclamar deles? Só porque todo mundo faz isso? Não creio. De fato, em minhas pesquisas sempre tive e tenho um olhar crítico para esses dicionários, admito, o que não faz, entretanto, com que eu deixe de reconhecer o valor inestimável deles.

Mas por que me incomodo então, por exemplo, quando vou em busca de uma palavra no Houaiss? Vou me ater a esse dicionário neste trabalho simplesmente para mudar de “alvo” e não bater sempre na mesma tecla em relação ao Aurélio.

Vejam, pois, um exemplo, escolhido de forma aleatória, de uma palavra-entrada do Houaiss (2009), sobre a qual eu desconhecia seu significado:

abacataia

☐ substantivo feminino

Rubrica: ictiologia. Regionalismo: Brasil.

f. não pref. de *abacatuiaia*

A primeira coisa que passaria pela minha cabeça em relação ao seu significado seria “ter algo a ver” com o abacate, pela indução que a própria raiz da palavra me oferece. E também porque, numa leitura vertical do dicionário, ela se encontra após **abacatada** (aquele creme de abacate batido com leite ou amassado com açúcar) e seguida por **abacatal** (o mesmo que abacateiral) e **abacate**.

Em sua rubrica, entretanto, vejo que estou no domínio da ictiologia (ramo da zoologia que estuda os peixes) e então me dou conta de que estou no reino animal e não mais no vegetal, como pensava até dois minutos atrás. Deduzo então que se trata de um peixe. Mas não quero deduções, estou à procura de informações verossímeis. Em seguida, procuro a definição de **abacataia** e qual não é a minha frustração quando percebo que não existe explicação sobre essa entrada, apenas uma redução da frase “forma não preferencial de” e a indicação de outra palavra, ou seja, **abacatuiaia**. Como também não sei o que é **abacatuiaia**, vou até o seu verbete (HOUAISS, 2009) e o que vejo é:

abacatuiaia *Datação*: c1631

☐ substantivo feminino

Rubrica: ictiologia.

1 Regionalismo: Brasil.

m.q. *xaréu-branco* (*Alectis ciliaris*)

2 Regionalismo: Pernambuco.

m.q. *peixe-galo* (*Selene vomer*)

Obs.: f. não pref.: *abacataia*, *abacatuia* e *abacutaia*

E agora sei que **abacatuiaia** é o mesmo que **xaréu-branco**, na rubrica “regionalismo Brasil”, e o mesmo que **peixe-galo**, na rubrica “regionalismo Pernambuco”, como se Brasil e Pernambuco fossem

dois países diferentes ou estados, não sei. Mas o problema das marcas de uso ou rubricas é outro e por si só demandaria algumas outras mesas-redondas específicas para o início da discussão... Enfim, consigo saber que **abacataia** é a forma não preferencial de **abacatuiaia** que por sua vez é o mesmo que **xaréu-branco** e **peixe-galo**. E sei que **abacatuiaia** é um peixe. E só. Qual o seu tamanho, cor, habitat, se é comestível ou não eu não sei. Porque, mais uma vez, não há definição. Ah, sei também que é do Brasil e “do” Pernambuco. Continuando, quero saber então como é o peixe **xaréu-branco** (HOUAISS, 2009), já que os cromônimos me interessam há pelo menos duas décadas. E então me deparo com:

xaréu-branco

□ substantivo masculino

Rubrica: ictiologia. Regionalismo: Brasil.

1 peixe teleósteo, perciforme, da fam. dos carangídeos (*Alectis ciliaris*), de distribuição circuntropical, com cerca de 1 m de comprimento, dorso verde-azulado, ventre prateado e nadadeiras dorsal, anal e ventrais com raios longos enegrecidos, semelhante ao peixe-galo; abacataia, abacatuiaia, abacatuia, abacutaia, abucataia, aletto, aracambé, aracanguira, galo, galo-de-fita, galo-do-alto, peixe-galo-do-brasil

2 m.q. *xaréu* (*Caranx hippos*)

3 m.q. *araximbora* (*Caranx guara*)

E me alegro, pois, finalmente, estou diante de uma definição, na primeira acepção do verbete. Ora, após a leitura da definição da entrada, tenho o perfeito conhecimento do que é um **xaréu-branco** que é o mesmo que **abacatuiaia** que é a forma preferencial de **abacataia**. Ledo engano, pois não tenho! Mas ainda tenho uma carta na manga que é o **peixe-galo**! E vou a sua procura (HOUAISS, 2009):

peixe-galo *Datação*: 1899

☐ substantivo masculino

Rubrica: ictiologia. Regionalismo: Brasil.

1 design. comum aos peixes teleósteos, perciformes, da fam. dos carangídeos, do gên. *Selene*, que possuem o corpo muito alto e extremamente comprimido; galo-do-morro

2 m.q. *galo-do-fundo* (*Zenopsis conchifer*)

Enfim, chegamos, após uma pequena viagem lexicográfica, à definição 1: “design. comum aos peixes teleósteos, perciformes, da fam. dos carangídeos, do gên. *Selene*, que possuem o corpo muito alto e extremamente comprimido”. E então, eis que surge “peixes teleósteos” e você não consegue decodificar o sentido porque não tem ideia do que seja o adjetivo “teleósteos”. Mais uma vez, volta-se ao dicionário (HOUAISS, 2009) e:

teleósteo *Datação*: 1899

☐ adjetivo

1 relativo aos teleósteos

☐ substantivo masculino

Rubrica: ictiologia.

2 espécime dos teleósteos

☐ *teleósteos*

☐ substantivo masculino plural

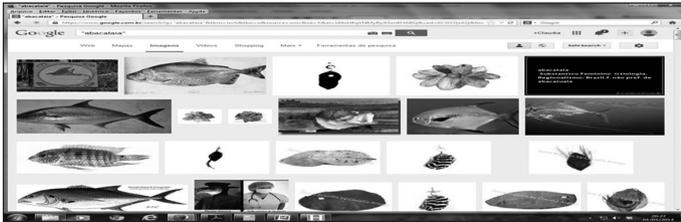
Rubrica: ictiologia.

3 subclasse dos actinopterígios, que abrange a maior parte de todos os peixes, com 38 ordens, 426 fam., 4.064 gên. e cerca de 23.637 spp.; são os peixes ósseos propriamente ditos

E após a leitura desse verbete, penso em desistir de querer saber o significado de **abacataia**! Além disso, percebi que ainda não tinha sequer em minha mente uma imagem desse tipo de peixe. No verbete **teleósteo**, estão as definições que eu poderia chamar de “balança caixão”, quer dizer, fazendo referência à brincadeira infantil, são aquelas que vão se esconder, no labirinto das palavras que é o dicionário. E acabo não tendo como entender o que é o peixe

**abacataia**, estando prestes a renunciar. Entretanto, penso ainda em outra estratégia que é a nossa arca de Noé atual: a Sra. Internet e o Sr. Google e jogo todas as minhas dúvidas lá, tanto em forma de palavras quanto em imagens. E vejamos:

**Figura 1 – Abacataia**



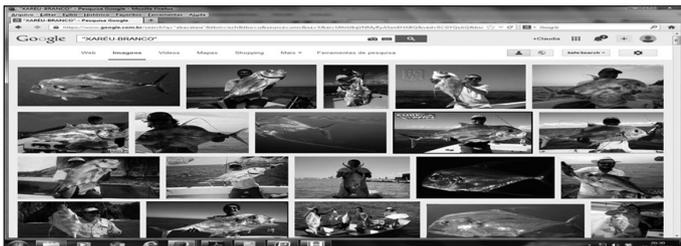
Fonte: Google Imagens.

**Figura 2 – Abacatuia**



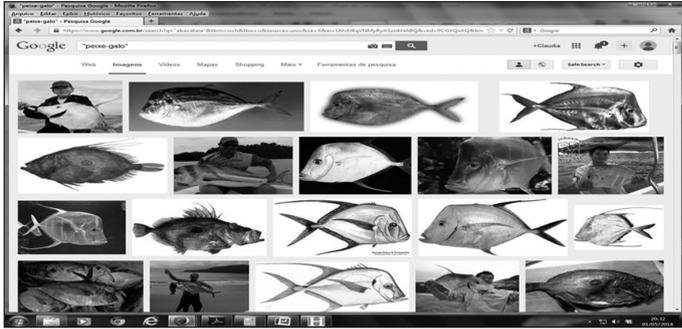
Fonte: Google Imagens.

**Figura 3 – Xaréu-branco**



Fonte: Google Imagens.

Figura 4 – Peixe-galo



Fonte: Google Imagens.

E de todas as imagens que capturei com as buscas que realizei, o máximo que consegui enxergar de “sintonia” nelas foi uma mesma foto em **abacatuia** e em **peixe-galo** que seria a que segue:

Figura 5 – Abacatuia e peixe-galo



Fonte: Google Imagens.

Além disso, essas imagens não conseguiram fazer com que eu atingisse meu objetivo, qual seja, saber, apenas, o significado de **abacataia**, e tampouco me fizeram ter em mente uma imagem “segura” dele. Sei apenas que se trata de um peixe.

Essa é uma pequena amostra do que vem a ser um sistema de remissivas em um dicionário, ou seja: procuramos a entrada **abacataia** que remete a **abacatuia** que por sua vez remete a **xaréu-branco** e **peixe-galo** e assim por diante, como vimos. Tudo isso para ilustrar o que vivencia um consultante de um dicionário monolíngue no nosso país.

E quando pensamos em dicionários bilíngues? Esse sistema de remissivas está presente nas definições? A propósito, as definições existem? O consulente se satisfaz com a busca que realiza para dirimir suas dúvidas? E indaguei-me: Incomodo-me quando vou em busca de uma palavra em Dicionários Bilíngues (DBs)?

Em 2001, escrevi um artigo em coautoria com Adriana Zavaglia, que foi publicado pela revista *Linguística* da ALFAL, no qual discutíamos, justamente, a questão da feitura de verbetes para dicionários bilíngues e/ou plurilíngues e das inadequações semânticas, os vazios linguísticos e as definições deficitárias nesse tipo de obra. Citando Schmitz (1998, p.162), um dicionário bilíngue tradicional pode ser de vários tamanhos, arrola um número pequeno de vocábulos e “conduz a uma superficialidade na apresentação das equivalências nas duas línguas”. De fato, os dicionários bilíngues privilegiam o uso de sinônimos da palavra-entrada, partindo-se do pressuposto que um termo de uma língua pode ser definido e compreendido a partir de um equivalente na língua de chegada, podendo, além disso, prescindir de explicações definitórias.

Como minha segunda língua é o italiano, usarei e farei exemplos de dicionários bilíngues para o par de línguas italiano e português. Trabalharei com quatro dicionários: Mea (2000), Martins Fontes (BENEDETTI, 2004), Michaelis *Online* (POLITO, 2009) e Dicionário Escolar WMF (2013).

Aleatoriamente, abro o dicionário Mea (2000), direção italiano-português, e encontro a entrada *dateria*, cuja entrada me remete a *dataria*:

**dateria** s.f. → **dataria**.

No lema *dataria*, encontro como equivalentes em português *dataria* e **dataria apostólica**, sob a rubrica de “religião”:

**dataria** s.f. RELIGIÃO *dataria*, *dataria apostólica*.

No Martins Fontes há somente a entrada **dataria**, com o equivalente **dataria** e nos outros dois dicionários, essa entrada não consta.

No Martins Fontes (BENEDETTI, 2004), a entrada *battitrice* me remete a *trebbiatrice* que por sua vez traz o equivalente **debulhadora**:

**battitrice** s.f. → **trebbiatrice**.  
**trebbiatrice** s.f. **debulhadora**.

O mesmo se verifica no Mea, que acrescenta o equivalente **laminadora**. Nos outros dois dicionários, essa palavra não consta.

No WMF (DICIONÁRIO ESCOLAR WMF, 2013), a entrada *giullare*, substantivo masculino, me oferece o equivalente **jogral**:

**giullare** s.m. **jogral**.

No Mea (2000), há: 1. Jogral, menestrel; 2. Bobo, bufão, fanfarrão, histrião:

**giullare** s.m. 1. Jogral. Menestrel; 2. Bobo, bufão, fanfarrão, histrião.

No Martins Fontes (BENEDETTI, 2004):

**giullare** s.m. jogral, menestrel. Saltimbanco. *Giullare di Dio*, autor de cantos de louvor a Deus. (spregiativo) bufão, truão.

No Michaelis (POLITO, 2009), temos:

**giullare** s.m. 1 menestrel, trovador. 2 bobo da corte, bufão.

Ora, se eu desconheço o que é **dataria**, **debulhadora** ou **laminadora**, **jogral**, **menestrel** e **trovador**, em português, na primeira pesquisa que fiz, esses dicionários pouco me ajudaram e não cumpriram sua função.

A ausência de definições é constante nos quatro dicionários, exceto quando ela é necessária para explicar a palavra na ausência de um equivalente para outras entradas que não essas. Note-se,

ademais, que os exemplos acima contemplados não trazem nenhum tipo de contextualização dos itens lexicais, seja por meio de abonações ou frases forjadas, o que poderia ajudar enormemente na compreensão de seu significado.

Sabemos que a Lexicografia plurilíngue foi a precursora da Lexicografia monolíngue, com os tabletes monolíngues em escrita cuneiforme, produzidos pelos sumérios, da antiga Mesopotâmia, em 2600 a.C. (repertoriando signos, nomes de profissões, de divindades e de objetos usuais). Entretanto, o que se observa, ainda hoje, é que a lexicografia plurilíngue não se desenvolveu ou evoluiu como aquela monolíngue, uma vez que nos atuais Dicionários Bilíngues (DBs) que se encontram em circulação, das mais variadas línguas, entre elas o italiano, o francês e o inglês, as lacunas existentes não são poucas, como acabamos de observar, dessa breve exposição.

A grosso modo, a tradução é o processo pelo qual uma mensagem em uma língua x é transferida para uma língua y. Estão em jogo: combinatórias gramaticais, morfologia, sintaxe, itens lexicais etc. de duas línguas diferentes, que devem “dar conta” de “significar” o que foi dito numa língua para outra, ou seja, é necessário reproduzir uma mensagem equivalente na língua de chegada e transferir um conceito dito.

Em termos de dicionários bilíngues, acredito que, como diz Rodrigues, a partir de Lefevere, que exista a **possibilidade de recuperação do significado de reprodução do conteúdo, subjacentes às abordagens linguísticas**. De fato, esse autor ainda aconselha os futuros tradutores a **buscar equivalentes em um dicionário bilíngue**, fazendo a ressalva de que **a língua-alvo pode não ter equivalentes exatos**, o que pressupõe que ele acredite que possa existir o equivalente exato (RODRIGUES, 2000 apud LEFEVERE, 1992).

É claro que esses equivalentes, em um dicionário, não correspondem a uma fixidez da língua; eles representam possibilidades de emprego para diferentes contextos, além de serem a imagem plena da busca incessante pela correspondência entre as línguas.

No âmbito da Lexicografia Bilíngue, o lexicógrafo não pode prescindir dos equivalentes, pois sem eles o seu trabalho não se concretizaria.

Nos estudos clássicos da tradução, o conceito de equivalência calcava-se numa relação biunívoca entre elementos de línguas diferentes. Em consonância, Milton (2010) explicita e discute algumas metáforas associadas ao tradutor e ao seu trabalho ao longo da história, muitas das quais são reveladoras de uma concepção conservadora de tradução segundo a qual o conteúdo do original pode ser transferido ao texto traduzido. Para esse autor, existe um elemento comum nessas metáforas: a discussão entre forma e conteúdo. Tais metáforas refletem que a importância da preservação e da transmissão do conteúdo do texto original deveria se sobrepor à forma como ele chegaria à língua-alvo.

Na década de 60, uma nova interpretação do conceito de equivalência, desatrelada da noção clássica de biunivocidade interlinguística, desponta especialmente com a abordagem desconstrutivista da tradução. Segundo os preceitos dessa teoria, os textos não devem ser entendidos como contendo significados fixos e estáveis neles inscritos, para os quais o tradutor precisa encontrar substitutos de mesmo valor na língua-alvo (MENDES, 2013).

Em relação à significação, Rodrigues (2000) parte do princípio de que não existe um significado universal e transcendental; não há um sentido “gravado” nos textos que se recupere imediatamente e reflita única e exclusivamente os supostos propósitos e intenções de um escritor ou autor. O sentido, dessa forma, seria resultado da inter-relação de diversos fatores, tais como: intenções do autor, material linguístico, situação contextual e comunicativa, entre outros, sendo a interpretação, especialmente, um fator primordial dessa equação.

Em Lexicografia Bilíngue, o uso do termo “equivalente” é abundante e o conceito de equivalência é empregado de maneira diferente da abordagem sugerida pelas teorias de tradução pós-modernas. Os chamados “equivalentes” correspondem à definição sinonímica de um dicionário apenas de forma parcial, uma vez que não há coincidência entre os escopos semânticos de itens lexicais de

sistemas diferentes. Assim, torna-se não exequível agrupar “todas” as possíveis traduções de um item lexical, em cada contexto específico para serem inseridas na microestrutura de um DB.

O papel da Lexicografia Bilíngue é, pois, aquele de tentar diminuir e de desfazer as barreiras interlinguísticas, sendo considerado “equivalente” um item lexical que contenha a maior quantidade possível de traços semânticos em comum com a unidade lexical de outra língua. Essa noção de equivalência torna, portanto, o termo adequado à LB.

Nessa linha, divido com vocês a feitura de verbetes para um projeto recente em Lexicografia Bilíngue do qual participei. Trata-se do *Dicionário Multilíngue de Regência Verbal: verbos preposicionados* [DMRV] (XATARA; ZAVAGLIA; SILVA, 2013).

Nesse dicionário, cada palavra-entrada possui:

- (i) uma entrada em língua portuguesa do Brasil;
- (ii) a acepção frequente com especificação esquemática dos complementos (direto e indireto) ou apenas do complemento direto, com destaque nas preposições exigidas e algum(ns) sinônimo(s) esclarecedor(e)s e elucidativo(s);
- (iii) um ou mais exemplos criados adequadamente para ilustrar o sentido, com contextos simples e
- (iv) os equivalentes em itálico nos seis idiomas.

Vejamos o *layout*:

**Figura 6** – Dicionário Multilíngue de Regência Verbal: verbos preposicionados

|  |   |
|--|---|
| <p><b>HABILITAR</b></p> <p><b>H</b></p> <p><b>HABILITAR</b></p> <p>1. alguém A, PARA algo (capacitar, preparar; tornar apto)</p> <p>→ Habitilaram-na ao (para o) novo cargo.</p> <p><b>A:</b> <i>jmdn auf etwas vorbereiten</i><br/> <b>E:</b> <i>capacitar a alguien para</i><br/> <b>F:</b> <i>préparer à</i><br/> <b>I:</b> <i>prepare for</i><br/> <b>It:</b> <i>preparare per</i><br/> <b>J:</b> <i>(だれか) wo (何か) no tameni junbi saseru</i> ~を ~のために 準備させる</p> <p>alguém A + verbo no infinitivo</p> <p>→ Todo o seu sofrimento habitou-a a enfrentar futuros obstáculos.</p> <p><b>A:</b> <i>jmdn auf etwas vorbereiten</i><br/> <b>E:</b> <i>preparar a alguien para</i><br/> <b>F:</b> <i>habilitar à</i><br/> <b>I:</b> <i>enable to</i><br/> <b>It:</b> <i>preparare per, a; rendere capace di</i><br/> <b>J:</b> <i>(だれか) wo (動詞) koto ni junbi saseru</i> ~を ~ことに準備させる</p> <p><b>HABITAR</b></p> <p>1. EM algum lugar (morar, residir)</p> <p>→ Habitava em uma pequena cidade do interior.</p> <p><b>A:</b> <i>wohnen in</i><br/> <b>E:</b> <i>vivir en</i><br/> <b>F:</b> <i>habiter à, dans, en, sur, Ø</i><br/> <b>I:</b> <i>live in</i><br/> <b>It:</b> <i>vivere, abitare in</i><br/> <b>J:</b> <i>(どこか) ni sumu</i> ~に住む</p> | <p><b>HABITUAR</b></p> <p>1. alguém A algo ou + verbo no infinitivo (acostumar, exercitar, familiarizar com)</p> <p>→ Habitaram os estagiários à nova metodologia da pesquisa.<br/>         → Desde pequena, a mãe habitou a filha a arrumar a cama.</p> <p><b>A:</b> <i>jmdn an etwas gewöhnen</i><br/> <b>E:</b> <i>habituat a alguien a</i><br/> <b>F:</b> <i>habituat à</i><br/> <b>I:</b> <i>accustom to</i><br/> <b>It:</b> <i>abituare a</i><br/> <b>J:</b> <i>(だれか) wo (何か) ni nasesaseru</i> ~を ~に 慣れさせる</p> <p><b>HARMONIZAR</b></p> <p>1. algo COM algo (conciliar, congraçar)</p> <p>→ Harmonizava seus conhecimentos anteriores com os recém-adquiridos.</p> <p><b>A:</b> <i>etwas mit etwas in Einklang bringen</i><br/> <b>E:</b> <i>armonizar con</i><br/> <b>F:</b> <i>harmoniser con</i><br/> <b>I:</b> <i>attune to</i><br/> <b>It:</b> <i>conciliare con</i><br/> <b>J:</b> <i>(何か) wo (何か) to chōwa-suru</i> ~を ~と 調和する</p> <p>2. ~ (-se) COM algo (estar em harmonia; não destoar de)</p> <p>→ A música harmoniza-se com o ambiente.</p> <p><b>A:</b> <i>mit etwas harmonieren; zu etwas passen</i><br/> <b>E:</b> <i>armonizarse, estar en armonía con</i><br/> <b>F:</b> <i>s'harmoniser à, avec</i><br/> <b>I:</b> <i>harmonise with</i><br/> <b>It:</b> <i>essere adatto a; essere in armonia con</i><br/> <b>J:</b> <i>(何か) to chōwa ga toreteiru</i> ~と調和が取れている</p> |
|--|---|

Fonte: Xatara; Zavaglia; Silva (2013, p. 351).

Para a proposta dos equivalentes nas línguas estrangeiras, o DMRV apregoa:

1. Propor um verbo equivalente para cada acepção apresentada, ou seja, uma tradução (mesmo parafrásica) para as línguas estrangeiras que corresponda ao significado indicado;
2. Esse equivalente pode ser um verbo que exija preposição ou não;

3. Caso o equivalente se refira, por exemplo, apenas a “algo” enquanto o verbo em português referir-se a “algo” e a “alguém”, isso deve ser explicitado mediante o pronome indefinido, escrito na LE.

Na busca por equivalentes em língua italiana no DMRV, que deveriam “corresponder ao significado indicado, com ou sem preposição” da entrada verbal, a tarefa não deixou de ser complexa, mesmo não existindo, em sua microestrutura, estruturas frasais ou idiomáticas; tratava-se da tradução, apenas, do verbo. E foi, justamente, esse o grande entrave dessa empreitada, ou seja, a falta de contextualização e a restrição de exemplos elaborados somente em língua portuguesa. De fato, muitas vezes, o exemplo forjado para o verbo não condizia com o significado indicado ou então não era realizável, daquele modo, no italiano.

Na entrada “Ferrar” (XATARA; ZAVAGLIA; SILVA, 2013), na acepção 3, a seguinte indicação era fornecida:

|   |
|---|
| 3. algo EM alguém (aplicar, dar, introduzir)<br>→ O ganso ferrou uma bicada no caçula da fazenda. |
|---|

O fato de no exemplo constar “ferrou uma bicada no caçula” fez com que a equipe a traduzisse por *mordere Ø*. Entretanto, a indicação semântica era “aplicar, dar, introduzir”, o que incitou a reflexão de que era possível, portanto, construções do tipo “alguém ferrar um beijo, um soco, um pontapé etc em alguém” e a tradução realizada em italiano com o verbo *mordere* (morder) não seria possível para as outras construções, além daquele contida no exemplo. De fato, não seria possível semanticamente em italiano *Il ragazzo ha morso un bacio nella madre\** (O garoto mordeu um beijo na mãe\*, numa tradução literal). Com isso, o correspondente tradutório em italiano passou a ser *dare a*, já que esse verbo comportaria as possibilidades de frases descritas anteriormente, em todos os níveis semânticos.

Já para a acepção 8, do verbo “Ficar” (XATARA; ZAVAGLIA; SILVA, 2013), a semântica atual do português influenciou na escolha

do equivalente italiano. A indicação relativa ao significado para essa acepção era:

8. COM alguém (namorar sem compromisso, por um curto espaço de tempo)  
→ Na festa, Rita ficou com o melhor amigo do ex-namorado.

que poderia levar a uma tradução como *flirtare*, *stare con*. Entretanto, hoje em dia, “ficar” deixou de ser apenas “namorar por pouco tempo ou sair com alguém e dar alguns beijos apenas por um dia”; hoje, “ficar” significa, frequentemente, “transar” com alguém de modo eventual ou por um curto espaço de tempo também. Desse modo, a opção foi contemplar essas nuances semânticas no italiano, sugerindo-se *stare*, *flirtare*, *andare con* como equivalentes, sendo que este último refletiria o “transar” em português, mesmo sabendo-se que “ficar em, durante uma festa” fosse mais adequado apenas *stare* o *flirtare* com alguém.

Para casos como os dos verbos “Fixar” e “Viver”, tomou-se a decisão de contar com o conhecimento e discernimento linguístico do usuário do dicionário para a escolha de qual das duas preposições oferecidas empregar em suas traduções. Os exemplos eram “Depois da faculdade, fixou-se em São Paulo” e “Antes eles viviam no interior; agora vivem em Porto Alegre”, ambos com a indicação de “em algum lugar”. Ocorre que em português a preposição empregada para os dois casos é “em”, ao passo que para a língua italiana, dependendo do lugar em que se mora ou se estabelece, emprega-se uma ou outra preposição, no caso, *a*, para cidades, ou *in*, para países, ilhas, estados. Dessa forma, os equivalentes foram: *stabilirsi a*, *in* para o verbo *fixar* e *vivere in*, *a* para o verbo *viver*, esperando que o consulente tenha em mente o uso das preposições para essas localidades no italiano. Caso parecido ocorreu com a tradução do verbo “Formar” com o sentido de “diplomar-se/graduar-se por algo”, como na frase “Formei-me por uma universidade de renome”, cuja tradução foi *laurearsi presso*, sem a indicação de que em seguida, na língua

italiana, deve-se usar um artigo, como por exemplo na frase *Mi sono laureato presso l'Università di Genova*.

Interessante relatar o caso do verbo “Foder” na acepção de “acabar, arruinar e desgraçar”, por meio da frase “Aquele casamento fodeu com a minha vida”, em que se verifica o uso figurado desse verbo em uma linguagem extremamente coloquial, até mesmo vulgar para muitos, cujo equivalente em italiano é *rovinare* Ø, que numa tradução livre significa “arruinar, destruir, corromper, estragar”, ou seja, um verbo que não possui toda a carga semântica daquele em português, visto que o verbo *fottere* (foder), em italiano, parece não possuir realizações com esses sentidos.<sup>1</sup>

Inúmeros foram os casos de equivalência zero, ou seja, a tradução por meio de verbos em italiano sem complemento preposicional, eis alguns exemplos:

**Tabela 1** – Casos de equivalência zero do DMRV

| em italiano   | em português                   |
|---|--------------------------------|
| <b>fragmentar em algo (dividir, fracionar)</b>                          | <i>frantumare</i> Ø            |
| <b>ganhar de alguém (vencer)</b>  | <i>battere, vincere</i> Ø      |
| <b>incidir em alguém (acometer, atacar, atingir)</b>                    | <i>colpire</i> Ø               |
| <b>incorrer em algo (incidir)</b>                                       | <i>commettere</i> Ø            |
| <b>influenciar em algo (influir)</b>                                    | <i>influenzare</i> Ø           |
| <b>judiar de algo (danificar)</b>                                       | <i>danneggiare</i> Ø           |
| <b>judiar de alguém ou animal (atormentar, fazer sofrer, maltratar)</b> | <i>maltrattare</i> Ø           |
| <b>lidar com algo (manipular, manobrar, operar)</b>                     | <i>manovrare, maneggiare</i> Ø |
| <b>mexer com algo (alterar, modificar, mudar)</b>                       | <i>modificare</i> Ø            |

**Fonte:** Xatara; Zavaglia; Silva (2013, p. 351).

Diante do exposto, é possível dizer que em Lexicografia Bilíngue lidamos com signos linguísticos providos de significação cuja relação de biunivocidade é um *desiderium* frequentemente não

<sup>1</sup> Pelo menos nas pesquisas realizadas no dicionário *online* da Treccani <<http://www.treccani.it/vocabolario/fottere/>>, no Zingarelli (2005) e na Internet.

alcançado. Entretanto, a tarefa de tradução é realizável, lançando-se mão de expedientes linguísticos dos mais variados tipos, visto que, sem ela, não existiriam os dicionários bilíngues, tão importantes e cada vez mais necessários no atual contexto globalizado em que vivemos.

Noções como “equivalência exata”, “significado real”, “correspondência gramatical, idiomática e um para um”, “equivalente literal” devem ser, senão esquecidas, postas de lado e/ou utilizadas com cautela pelo lexicógrafo tradutor, visto não corresponderem à realidade no processo tradutório entre duas línguas, ou melhor, dois sistemas linguísticos diferentes.

E, para concluir, em que diferem ou se assemelham os verbetes dos quatro dicionários mencionados anteriormente com os desse último dicionário?

Antes de mais nada, trata-se de um conjunto de obras lexicográficas completamente diferente entre si, mas que se torna analisável a partir do momento que os quatro primeiros dicionários são obras disponíveis no mercado, para não dizer as únicas, e o último é um dicionário do qual participei ativamente da elaboração.

As semelhanças existem, e não são poucas. A primeira delas e a única sobre a qual vou falar é que, mesmo que estejamos nos esforçando por uma Lexicografia Bilíngue mais consciente e abrangente, ainda estamos produzindo verbetes para dicionários bilíngues sem o devido cuidado com a inserção e elaboração de definições, além, muitas vezes, da ausência total de exemplos autênticos na língua estrangeira, como é o caso de todos os dicionários aqui citados.

O ideal seria o planejamento, a compilação e a elaboração de dicionários bilíngues, no caso, para o par de línguas italiano e português do Brasil, a partir de grandes *córpus* linguísticos, máxime se paralelos, para que a busca de correspondentes tradutórios e contextos definicionais pudesse ser pautada em dados autênticos e em uso para ambas as línguas.

Além disso, seria desejável que o sistema de remissivas, que mostrei no início da minha fala, deixasse de existir em dicionários bilíngues, além dos monolíngues, de uma forma tão contundente e frequente e que pudéssemos ler verbetes completos e que satis-

fizessem as nossas dúvidas e nossos anseios no momento de nossa pesquisa.

Enquanto essa realidade não chega, consciente da sua complexidade, continuo a acreditar que os lexicógrafos, quando optam por produzir obras bilíngues, sabem da falta de clareza e das lacunas que poderão existir em seus verbetes, mas, nem por isso, deixam de confiar nessa insana tarefa, tampouco negligenciá-la, e nem desistir de cumpri-la.

E faço a *mea-culpa* também, pois ainda estamos caminhando para essa Lexicografia Bilíngue consciente, planejada e estruturada, mas que não deixa de ser um desejo, acredito, de todos aqueles que trabalham na feitura de dicionários com mais de duas línguas no Brasil.

## REFERÊNCIAS

BENEDETTI, I. C. **Dicionário Martins Fontes italiano-português**. São Paulo: Martins Fontes, 2004.

DICIONÁRIO ESCOLAR WMF: italiano-português, português-italiano. Tradução de Ivone Castilho Benedetti e Letizia Zini. São Paulo: Martins Fontes, 2013.

HOUAISS, A. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009.

MEA, G. **Dicionário de italiano-português**. Porto: Porto Editora, 2000.

MENDES, R. **Aspectos comparativos da regência verbal em português e inglês para a elaboração de um dicionário bilíngue**. 134f. 2013. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2013.

MILTON, J. **Tradução: teoria e prática**. 3. ed. São Paulo: Martins Fontes, 2010.

POLITO, A. G. **Dicionário eEscolar Italiano**. 2003. Disponível em: <<http://michaelis.uol.com.br/escolar-italiano/creditos/>>. Acesso em: 03 mar. 2018.

RODRIGUES, C. C. **Tradução e diferença**. São Paulo: EDUNESP, 2000.

SCHMITZ, J. R. A problemática dos dicionários bilíngues. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (Org.). **As ciências do léxico**: lexicologia, lexicografia e terminologia. Campo Grande: Ed. da UFMS, 1998. p.159-168.

XATARA, C.; ZAVAGLIA, C; SILVA, R. M. da. **Dicionário Multilíngue de Regência Verbal**: verbos preposicionados. Barueri: DISAL, 2013.

ZINGARELLI, N. **Lo Zingarelli**: Vocabolario della lingua italiana. Bologna: Zanichelli, 2005.

# DICIONÁRIOS QUE FIZEMOS: ALGUMAS REFLEXÕES SOBRE O DICIONÁRIO HISTÓRICO DO PORTUGUÊS DO BRASIL – SÉCULOS XVI, XVII E XVIII (DHPB)

Clotilde de Almeida Azevedo MURAKAWA

## Introdução

Este texto é o resultado de apresentação feita durante o I Congresso Internacional de Estudos do Léxico e suas Interfaces (CINELI), em mesa-redonda cujo tema foi **Dicionários que queremos... Dicionários que fizemos: reflexões sobre Lexicografia**. Pontua-se, a seguir, alguns procedimentos adotados para a redação de verbetes no DHPB.

A 1ª etapa a ser estabelecida para a elaboração do DHPB foi a construção do banco de dados, hoje ponto essencial e de honra para a elaboração de toda e qualquer obra lexicográfica. Cabe à Linguística de Corpus o mérito de tudo isso. Com o DHPB não poderia ser diferente.

O banco de dados, com aproximadamente 7 milhões e meio de ocorrências, serviu de base para a elaboração da nomenclatura e a partir dela a redação dos verbetes. Este banco reúne documentos/textos dos séculos XVI, XVII, XVIII e começo do XIX até o ano de 1808, que falam sobre o Brasil no período.

De posse do aparato teórico lexicográfico, buscou-se organizar um modelo de verbete que correspondesse aos objetivos do Dicionário Histórico, ou seja, que informações linguísticas e históricas deveriam fazer parte do verbete. À medida que a consulta ia sendo feita ao banco de textos começaram a surgir as dúvidas: tal informação deve constar ou não do verbete? Que procedimento dar a determinados casos tão bem representados no banco? A partir de alguns questionamentos, soluções tiveram que ser tomadas. Algumas dessas soluções vão ser relatadas a seguir, tendo sempre como ponto de referência a informação registrada no banco de dados.

## Variante

Com esta etiqueta **variante**, foram reunidas todas as variantes gráficas, morfológicas e fonéticas que o banco apresentou. Assim, a palavra-entrada veio acompanhada da classe gramatical e em seguida, as variantes. Estas foram obtidas através de um procedimento de busca, denominado **busca por similaridade** que o programa Philologic permite e também foram criadas outras (a partir do conhecimento que se tem da História da Língua Portuguesa) numa tentativa de encontrar mais variantes, além daquelas que, num 1º momento, foram localizadas no banco de dados:

Search for:

|       |                                       |                                      |
|-------|---------------------------------------|--------------------------------------|
| razão | <input type="button" value="Search"/> | <input type="button" value="Clear"/> |
|-------|---------------------------------------|--------------------------------------|

Display:  Context  KWIC  Similarity Search

Search Context:

Word or Phrase  Phrase separated by  words

Proximity Search in:  Sentence  Paragraph

Bibliographic Search Fields:

Title:

(e.g. 'DA EXSICCAÇÃO DA CANELLA')

Author:

(e.g. 'MARTINHO DE MENDONÇA DE PINA E DE PROENÇA')

Date:

(e.g. '1627')

**Found 15 matches**, shown with frequencies in entire database.

Select words to search in the entire database. Select output options and bibliographic criteria below.

or

- 
- 2  arazão
  - 22  brazão
  - 3  cazão
  - 1  erazão
  - 1  frazão
  - 1  iazão
  - 3  ralão
  - 63**  **raszão**
  - 1  razoão
  - 1  razáo
  - 1  razão
  - 2278**  **razão**
  - 517**  **rezão**
  - 4  sazão
  - 7  vazão

Como se pode observar na relação acima não aparecem outras variantes como **razom**, ou ainda **razam** (26 ocorrências.) ou **rasam** (06 ocorrências.) e **rezam** (88 ocorrências). Foram feitas, então, as tentativas e estas variantes foram encontradas e mais a variante **rrezão** (03 ocorrências.). Já **razão** teve 2278 ocorrências no banco. As demais variantes registradas na busca por similaridade com ocorrência 1, foram consideradas, de acordo com o contexto onde se encontram, como erro de impressão ou de correção do texto.

Surgiu, então, a dúvida: qual o valor e importância para o estudo da língua registrar apenas as variantes encontradas? E os contextos onde estão inseridas? Diante de tal importância a solução foi registrar de todas as variantes o contexto com todas as informações bibliográficas necessárias. Com este procedimento, ficaram recuperados as variantes, os contextos, o tipo de texto, o autor, a época e a localização da unidade lexical no conjunto da obra. Veja-se abaixo um exemplo de algumas das variantes **razão**, **rrezão** e **rezam**:<sup>1</sup>

Em sua maneira de viver não era fóra da lei natural e da **razão**, o que em mui poucos gentios tenho visto n'esta terra. Ficou um seu Irmão por principal, o qual tem por nome Simão, e o morto D. João, como qual mettemos cá em vergonha os maus christãos, porque é mui virtuoso e fóra dos costumes dos outros, PADRE ANTONIO PIRES. (1865) [1551], CARTA QUE O PADRE ANTONIO PIRES ESCREVEU DO BRASIL, DA CAPITANIA DE PERNAMBUCO, AOS IRMÃOS DA COMPANHIA, DE 2 DE AGOSTO DE 1551. [A00\_0721 p. 97].

[...] o que se faria bem se a justiça secular e eclesiastica fosse mais zelosa, como convem à honrra de N. Senhor e bem comum da terra, e desta maneira podião hir cada día ganhando gente e sogeitando-a ao jugo da **rrezão**. E os que não quizesem recebê-lo, sujeitá-los e fazê-los tributarios ao serviço

---

<sup>1</sup> Os exemplos deste texto foram extraídos do *Dicionário Histórico do Português do Brasil – séculos XVI, XVII e XVIII*. A informação bibliográfica traz o autor, data da edição utilizada entre parênteses, data ou século em que o autor escreveu a obra entre colchetes e em negrito, título da obra ou capítulo e numeração do arquivo e página conforme o banco de dados do DHPB.

d'El-Rei e dos christãos, que os ajudassem a senhoriar, como se fes em todas as terras novas que são conquistadas, [...]. P. MANUEL DA NÓBREGA (1956) [1557], CARTA DO P. MANUEL DA NÓBREGA AO P. MIGUEL DE TORRES, [BAÍA (RIO VERMELHO) AGOSTO] 1557. [A00\_0026 p. 401].

Aqui estivemos tomando água e lenha e corregendo as naos, que dos temporaes que nos días passados nos deram vinham desaparelhadas. Nesta baía achámos um homem português que 20 havia 22 anos que estava nesta terra e deu **rezam** larga do que nela havia. Os principaes homens da terra vieram fazer obediência ao capitam I. e nos trouxeram muito mantimento e fizeram grandes festas e bailos, amostrando muito praz PÊRO LOPES DE SOUSA (1968) [1530], DIÁRIO DA NAVEGAÇÃO DE PÊRO LOPES DE SOUSA [A00\_0078 p. 47].

Seguindo este procedimento, todos os verbetes foram contemplados com suas variantes e acompanhados do referido contexto.

## **Homonímia**

Um 2º ponto que merece ser destacado com relação a certos desvios no procedimento geral estabelecido foi com relação à homonímia. Que critério adotar para a inserção em entradas separadas de unidades lexicais homógrafas e homófonas?

A prática lexicográfica tem adotado o critério etimológico como o mais usual e que tem sido seguido pela maior parte dos lexicógrafos. Entretanto, mais uma vez o banco de dados mostrou-se valioso neste sentido. A consulta à nomenclatura deixava claro que os três critérios propostos pela Lexicografia – critério etimológico, gramatical ou funcional e semântico – poderiam estar contemplados nos verbetes do DHPB, oferecendo ao consulente não só a diferença de étimo, mas também o aspecto gramatical ou funcional da palavra-entrada, assim como a mudança semântica que, proveniente de uma polissemia, acabou se apresentando como uma homonímia. Seria difícil desconsiderar a mudança de função de

um substantivo para adjetivo ou deste para aquele, tão bem empregado em obras do Pe. Vieira e de outros tantos oradores sacros do período. Foram, então, adotados os três critérios e as entradas foram separadas com a numeração acima <sup>1,2,3</sup>.

Para a seleção dos homônimos e das acepções, utilizou-se o procedimento de **busca KWIC Report** que permitiu a visualização das unidades e uma pequena parte do contexto. Abaixo um exemplo da variante **rezão**: conforme aparece no Philologic:

Your search found **517** occurrences

More search results (batches of 100)

1 2 3 4 5 6

[Retrieve all occurrences](#) (This may take some time to download)

[Click here for a Concordance Report](#)

**Occurrences 1-100:**

1. **A00\_1589** ([bib:p.0](#))nom hia tão mall que com a muita **rezão** que llevava me non atrevesse e Deus ajudar
2. **A00\_1590** ([bib:p.0](#))e gerall em tudo o que me parece **rezão** por minha consciencia dizello e V. A. gasta
3. **A00\_0007** ([bib:p.0](#))o mais do coração, o qual não hé **rezão**, que seja senhor delle senão ho mesmo que
4. **A00\_0010** ([bib:p.0](#))ior cosmografo, terá V. A. muita **rezão**, que eu não sey nada disto, senão deseya
5. **A00\_1591** ([bib:p.0](#)) se lloguo prover nisto quem com **rezão** o deve fazer porque doutra maneira estão
6. **A00\_1591** ([bib:p.0](#))pior cosmografo terá V. A. muita **rezão** que eu não sey nada disto se não deseyar
7. **A00\_0015** ([bib:p.0](#))stamdo tão longe de V. A. parece **rezão** que quamdo hum homem for avexado sem causa
8. **A00\_0015** ([bib:p.0](#))s e achara que nom tynha nenhuma **rezão** pera mamdar pasar a dita carta d'escomun
9. **A00\_1592** ([bib:p.0](#))nao nos Petiguares sem gemte e a **rezão** era por a gemte da tera ter fomes e os pro
10. **A00\_0013** ([bib:p.0](#))trou mais o amor que nos tinha, **rezão** hee que lho paguemos ao menos algum

Seguem alguns exemplos de verbetes de acordo com os três critérios:

### **Critério etimológico**

Neste critério, o que leva a entradas separadas é o étimo. Assim em **escalar**<sup>1</sup> e **escalar**<sup>2</sup> são os étimos que “autorizam” a inserção em entradas separadas. Nestes casos, a 1ª datação é colocada nas 2 entradas.

#### **escalar**<sup>1</sup> v.

variante: escallar.

#### **Intransitivo**

1. Abrir e limpar um peixe com uso de faca, para retirar suas espinhas, preparando-o para o consumo.

[...] e he tãta a quantidade q' mataõ q' vi por vezes perto de duas mil pessoas cõ ca noas ao matar e fazerẽ mta deligêçia pa o salgar como era leuarẽ, o sal pizado as facas pera **escalar** mtas e afiadas, mta lenha jũta pa o asar e cõtudo mais era o q' se perdia do q' se aproueitaua [...]. FRANCISCO SOARES (1966) [1591], COISAS NOTÁVEIS DO BRASIL – MANUSCRITO DE COIMBRA [A00\_0065 p. 195].

#### **Transitivo direto**

2. Acutilar; golpear com faca; agredir; ferir.

[...] e vão os doentes parar nas mãos das velhas comedeyras, que não tem outro officio; pois a huns os **escallaõ** com esfregaçoens pelos braços com tal força, que se lhe levantaõ caroços por elles, e entãõ he que dizem, que por ter caroços, tem a espinhella cahida, e que tenha paciencia para o fazer ir ao outro mundo com taes dores, que lhe impedem a respiraçãõ [...]. LUIS GOMES FERREIRA (1735) [1735], DA MISCELLANIA DE VARIOS REMEDIOS, AÍM EXPERIMENTADOS, E INVENTADOS PELO AUTOR, COMO ESCOLHIDOS DE VARIOS PARA DIVERÍAS ENFERMIDADES [B00\_0031 p. 204].

**1ª. datação [1591]**

Aqui esteue Do flores Baldes 6. ou 7 mezes e nũa se sêtio falta cõ estarẽ algũs 4000 soldados q' hiaõ pera o estreito tẽ peixe de aRibaçaõ q' saõ como grãdes tainhas de pezo de 4 arratẽs e he tâto q' o cercaõ no Ryo e toda a terra se poẽ a **escalar** e salgar de dia e de nojte e asar e mtas vezes [...]. FRANCISCO SOARES (1966) [1591], COISAS NOTÁVEIS DO BRASIL - MANUSCRITO DECOIMBRA [A00\_0065 p. 49].

**escalar<sup>2</sup> v.**

variante: escallar.

**Intransitivo**

1. Assaltar, saquear; transpondo muros e obstáculos.

[...] (tomou) leuando consigo hũ Atambor e hũ Sold.o com [húa] band.ra branca na mão, e se foi m.to seguro ao engenho onde determinaua fazer emenda de todo o comettido atalhando com suaprezença a resoluçam com q' estauam os q' **escalar** a força [e queriam lançar fogo]. desconhecido (1935) [séc. XVII], RELAÇÃO DO DR. ANTÔNIO DA SILVA E SOUSA SOBRE A REBELIÃO DE PERNAMBUCO 1645 – AO LEITOR [A00\_2116 p. 105].

**Transitivo direto**

2. Subir em lugar elevado ou íngreme; montar; trepar.

[...] sabendo fervirfê da valentia do feu animo, vendo-fê entre dous perigos, escolheo o mayor, **efcalando** a trincheira, que fê lhe oppunha diante dos olhos com hum fatal efrago da multidaõ de barbaros, que a guarnecia; e afugentados todos os mais, que lhe disputavaõ a campanha, ficou fenhor della. BERNARDO PEREIRA DE BERREDO (1749) [1718], ANNAES HISTORICOS DO ESTADO DO MARANHAÕ – LIVRO XIII [A00\_2525 p. 434].

3. Designar pessoas para determinado serviço, trabalho ou tarefa.

[...] E a Scena portentoza a cada instante Se muda e se converte; esta diante Huma extensaõ larguissima de montes, Que cortaõ varios rios, lagos e fontes; Densos mattos a cobrem; vem-se as serras De escabrozos roxedos novas guerras Tentar, buscando os Céos, como tentava Briareo, quando aos Deozes **escallara**. CLÁUDIO MANUEL DA COSTA (1773) [1773], CANTO 5º [G00\_0007 p. 22].

4. Percorrer.

[...] tanto em prejuizo do Serviço de S. Mag.de e de sua real faz.da e bem comú daquelles Povos por se não lavrarem os descubrim.tos do ouro no Ribeí.o do peixe, **escaland**-se este sem se repartir, nem tirar a datta de Sua Mag.de a q.e se seguirão varios tumultos com Armas, desobedecendo-se ás ordres do Guarda mor [...]. ANTONIO DE ALBUQUERQUE COELHO DE CARVALHO (1897) [1711], XIII – DIVERSAS CARTAS – PATENTES, ORDENS, BANDOS, ETC, DO GOVERNADOR ANTONIO DE ALBUQUERQUE COELHO DE CARVALHO [A00\_0774 p. 783].

**Transitivo indireto**

5. Fazer parada ou escala em.

Aos quaes navios de efcravos, ferà de grandifsima validade **efcallar** no Maranhão, pellas muitas mais cômodidades que alj tẽ, que em nenhũa outra parte. A primeira, he ficarem dalj mais nauegados em Indias, & haverem de chegar là cõ as peças,que aqui refrescarem muito inteiras, & vendaveis, o que não tem nos outros portos do Brazil; porque para focorro eftão muito cedo, luftre das peças ficão longe. CAP. SYMÃO ESTACIO DA SYLVEIRA (1624) [1624], RELAÇÃO SV MARIA. AS COVSAS DO MARANHÃO [B00\_0021 p. 12]. (1ª. datação).

Os verbos **escalar**<sup>1</sup> e **escalar**<sup>2</sup> têm étimos diferentes. Segundo Cunha (1996) **escalar**<sup>1</sup> vem de calar, significando “abaixar, abater”; já **escalar**<sup>2</sup> vem de “escada”, do baixo latim *scalata*, derivado de *scala-ae*.

**ata**<sup>1</sup> s.f.

variante: acta.

Documento em que se registram decisões ou determinações realizadas em reuniões ou sessões públicas.

Feyta a eleyção de Guardiães, Prefidêtes, Confellôres, & mais officiaes, se fará taboa, & se mandará ler pelos Conventos do mesmo modo, que se fez no Capitulo; & adverte-se q̃ na Congregação não se podem fazer Constituições, & **Actas** que

tenham forã de Iey, mas podem mandar alguns apontamentos, ã fe obfervem para mayor perfeção da Religiaõ, os quaes acabam com o Diffinitorio, que os infituhio. desconhecido (1709) [1709], BREVE DA CONFIRMAÇAM DESTES ESTATVTOS [A00\_2499 p. 92].

**1ª. datação [1648]**

Pesah (H.) – Páscoa dos judeus, de 15 a 22 do Nisan. Pinkes (H.) — Livro de **atas** da comunidade judaica. DR. ARNOLD WIZNITZER (1952) [1648], 3 - O LIVRO DE ATAS DA CONGREGAÇÕES JUDAICAS “ZUR ISRAEL” EM RECIFE E “MAGEN ABRAHAM” EM MAURÍCIA, BRASIL, 1648-1653 - TRANSCRIÇÃO DO MANUSCRITO ORIGINAL, INTRODUÇÃO, NOTAS E GLOSSÁRIO, PELO DR. ARNOLD WIZNITZER [A00\_1235 p. 239].

**ata<sup>2</sup> s.f.**

1. Fruta doce, de origem asiática, semelhante à pinha mas com casca menos endurecida, formada por várias sementes, em formato de pinhões, cujas superfícies são recobertas por uma camada macia e branca.

Ha pouco tempo, que pela capitania do Mato-Grosso se introduziram no Rio-Negro o jambo e o tamarindo. O figo, a laranja, a lima, o limão doce, a abobora, a melancia, o melão, o pepino, a beringella, e o tomate foram transplantados da Europa. A taqueira, ou gerumun de machado, e o calombro foram introduzidos pelos ilhéos. O côco, e a **ata** vieram da Asia. ALEXANDRE RODRIGUES FERREIRA [séc. XVIII], 2.<sup>a</sup> PARTE: BAIXO RIO NEGRO - PARTICIPAÇÃO SÉTIMA: PARTICIPAÇÃO GERAL DO RIO NEGRO [A00\_2243 p. 701].

2. Árvore pequena e pouco frondosa que produz essa fruta.

**Ata** he huma arvore pequena mas engraçada he o seu fructo semelhante a huma pinha, cheia de huma massa alvissima e branda, cheia porem de sementes pretas como feijoens: o gosto pode semelhar-se a requeijão fresco com açúcar. LUIZ DOS SANTOS VILHENA (1921) [1801], CARTA VIGESIMA [A00\_0846 p. 745].

**1ª. datação [1757]**

Ainda é mais deliciosa a **ata**, que poderia ser a rainha das fructas, se não as deslustrassem as suas muitas sementes. PE. JOÃO DANIEL (1976) [1757], PARTE TERCEIRA - DÁ NOTICIA DA SUA MUITA RIQUEZA NAS SUAS MINAS NOS SEUS MUITOS, E PRECIOSOS HAVERES, E NA MUITA FERTILIDADE DAS SUAS MARGENS - TRATADO PRIMEIRO - DAS MINAS DE OURO E PRATA, E DIAMANTES DA REGIÃO AMAZÔNICA - CAP. 6º - PROSEGUE-SE A MESMA MATÉRIA [A00\_1858 p. 327].

Segundo Cunha (1996) **ata**<sup>1</sup>, vem do latim *acta-orum* significando “coisas feitas”, ou seja, relato escrito onde se registra o que se passou numa reunião, congresso, enquanto que **ata**<sup>2</sup>, “fruta”, tem origem controvertida. Já Machado (1967) dá como origem a palavra antilhana *ahate*, acompanhada da interrogação que indica dúvida quanto à origem.

### **Critério gramatical ou funcional**

Os verbetes seguintes exemplificam a mudança gramatical que substantivos e adjetivos podem sofrer em diferentes contextos; o étimo é o mesmo, entretanto, sofrem uma conversão, ou seja, mudam de classe gramatical.

**manuscrito**<sup>1</sup> *s.m.*

variantes: manuscrito, manuscripto.

1. Obra escrita à mão.

O **manuscrito** de Dionysio da Costa diz que a entrada foi pela Bertioça: isto mesmo dicta a boa razão, e contesta a Fortaleza que Martim Affonso mandou levantar n'aquelle porto, [...]. FREI GASPAR DA MADRE DE DEUS (1920) [1767], MEMORIAS PARA A HISTORIA DA CAPITANIA DE S. VICENTE HOJE CHAMADA DE S. PAULO [A00\_0169 p. 123].

2. Original, escrito à mão de um texto impresso ou de outro manuscrito.

[...] a erigio o famoso Diogo Luis de Olivr.<sup>a</sup> 14.<sup>o</sup> G.<sup>or</sup> deste Estado, como consta de hum **manuscrito** antigo, e Verificaç<sup>o</sup> as Provizo<sup>ens</sup> q.<sup>c</sup> se citao<sup>o</sup> nas açço<sup>ens</sup> do sobred.<sup>o</sup> Diogo Luis de Olivr.<sup>a</sup>. JOZÉ DE MIRALES (1900) [séc. XVIII], I - HISTORIA MILITAR DO BRASIL [B00\_0010 p. 72].

**1<sup>a</sup>. datação [1607]**

Vae copiado este **manuscrito**, talvez do original, na linguaagem, ou, melhor, orthographia actual, com excepção de algumas palavras que, por caracterisarem o seculo, deixei ficar sem transposição para as que hoje usamos, como: giolho por joelho, antre por entre, devaç<sup>o</sup> por devoç<sup>o</sup>, etc, etc. PADRE PEDRO RODRIGUES (1897) [1607], I. - VIDA DO PADRE JOSÉ DE ANCHIETA, PELO PADRE PEDRO RODRIGUES [A00\_1316 p. 47].

**manuscrito<sup>2</sup> adj.**

Que está escrito a mão.

De huma relação **manuscrita** do Governador Artur de Sá e Menezes colhemos tudo o que a este repeito se plica ao heor, e talvez estes trabalhos taó generosamente sofridos, daó hum character da grandeza do espirito, [...]. CLÁUDIO MANUEL DA COSTA (1773) [1773], CANTO 1<sup>o</sup> [G00\_0003 p. 12].

**marchante<sup>1</sup> adj.**

Que transporta, carrega mercadorias.

Naõ cõsidero de menos inportância aprestaremse como disemos os Navios **Marchantes**; porque fasê assy outra Armada muito mayor que a da Cõpanhia: [...]. VÁRIOS AUTORES (1961) [1654], ECONOMIA - I PARTE - PARECER DE FRANCISCO DE BRITO FREIRE, DIRIGIDO A D. JOÃO IV, SOBRE OS MEIOS DA CONSERVAÇÃO DO BRASIL, COM APRECIACÕES ACERCA DO VALOR DAS MINAS DE S. PAULO E DO CARÁCTER DOS MORADORES DESSA CAPITANIA, A QUE CHAMA A ROCHELLA DO SUL / 1654 - SETEMBRO - 2 [A00\_2383 p. 53].

**1<sup>a</sup>. datação [1587]**

Logo no anno seguinte de 1550 se ordenou outra armada, com gente e mantimentos, em soccorro d'esta nova cidade,

da qual foi por capitão Simão da Gama de Andrade com o galeão velho muito afamado e outros navios **marchantes**, em a qual foi o bispo D. Pedro Fernandes Sardinha, pessoa de muita autoridade, grande exemplo e estremado pregador, [...]. GABRIEL SOARES DE SOUSA (1938) [1587], MEMORIAL E DECLARAÇÃO DAS GRANDEZAS DA BAHIA DE TODOS OS SANTOS, DE SUA FERTILIDADE E DAS NOTAVEIS PARTES QUE TEM (PARTE SEGUNDA -TITULO I) [A00\_0177 p. 129].

**marchante**<sup>2</sup> *s.m.*

Indivíduo que negocia gado para açougues.

E que o **marchante** | do asougue da misericórdia | a que se tem consedido lisença | per hum anno ou enquanto ella | durar sera obrigado se quiser | continuar en cortar g[ado] no | dito talho não podera comprar | senão aos criadores que tiuerem | gado uindo [...]. RUI DE CARUALHO PINHEIRO (1949) [1661], TERMO DE REZULUSÃO QUE | SE FES COM OS CRIADORES [SO]BRE | A FORMA DA DESTREBUISAM DOS | TALHOS [A00\_1454 p. 77].

### **Critério semântico**

Nos verbetes a seguir, as entradas **assunto** e **assombrado** são casos de polissemia, ou seja, **assunto**<sup>1</sup> **assunto**<sup>2</sup> têm o mesmo étimo e semanticamente se distanciaram, o mesmo acontecendo com **assombrado**<sup>1</sup> e **assombrado**<sup>2</sup>.

**assunto**<sup>1</sup> *s.m.*

variantes: assumpto, asumpto.

Aquilo sobre o que se conversa, fala ou escreve; tema; matéria. [...] e ha Escritores cuja mizeria os precisa lançar mão de qualquer ficção para comprovar seos **asumptos**; [...]. FREI DOMINGOS DE LORETO COUTO (1904) [1757], LIVRO QUARTO – PERNAMBUCO ILLUSTRADO COM VIRTUDES / CAP. VIII – DE OUTROS SERVOS DO SENHOR, QUE AUTHORISARÃO A PÁTRIA COM VIRTUDES PRECLARAS. N. 46 [A00\_0662 p. 255].

**1ª. datação [1584]**

Êste colégio foi o segundo da Companhia do Brasil, e como a cidade da Bahia teve grandes aumentos nos engenhos de açúcar e fazendas e muito trato de portugueses, e como é o **assunto** dos Governadores e Bispos, assim êle também cresceu muito [...]. JOSÉ DE ANCHIETA (1964) [1584], INFORMAÇÃO DO BRASIL E DE SUAS CAPITANIAS [A00\_0079 p. 42].

**assunto**<sup>2</sup> *adj.*

Que se elevou; elevado.

O Padre João Felipe Beltendorf, em hum dos seus escriptos [...] afirma, que indo ao Reino, por Procurador de toda a missão em 1685, alcançara do Serenissimo Senhor D. Pedro II, novamente **assumpto** ao throno, huma concessão absoluta da aldeia do Maracú [...]. PADRE JOSÉ DE MORAES (1860) [1759], LIVRO V, DE OUTRAS ACCÇÕES DOS NOSSOS MISSIONARIOS NO ESTADO DO MARANHÃO E DAS DO GRANDE PADRE ANTONIO VIEIRA ATÉ A SUA PARTIDA PARA O PARÁ. CAPITULO IV, CONTINUAÇÃO DA MESMA MATERIA (v. CAPÍTULO III) [A00\_0290 p. 417].

**assombrado**<sup>1</sup> *adj.*

variantes: asombrado, assôbrado, asóbrado.

Cheio de medo, admiração, receio.

Amanheceo o dia de 15. de Janeiro, em que se festeja S. Amaro, mui sombrio, &. o inimigo mais **assóbrado** por nos ver alojados tam perto, & com tanta obra feita em tam poucas horas, & medindo com ella o numero dos nossos soldados, acrecêtou em os seus grãde temor. FRANCISCO BARRETO (1899) [1653], RELAC,AM | DIARIA | DO SÍTIO, E TOMADA | DA FORTE PRAÇA DO RECIFE, RECUPERA- | ÇÃO DAS CAPITANIAS DE ITAMARACÀ, PA- | RAIBA, RIO GRANDE, CIARÁ, & ILHA DE | FERNAÓ DE NORONHA, POR FRANCISCO | BARRETO MESTRE DE CAMPO GENE- | RAL DO ESTADO DO BRASIL, & | GOVERNADOR DE PERNAMBUCO [A00\_1128 p. 189].

**1ª. datação [1559]**

[...] foy arreatado do spiritu maligno, segundo que todos julgamos pelos sinaes que fez naquela ora e tres dias continuos;

e elle mesmo como **asombrado** das visões que via, brama-va e não queria estar senão com os olhos tapados, [...]. P. MANUEL DA NÓBREGA (1956) [1559], CARTA DO P. MANUEL DA NÓBREGA AO P. MIGUEL DE TORRES E PADRES E IRMÃOS DE PORTUGAL, BAÍA 5 DE JULHO 1559 [A00\_0033 p. 64].

**assombrado**<sup>2</sup> *adj.*

1. Que tem sombra; que faz sombra.

Tinha de largura n'este tempo e logar, onde o passámos, oito braços, com 4 para 5 palmos de fundo: suas aguas são crys-tallinas, mas porque corre **assombrado** de verdes matas, e como profundamente entalhado entre empinados montes, [...]. JOSÉ VIEIRA COUTO (1842) [1801], ITINERARIO DO RIO DE S. FRANCISCO ATÉ AO RIO ABAITÉ [A00\_2357 p. 54].

2. Que exterioriza alegria ou satisfação interior.

[...] mas sempre o imigo se atrauesa p<sup>a</sup> lhe estrouar sua salu-ação saõ os brasis bem despostos das mesmas feicons estatu-ra dos portugezes mas baços as molheres bem **asóbradas** tem os cabelos como os portugezes [...]. FRANCISCO SOARES (1966) [1590], COISAS NOTÁVEIS DO BRASIL - MANUSCRITO DE MADRID [A00\_0064 p. 7].

## **Datação**

Um 3º e último aspecto a ser considerado diz respeito à datação, informação importante que permite retrodatar todas as entradas que compuseram o DHPB. Há que se considerar que tal datação diz respeito ao conjunto de todas as obras que entraram no banco de dados no período de mais de 3 séculos, ou seja, do ano de 1500 ao ano de 1808 e não à 1ª vez que apareceu na língua portuguesa. O documento mais antigo do conjunto que registrar a palavra-entrada quer como variante, quer flexionada, foi considerado o mais antigo. No caso de não se ter a data do documento foi registrado o século em que a obra foi escrita; e existindo um único significado registrou-se a 1ª datação ao final do contexto. Vejam-se alguns exemplos a seguir:

**nau** *s.f.*

variantes: náu, nao, náo.

**1ª. datação [1500]**

E o capitã mandou a dous degradados e ado dijz que fosse la a aldeia e a outras se ouuesem delas nouas e q̃ ã toda maneira nõ se viesem adormjr aas **naos** ajnda que os eles mandasem e asy se fora. PERO VAZ DE CAMINHA (1964) [1500], CARTA DE PERO VAZ DE CAMINHA [A00\_0335 p. 9].

**navalha** *s.f.*

variante: nabalha.

**1ª. datação [1591]**

[...] q' chupãõ o sangue muj sutil m<sup>te</sup> alguãs vezes inda q' poucas e so vi hũ homẽ que tinha o cabello trosquia-do delles q' parecia a **nabalha** e nõqua lhe mais naceo nõ diz q' naçe mais tem grãdez vnhas e come carne Coruos, Vrubus. FRANCISCO SOARES (1966) [1591], COISAS NOTÁVEIS DO BRASIL - MANUSCRITO DECOIMBRA [A00\_0065 p. 143].

**vigária** *s.f.*

variante: vigaria.

Cargo que têm as mulheres, nas ordens terceiras; faz as vezes de superiora.

[...] mandou vir do Mosteiro do Desterro da Bahia quatro com os empregos de Abbadessa, **Vigaria**, Mestra de noviças, e Porteira, as quaes assistirão alguns mezes no Hospicio de Jerusalem, e depois com doze noviças fôrão dar principio a vida regular no convento novo em 1750. FREI GASPAR DA MADRE DE DEUS (1920) [1767], MEMORIAS PARA A HISTORIA DA CAPITANIA DE S. VICENTE HOJE CHAMADA DE S. PAULO [A00\_0170 p. 381].

**(1ª. datação)**

**abieiro** *s.m.*

Árvore frondosa, de madeira compacta e bela, cujo principal valor está na produção de seu fruto, o abio.

É também digna das quintas, pomares, e jardins a árvore **abieiro**, e na verdade é das mais cultivadas no Estado amazonico, não só pela bizzaria do seu fructo, senão também pela pomposa gala das suas folhas. É árvore mediana, mas muito copada, e por isso muito sombria. PE. JOÃO DANIEL

(1976) [1757], PARTE TERCEIRA - DÁ NOTICIA DA SUA MUITA RIQUEZA NAS SUAS MINAS NOS SEUS MUITOS, E PRECIOSOS HAVERES, E NA MUITA FERTILIDADE DAS SUAS MARGENS - TRATADO PRIMEIRO - DAS MINAS DE OURO E PRATA, E DIAMANTES DA REGIÃO AMAZÔNICA - CAP. 6º - PROSEGUE-SE A MESMA MATÉRIA [A00\_1858 p. 327].  
(1ª. datação)

## **Conclusão**

Todas as soluções encontradas e outras mais que poderiam ser mencionadas neste texto, diante do material linguístico que o banco de dados ofereceu, sempre tiveram como orientação o tipo de dicionário que estava sendo construído, o público a que ele foi destinado, e principalmente o período de abrangência do dicionário. A partir destes focos todas as soluções foram tomadas, tendo como referencial o que estava documentado no banco de dados.

## **REFERÊNCIAS**

- CUNHA, A. G. da. **Dicionário Etimológico da Língua Portuguesa**. 2. ed. Rio de Janeiro: Nova Fronteira, 1996.
- MACHADO, J. P. **Dicionário Etimológico da Língua Portuguesa**. 2. ed. Lisboa: Editorial Confluência, 1967.



# LÉXICO HISTÓRICO DO PORTUGUÊS BRASILEIRO: CAMINHOS PERCORRIDOS, METAS A ATINGIR

**Vanderci de Andrade AGUILERA**  
**Fabiane Cristina ALTINO**

## **Introdução**

O *Léxico Histórico do Português Brasileiro*<sup>1</sup> (LHisPB) é um Projeto do Departamento de Letras Vernáculas e Clássicas e do Programa de Pós-Graduação em Estudos da Linguagem, com sede na Universidade Estadual de Londrina, sob a nossa coordenação. Vinculado ao Projeto nacional e interinstitucional *Para a História do Português Brasileiro* (PHPB), tem como propósito elaborar um léxico a partir de manuscritos oriundos das diversas vilas e províncias do Brasil Colônia e Brasil Império, isto é, de documentos datados entre o início do século XVIII e o final do século XIX.

Nossa proposta é analisar os manuscritos selecionados e editados pelas equipes regionais de pesquisadores do PHPB<sup>2</sup>, principalmente do Rio de Janeiro, Bahia, Paraná, Ceará, Paraíba, Pernambuco, Rio Grande do Norte e São Paulo.

---

<sup>1</sup> Este Projeto foi aprovado pelo CNPq e conta com auxílio financeiro e bolsa produtividade.

<sup>2</sup> São as equipes que, até o presente momento, nos enviaram os manuscritos editados.

Ao publicar o LHisPB, esperamos: (i) oferecer, sob a forma de banco de dados informatizado, um vasto material lexicográfico para estudos sincrônicos e diacrônicos do léxico registrado em manuscritos dos três séculos mencionados; (ii) proporcionar um instrumento de consulta rápida do uso de vocábulos vigentes na época e em várias regiões do Brasil Colônia e do Brasil Império; (iii) apresentar, com a frequência de uso, os itens lexicais e suas variantes gráficas registrados em cerca de 2000 documentos emanados das então províncias e vilas sediadas nos atuais estados brasileiros.

### **Fontes e condições do material que compõe o *corpus* do PHPB**

O acervo dos manuscritos do Rio de Janeiro está disponível no *site*<sup>3</sup> e compõe-se de 141 cartas de natureza pessoal e oficial, datadas dos séculos XVIII e XIX, a primeira das quais de 1720 e a última de 1896. Do *corpus* desse projeto, selecionamos: os documentos da Administração Pública, documentos da Administração Privada e documentos particulares enviados a Rui Barbosa<sup>4</sup>. O acervo lexical dos manuscritos do RJ conta com mais de 4.000 formas, das quais já estão lexicalizadas, sob a forma de verbetes, da letra A até a M. Não dispomos dos fac-símiles uma vez que os documentos estão disponíveis no modelo semidiplomático aos quais demos o tratamento de limpeza (*nettoyage*), retirando as caixas de identificação e as informações externas ao conteúdo do documento para submetê-los ao programa Léxico 3.

---

<sup>3</sup> Disponível em: <<https://sites.google.com/site/corporaphpb/home/corpora-manuscritos/tabela-manuscritos-rio-de-janeiro>>. Acesso em: 22 fev. 2018.

<sup>4</sup> “Ruy Barbosa de Oliveira (Salvador, 5 de novembro de 1849 – Petrópolis, 1 de março de 1923) foi um polímata brasileiro, tendo se destacado principalmente como jurista, político, diplomata, escritor, filólogo, tradutor e orador. Um dos intelectuais mais brilhantes do seu tempo, foi um dos organizadores da República e coautor da constituição da Primeira República juntamente com Prudente de Moraes. Ruy Barbosa atuou na defesa do federalismo, do abolicionismo e na promoção dos direitos e garantias individuais. Destacou-se, também, como jornalista e advogado. Foi deputado, senador, ministro. Em duas ocasiões, foi candidato à Presidência da República. Notável orador e estudioso da língua portuguesa, foi membro fundador da Academia Brasileira de Letras, sendo presidente entre 1908 e 1919”.

Disponível em: <[http://pt.wikipedia.org/wiki/Ruy\\_Barbosa](http://pt.wikipedia.org/wiki/Ruy_Barbosa)>. Acesso em: 15 fev. 2018.

Da Bahia, para integrar o LHisPB, buscamos os documentos oficiais, inventários, testamentos e declarações, que constam do projeto *Contribuição para a constituição de um banco de textos e de um banco de dados para o estudo da história do português no Brasil, do século XVII ao XIX*, datados entre os anos de 1698 e 1828, editados por Zenaide de Oliveira Novais e Norma Lúcia Fernandes de Almeida Carneiro e apresentados em *CD-ROM*.

O *corpus* do Paraná consta de 730 manuscritos oficiais, obtidos junto ao Arquivo Público do Estado de São Paulo, lavrados entre os séculos XVIII e XIX, mais especificamente de novembro de 1721 a dezembro de 1853, no âmbito das antigas vilas, como as de Paranaguá, Curitiba, Guaratuba, Antonina, Morretes, Iapó (atualmente Castro), Vila Nova do Príncipe (hoje Lapa), São José, entre outras. Os fac-símiles estarão disponíveis no *site* do Projeto do Léxico Histórico do Paraná (LhisPar). O *corpus* consta de cerca de 130.600 palavras. Parte do *corpus* paranaense foi editado conforme consta de Almeida-Baronas (2006), Aguilera e Almeida-Baronas (2007), Aguilera e Vasconcelos (2007) e Almeida-Baronas e Almeida (2009).

Para constituir o *corpus* do Ceará, recorreremos à tese de Expedito Ximenes (2009), que traz 45 *Autos de Querrela*, datados entre 1802 e 1829 e escritos nos termos da vila de Fortaleza. Os *Autos* contêm cerca de 79.000 palavras.

O *corpus* da Paraíba, com aproximadamente 21.000 palavras, compõe-se de 203 cartas oficiais dos séculos XVIII e XIX, que foram editadas e organizadas por Fonseca (2004). A primeira carta data de abril de 1774 e a penúltima<sup>5</sup>, de agosto de 1874.

Sobre o *corpus* de Pernambuco, totalizando cerca de 17.500 palavras, temos 21 cartas oficiais do Século XVIII, datadas de 24 de Julho de 1704 a 23 de abril de 1784 e, do século XIX, 30 cartas oficiais, datadas de 22 de outubro de 1801 a 5 de julho de 1884.

Para compor o *corpus* de São Paulo, buscamos: (i) Edição semidiplomática de *Memória Histórica da Capitania de São Paulo e todos os seus Memoráveis Sucessos desde o anno de 1531 thê*

---

<sup>5</sup> Na carta nº 203 não consta a data.

o presente de 1796, de Manuel Cardoso de Abreu; (ii) sete Autos depositados no Arquivo do Tribunal de Justiça do Estado de São Paulo e (iii) 26 cartas pessoais, do Museu Paulista, referentes às *Correspondências particulares do Fundo Rafael Tobias de Aguiar*, editadas por Alexandra de Souza Marques, sob a supervisão de Marcelo Módolo, da Universidade de São Paulo. Sobre o *Memória Histórica*, Costa (2007, p.1), esclarece:

*Memória Histórica da Capitania de São Paulo e todos os seus Memoráveis Sucessos desde o anno de 1531 até o presente de 1796* é um livro manuscrito pertencente ao Arquivo do Estado de São Paulo, cota E11571, de autoria de Manuel Cardoso de Abreu, um sertanista paulista, nascido na freguesia de Araraitaguaba, atual Porto Feliz, que exercia o cargo de oficial maior da secretaria da Capitania de São Paulo. Datado de 1796, fim do século XVIII, o manuscrito, composto de 163 fólhos escritos em frente e verso, com exceção da folha de rosto e do fólho final, traz um levantamento da história da Capitania de São Paulo, antes chamada Capitania de São Vicente, desde sua fundação até o ano de 1796.

Ao todo temos um *corpus* paulista com cerca de 27.000 palavras.

Resumimos no quadro 1, a seguir, uma descrição dos *corpora* constituídos para a elaboração do Léxico Histórico do Português Brasileiro<sup>6</sup> (LHisPB).

---

<sup>6</sup> Para esta etapa do Projeto estamos trabalhando com os *corpora* de sete estados. Não incluímos o *corpus* de Minas Gerais pelo fato de a equipe regional já estar elaborando o Léxico Histórico Mineiro.

**Quadro 1** – Resumo descritivo dos *corpora* de acordo com a regional de origem

| Regional          | Natureza dos documentos | nº docs. | nº palavras | datas     | Estágio projeto        |
|-------------------|-------------------------|----------|-------------|-----------|------------------------|
| RJ                | pessoal e oficial       | 141      | 50.000      | 1720-1896 | revisão                |
| BA                | oficial                 | 28       | 17.000      | 1698-1828 | revisão                |
| PR                | oficial                 | 730      | 130.600     | 1721-1853 | revisão                |
| CE                | oficial                 | 45       | 79.000      | 1802-1829 | não iniciado           |
| PB                | oficial                 | 203      | 21.000      | 1774-1874 | revisão                |
| PE                | oficial                 | 51       | 17.500      | 1704-1874 | Elaboração de verbetes |
| SP                | pessoal e oficial       | 196      | 27.000      | 1796-1893 | revisão                |
| total de palavras |                         | 1394     | 296.000     |           |                        |

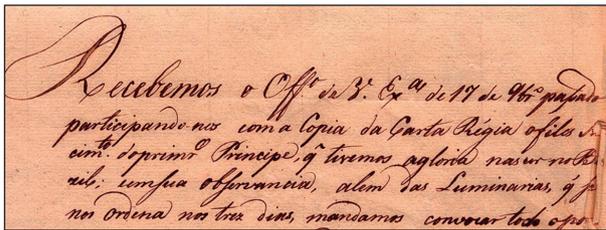
**Fonte:** Elaboração própria.

Os *corpora* de todos os estados, exceto o *corpus* do Paraná, foram editados pelas equipes regionais. Desse modo, tratamos apenas das condições de legibilidade dos manuscritos paranaenses. A propósito, cerca de 90% deles ofereceu um grau mediano de dificuldade, isto é, letra legível ou parcialmente legível; 8% se encontram em bom estado, mas de difícil leitura e 2% apresentam partes rasgadas, dobradas, corroídas por insetos ou danificadas pelo tempo, ou escritas na frente e no verso de papel fino e transparente, com tinta espessa, dificultando muitíssimo a leitura. Ilustramos a seguir com alguns excertos de documentos:

O excerto da Figura 1 foi destacado do documento 456, lavrado na câmara de vereadores da vila de Antonina em 31 de dezembro de 1811 e transcrito na sequência em edição semidiplomática. Trata do comunicado do nascimento do *primeiro príncipe nascido no Brasil*<sup>7</sup>.

<sup>7</sup> Não pudemos identificar a que “primeiro príncipe nascido no Brasil” a carta se refere, uma vez que a nona filha de D. João VI com D<sup>a</sup>. Carlota Joaquina, Ana de Jesus Maria de Bragança, nasceu em 1806. D. Pedro I, na época, tinha apenas 13 anos de idade.

**Figura 1** – Início da carta oficial correspondente ao fôlio 1 nº 456 do LHisPar.



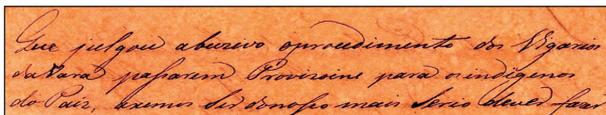
Recebemos o Offício de Vossa Excelência de 17 de novembro passado participando-nos com a Cópia da Carta Régia ofiles [nas] cimento doprimeiro Principe, que tivemos agloria nascer no B[ra]zil; eemsua observancia, alem das Luminarias, que [se] nos ordena nos trez dias, mandamos convocar todo opo[vo]

**Fonte:** Acervo depositado no Banco de dados do LHisPar-UEL.

Embora a letra seja bem legível, a dificuldade na leitura do presente fôlio diz respeito à dobra do papel no lado direito que leva o leitor a fazer pressuposições de acordo com o assunto.

O excerto da Figura 2, a seguir, faz parte do documento 147, lavrado na câmara de vereadores da vila de *Coritiba* em 12 de dezembro de 1821 e transcrito na sequência em edição semidiplomática. Trata-se de uma denúncia contra o pagamento de custas, de várias naturezas, impostas pelos vigários da paróquia local.

**Figura 2** – Extraído da Carta oficial correspondente ao fôlio 1 nº 147 do LHisPar.



Que julgou abuzivo o procedimento dos Vigários da Vara passarem Provisions para os indigenas do Pais, axemos Ser donosso mais Serio dever fazer

**Fonte:** Acervo depositado no Banco de dados do LHisPar-UEL.

A letra é razoavelmente legível, no entanto, a dificuldade na leitura do presente fólio diz respeito à indistinção no traçado de algumas letras, como [u] de *abuzivo* e a sílaba [ce] de *procedimento*; as letras [a], [e] e [o] nas palavras: *procedimento*, *passarem*, *axamos/laxemos(?)*, bem como a grafiação do /s/ e do /z/, em *abuzivo*, *Provisoins*, *fazer*.

O excerto da Figura 3, na sequência, faz parte do documento 533, lavrado na câmara de vereadores da vila de Paranaguá, no dia 23 de junho de 1810 e transcrito na sequência em edição semidiplomática.

**Figura 3** – Extraído da Carta oficial correspondente ao fólio 1 nº 533 do LHisPar.



[des]ta Villa freguesia do Pilar, creandose a Villa Antonina Separando [a]tte as Pessoas que eraõ proprietarios naquele destrito sempre no le[mite] deste tempo se fizeraõ Cameras sem percizaõ dos Meleccianos.

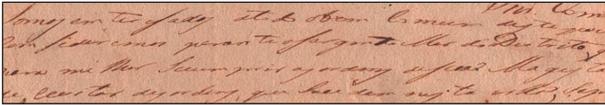
**Fonte:** Acervo depositado no Banco de dados do LHisPar-Uel.

Os fólhos 533 e 534 ofereceram um alto grau de dificuldade devido à escrita com tinta espessa, na frente (recto) e no verso de folha de papel fino, além de trechos corroídos por insetos. Trata-se de carta oficial enviada pela Câmara de Vereadores de Paranaguá, datada de 23 de junho de 1810, pela qual denunciam ao capitão-general governador da capitania de São Paulo, Antonio José da Franca e Horta, constrangimentos praticados por Manoel Álvares aos milicianos daquela corporação.

O excerto da Figura 4, a seguir, faz parte do documento 38, lavrado na câmara de vereadores da vila de Curitiba, no dia 15 de outubro de 1797 e transcrito na sequência em edição semidiplomática. Trata-se de carta oficial enviada pela Câmara de Vereadores de Curitiba ao Governador da Província de São Paulo, Antonio

Manuel de Castro e Mendonça<sup>8</sup>. A carta informa ao Governador da Província a possibilidade de remoção de um Guarda para patrulhar a saída de produtos da Vila até o rio Mogiguaçu.

**Figura 4** – Extraído da Carta oficial correspondente ao fólio 38 do LHisPar.



Somos emteressados a todo o bem comum deste povo consideramos perante o Sargento Mor [que] para melhor se cumprir as ordens de Sua Magestade e evitar desordens que sucedem

**Fonte:** Acervo depositado no Banco de dados do LHisPar-UEL.

O fólio 38 é um exemplo de manuscrito que oferece dificuldades de leitura devido à letra inclinada, palavras coladas (*atodo*) e outras descoladas (*mi lhor*) e indistinção de algumas letras, como em *eevitar*, que, além de a conjunção estar colada ao verbo, as três letras iniciais “*e e v*” podem ter outra leitura como “*c e u*”, por exemplo.

As abreviaturas, de início, ofereceram grande dificuldade, mas, à medida que nos familiarizávamos com os documentos, e a partir de consultas à obra de Flexor (2008), fomos fazendo associações com casos já detectados e desfazendo as ambiguidades, por exemplo: *V<sup>a</sup>* tanto poderia remeter a *Vossa* como a *Villa* e, neste caso, só o contexto elucidava a dúvida.

---

<sup>8</sup> “António Manuel de Melo e Castro de Mendonça foi um nobre e administrador colonial português. Tomou posse do governo de São Paulo em 28 de Junho de 1797, tendo sido o décimo quarto capitão-general da capitania. O seu mandato estendeu-se até 10 de Novembro de 1802. Posteriormente, foi enviado pela Coroa, a partir da Corte no Rio de Janeiro, como 32º capitão-general, ou Governador-geral de Moçambique, entre 14 de Agosto de 1809 e Agosto de 1812, tendo também sucedido a seu pai (interinamente?) no cargo de capitão-general dos Açores”. Disponível em: <[https://pt.wikipedia.org/wiki/Ant%C3%B3nio\\_Manuel\\_de\\_Melo\\_e\\_Castro\\_de\\_Mendon%C3%A7a](https://pt.wikipedia.org/wiki/Ant%C3%B3nio_Manuel_de_Melo_e_Castro_de_Mendon%C3%A7a)>. Acesso em: 16 fev. 2018.

## **Caminhos percorridos**

O Quadro 1 mostra sinteticamente o estágio atual do trabalho rumo à elaboração do LHisPB: (i) leitura e seleção de 1394 manuscritos obtidos das sete equipes regionais (Rio de Janeiro, Bahia, Paraíba, Paraná, Pernambuco, Ceará e São Paulo) ; (ii) conclusão da limpeza (*nettoyage*) de 1102 deles, isto é, foram retiradas palavras e expressões que não faziam parte dos documentos, como anotações de data, cidade, e outras informações acrescentadas *a posteriori*; (iii) junção e descolamento de palavras para adequação à unidade gráfica atual sem alterar a grafia da época (presença/ausência de diacríticos, letras dobradas, oscilação gráfica das vogais e consoantes...) e retirada de sinais de pontuação; (iv) classificação das palavras homônimas e homógrafas, tais como *a* artigo, pronome demonstrativo, pronome oblíquo, preposição, contração, 3ª pessoa do presente do indicativo do verbo haver quando grafado sem o h; (v) organização de cada *corpus* em arquivo separado, salvo em caixa alta e submetido à ferramenta Léxico 3; (vi) elaboração, concluída ou em andamento, de verbetes com os itens lexicais extraídos dos manuscritos do Rio de Janeiro, Bahia, Paraíba, Ceará, Pernambuco, São Paulo e do Paraná.

O Léxico Histórico do Paraná está, praticamente, concluído – foram elaborados 4130 verbetes – em fase de revisão para ser colocado no *site* do Projeto. Dentre os seis *corpora* restantes, os que são mais longos, isto é, Rio de Janeiro e São Paulo, com 50.000 e 79.000 palavras respectivamente, demandarão mais tempo para a sua execução, uma vez que os quatro outros, Bahia, Paraíba, Ceará e Pernambuco, apresentam um *corpus* mais reduzido com um número de palavras entre 27.000 e 17.000 e não exigirão muito tempo para a elaboração dos verbetes.

Na sequência, apresentamos alguns verbetes do LHisPar referentes a palavras lexicais (substantivos, adjetivos, verbos e advérbios) e a palavras gramaticais (preposições e conjunções).

Selecionamos para este trabalho o verbeito correspondente ao substantivo *manhã*:

1. **manhá**. Recebe esta parada pellas des horas da *manham* sem defeyto algum no dia abaixo declarado. (CAN, 1808, 523, 12) tem nesta Vila hua dearia guarda de 35 homens, para Com [ela] fornecer a da cadeia, com sinco as da terra e com seis [inint.] o contrato das passagens em que ficaõ no quartel com manejo de arma de *menhaã* e de tarde de peça. (PGA, 1810, 534<sup>9</sup>, 12); logo mandei mais soldados para aquella guarda por quem me trouce a parte Em mesmo dia de *minham* mandei o Cabo Comandante deste Destacamento co mais soldados a reconhecer o dito Navio. (PGA, 1803, 721, 06).

FORMAS DOCUMENTADAS: *manham* (1), *menhaã* (1), *minham* (1).

FREQUÊNCIA GLOBAL: 3

O verbete mostra a entrada em negrito, na forma ortográfica atual, seguido de ponto final. Apresentam-se as abonações, separadas por ponto e vírgula, que buscam documentar todas as variantes gráficas, em destaque na cor azul e em itálico, apresentadas no *corpus* e indicadas pelo Léxico 3. Ao fim de cada abonação e entre parênteses, temos a abreviatura da localidade de onde foi extraído o excerto, o ano, o nº do documento e a linha onde se encontra a palavra em questão. Na primeira abonação, por exemplo, o excerto foi extraído de documento emitido em Cananéia, no ano de 1808, no documento de nº 523 e na linha 12. As Formas Documentadas e a Frequência Global são extraídas da relação dada pelo Léxico 3. De cada variante será selecionada, pelo menos, uma abonação. No caso de o item lexical apresentar apenas uma variante e for bastante recorrente no *corpus*, procuramos documentar com vários excertos para mostrar os diversos contextos de uso em épocas diferentes.

---

<sup>9</sup> Os documentos 533 e 534, devido ao mau estado de conservação dos fólhos e à escrita na frente e no verso do papel transparente, foram, em princípio, descartados pela dificuldade de compor um texto completo e coerente. No entanto a frase acima foi considerada importante para ilustrar uma variante gráfica e fonética de *manhá* = *menhá* = forma frequente no Português popular falado por pessoas mais idosas, geralmente rurais, conforme ilustram os trabalhos na área da Dialetologia e da Geolinguística (AMARAL, 1922; AGUILERA, 1994).

Selecionamos o adjetivo *salutífero*, para ilustrar o tratamento dado aos itens dessa classe gramatical, isto é, independentemente do gênero e do número com que apareça nos documentos, a entrada do adjetivo, quando biforme, será feita pela forma masculina singular.

**2. salutífero.** como então, tão saudaveis dispuziçoens, reimentadas a sua cautella pelos referidos Provimmentos e Carta Regia, emquanto senão determinaõ mais *salutíferos* remedios pelo Codigo regenerativo da Naçaõ, cujos iscolhidos Repreentantes simentaõ o mais solido Edeficio da nossa felecidade (SL, SD, 151, 19).

FORMAS DOCUMENTADAS: *salutíferos* (1).

FREQUÊNCIA GLOBAL: 1

As abreviaturas que estão no final do verbete e entre parênteses significam que, neste caso, não constava do documento a localidade de origem e a data da redação do manuscrito. Os números 151 e 19, como já explicitado, indicam o número do documento no acervo e a linha em que se encontra o item lexical analisado.

Escolhemos o verbo **vedar** para ilustrar a elaboração de verbetes dessa classe gramatical. São extraídas do *corpus* todas as formas flexionadas do verbo para constar das formas documentadas. Nos casos de adjetivos e particípios passados, quando se registra a dificuldade de distinguir uma classe da outra, o item é colocado no verbete como verbo:

**3. vedar.** sendo na verdade certo que as madeiras proprias que para este fim se achão *vedadas* na capitania do Rio de Janeiro são somente o Tapinhoá, e Peroba (ANT, 1798, 401, 05); o dito Governador manda escrever, que na sua jurisdição não cabe proibir aos meus vasalos a liberdade de navegar, e comerciar para aquelles portos que não são *vedados* por mim (PGA, 1803, 707, 34); pôs em execução os repetidos provimmentos que *vedaõ* a entrada de Tropas na ditta Ronda (CAS,

1822, 338, 03); tanto assim que querendo o Suplicante **vedar** hum Caminho exquizado que os mesmos tropeiros tem feito, Logo que o fexa, os mesmos abrem (CTB, 1797, 40, 19).

FORMAS DOCUMENTADAS: *vedadas* (1), *vedados* (2), *vedaão* (1), *vedar* (2).

FREQUÊNCIA GLOBAL: 6

A entrada dos advérbios em –mente é feita de acordo com a grafia atual, isto é, compondo uma única palavra, independentemente de que estejam os adjetivos colados ou não ao sufixo **mente**. Ilustramos com o advérbio:

4. **direitamente**. e haverá o Ordenado se o tiver e todos osproes, e precalços, que **direitamente** lhe pertencerem (SPA, 1796, 395, 16); que Se lhe defferise lhe haver Ministro, que **Direyta mente** provese no caso (CTB, 1764, 19, 22)

FORMAS DOCUMENTADAS: *direitamente* (2), *direyta mente* (1)

FREQUÊNCIA GLOBAL: 3.

Dentre as palavras gramaticais selecionamos a preposição **de** e a conjunção **porquanto** para ilustrar a composição do verbete de cada uma dessas classes gramaticais.

5. **de**. Senhor General Conde **de** Sarzedas (CTB, 1733, 1, 9); que por serem limitados os rendimentos desta Camara, e naõ chegarem para as dispezas **de** tantos Caminhos e pontes **de** que se preciza (CTB, 1764, 4, 7); tenho feito arrematar em Praça sallariados muitos orfaons **de** ambos os sexos (S/D, S/D, 81, 5); aVera tres mezes que puzemos na presenca **de** Vossa Excelencia o mais Lamentavel estado a que podia chegar o miseravel povo do destrito (,1791, 154, 2); cuja inteligencia com mais aferro preocupou as minhas ideias na presenca das ordens **de** Vossa Excelencia e Senhorias de 18 de novembro **de**

1814 (CTB, 1818, 390, 3); hoje não ha neste distrito hum lavrador **di** profiçãõ (CAN, 1798, 170, 20).

FORMAS DOCUMENTADAS: *de* (7534), *di* (2).

FREQUÊNCIA GLOBAL: 7536

O formato dos verbetes das palavras gramaticais segue o mesmo modelo adotado para as palavras lexicais. A preposição **de** é a mais frequente no *corpus* seguida da conjunção **e**, com 5495 registros e o **que** pronome relativo, com 3024. Da entrada da preposição **de**, não constam as combinações com os artigos (**do, da, dos, das**) e pronomes demonstrativos (**deste, disso, disto, daquele, daquilo**) que vão compor outros verbetes.

**6. porquanto.** não podera Seruir ditos offiços Sem pRouiZaõ do dito gouerno pasada todos os seis meses **por Coanto** tudo o que pRoSeSarem Sem as ditas pRouiZoins he nullo. (PGA, 1725, 612, 04); aos officiaes da Camara dizendo ser bem cómum; o que he pelo contrario, **por quanto** os supplicantes tem experimenttado notáveis prijuizos. (CTB, 1798, 73, 04); mencionada Vila de Parananguã, e seus moradores fica sendo inteiramente prejudicial a estes, e só de utilidade aos de Santos, **por quanto** primeiramente fica com esta prohibiçaõ. (SL, SD, 710, 12); não pode o mesmo recorrente evitar estes, e outros projuizos, **por quanto** suposto que este o fexe logo he aberto e tranquiado o mesmo passo pelos Tropeiros e andantes (CTB, 1798, 48, 10).

FORMAS DOCUMENTADAS: *por coanto* (1), *por quanto* (9).

FREQUÊNCIA GLOBAL: 10

## **Próximas etapas**

Concluída a revisão do LHisPar, as próximas etapas, que contemplam os Léxicos do Rio de Janeiro, Bahia, Paraíba, São Paulo, Pernambuco e Ceará, inspiradas em Company e Melis (2002),

seguem a mesma metodologia adotada para esse léxico, ou seja: (i) *netoyage* de cada fôlio, eliminando informações alheias ao conteúdo do documento; (ii) identificação das palavras homônimas, em cada um dos léxicos regionais, mediante numeração convencionalizada pela equipe; (iii) reunião de todos os fôlios em um só arquivo; (iv) gravação de arquivo em caixa alta e submissão à ferramenta Léxico 3; (v) elaboração dos verbetes, em ordem alfabética, a partir das formas elencadas pelo Léxico 3; (vi) revisão de todo o Léxico.

Uma vez concluídos os Léxicos de cada equipe regional, todos os *corpora* do banco de dados serão reunidos para compor um só o qual dará origem ao Léxico Histórico do Português Brasileiro, com aproximados 6 000 verbetes.

### Considerações finais

O trabalho até agora realizado demandou a preparação de uma equipe especializada composta por alunos de Iniciação Científica, de Especialização, de Mestrado e de Doutorado, bolsistas e não bolsistas, que se dedicaram durante anos auxiliando em todas as etapas do Projeto. Desta forma, além do LHisPar, LHisRJ, LHisBA, LHisPB, LHisPE, LHisSP e LHisCE, muitos trabalhos foram e serão desenvolvidos nos vários níveis acadêmicos a partir dos *corpora* disponíveis no banco de dados do Projeto.

### REFERÊNCIAS

AGUILERA, V.de A. **Atlas lingüístico do Paraná**: ALPR. Curitiba: Imprensa Oficial do Estado, 1994.

AGUILERA, V. de A.; ALMEIDA-BARONAS, J. E. (Org.). **Scripturae nas villas de São Luiz de Goaratuba e Antonina**: manuscritos setecentistas e oitocentistas. Londrina: EdUEL, 2007.

AGUILERA, V. de A.; VASCONCELOS, C. (Org.). **Scripturae na villa de Pernagoa**: manuscritos setecentistas. Londrina: EdUEL, 2007.

ALMEIDA-BARONAS, J. E. Manuscritos paranaenses: fontes para um estudo diacrônico. In: SELISIGNO, 5.; SIMPÓSIO DE LEITURA, 6., 2006, Londrina. **Anais...** Londrina, PR: UEL, 2006. p.1-8.

ALMEIDA-BARONAS, J. E.; ALMEIDA, P. L. C. de. A escrita dos séculos XVIII e XIX: em busca de dados diacrônicos. In: SEMINÁRIO DO CELLIP: Centro de Estudos Linguísticos e Literários do Paraná, 19., 2009, Cascavel. **Anais...** Cascavel, PR: UNIOESTE, 2009. p. 1-8.

AMARAL, A. **O dialeto caipira**. São Paulo: HUCITEC: Secretaria da Cultura, 1922.

COMPANY, C.; MELIS, C. **Léxico histórico del español de México**. México: Universidad Autónoma de México, 2002.

COSTA, R. F. **Edição semidiplomática de Memória Histórica da Capitania de São Paulo e todos os seus Memoráveis Sucessos desde o anno de 1531 até o presente de 1796, de Manuel Cardoso de Abreu**. 2007. Disponível em: <<http://principo.org/costa-renata-ferreira-2007-edico-semidiplomtica-de.html>>. Acesso em: 22 fev. 2018.

FLEXOR, M. H. O. **Abreviaturas**: manuscritos dos séculos XVI ao XIX. 3. ed. Rio de Janeiro: Arquivo Nacional, 2008.

FONSECA, M. C. de A. (Org.). **Cartas oficiais da Paraíba dos séculos XVIII e XIX**. João Pessoa: Ideia, 2004.

XIMENES, E. **Estudo Filológico e Linguístico das Unidades Fraseológicas da Linguagem Jurídico-criminal da Capitania do Ceará nos Séculos XVIII e XIX**. 2009. 413f. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Ciências e Letras, Universidade Federal do Ceará, Ceará, 2009.



# O DESAFIO DA RETRODATAÇÃO PARA OS ESTUDOS ETIMOLÓGICOS DE LÍNGUA PORTUGUESA

**Mário Eduardo VIARO<sup>1</sup>**

**Aldo Luiz BIZZOCCHI**

**Mariana Giacomini BOTTA**

**Marco Dimas GUBITOSO**

**Gustavo Luiz VIEIRA**

## **Introdução**

A discussão sobre a origem das palavras existe desde a Antiguidade e sempre foi pauta para inúmeras especulações, uma vez que todo falante observa e procura entender as mudanças que ocorrem em sua língua, como a substituição de uma palavra por outras, ou mesmo a alteração de um significado.

Mas a técnica etimológica de cunho científico desenvolveu-se lentamente. Na Península Ibérica, o primeiro passo foi dado por Nebrija, em 1492, com o detalhamento das alterações que o espanhol teria sofrido, a partir do latim, para adquirir a forma daquele momento. Em 1606, Duarte Nunes de Leão descreve uma certa

---

<sup>1</sup> Pesquisador CNPQ Pq 2.

regularidade dessas “corrupções”, transformações pelas quais teria passado a língua portuguesa.

Autores preocupados com as regularidades dos fenômenos fonéticos, que formam a base de toda etimologia científica, surgiram apenas em meados do século XVIII e, sobretudo, no século XIX. Podem ser citados o húngaro Gyarmathi, o dinamarquês Rask e os alemães Grimm e Bopp, que alicerçaram as bases das regras que futuramente seriam chamadas “leis fonéticas”. No final do século XIX, acreditava-se *grosso modo* que tudo que não era explicável pelas leis fonéticas encontrava sua explicação nas analogias.

Posteriormente às duas guerras mundiais, as questões etimológicas perderam força nos estudos acadêmicos. Apesar do grande avanço esperado, a Etimologia foi esquecida no interregno do surgimento da Linguística Moderna e deixou de ser uma disciplina dos cursos de Letras, tendo desaparecido da formação dos estudiosos e sendo abandonada em seguida, confundindo-se apenas com a pseudociência de mesmo nome, que nada mais é que um entretenimento fundamentado apenas em subjetivismos e intuições sem investigação histórica e sem rigor metodológico. Os estudos científicos foram retomados somente no final do século XX (VIARO, 2011).

Além disso, a mídia e a internet têm contribuído para a divulgação de soluções etimológicas fantasiosas, sem que os estudos acadêmicos na área de Linguística Histórica e Filologia consigam expor suas conclusões. É importante lembrar que hoje há ferramentas importantes, tanto teóricas quanto tecnológicas, para se desenvolverem pesquisas de qualidade no nível acadêmico, que podem contribuir para a mudança da visão da sociedade sobre o tema.

Os estudos sobre a língua portuguesa ressentem-se desses acidentes históricos, pois esta passou a ser estudada historicamente somente no final do século XIX e início do XX. Tudo o que se tem nesta área se deve ao esforço individual de poucos pesquisadores, como José Pedro Machado (1952-1977, 1981) em Portugal e Antônio Geraldo da Cunha (1982, 1989, 2003, 2006) no Brasil. A língua portuguesa atualmente está na triste condição de idioma neolatino menos conhecido do ponto de vista histórico.

Visando mudar essa situação, foi criado, em 2013, o Núcleo de Apoio à Pesquisa em Etimologia e História da Língua Portuguesa<sup>2</sup>, que pretende aliar todo o conhecimento atual da área às inovações tecnológicas da Linguística computacional e Lexicometria. Dentre os projetos desenvolvidos destaca-se o *Dicionário Etimológico da Língua Portuguesa* (DELPo), sob a direção de Mário Eduardo Viaro, que, num primeiro momento, propõe realizar a datação das primeiras ocorrências do maior número possível de palavras do português para, em seguida, serem elaboradas hipóteses etimológicas baseadas em dados fidedignos.

## **Os dicionários etimológicos da língua portuguesa**

As informações que se têm sobre a etimologia do Português são incompletas e fragmentadas. Os dicionários etimológicos apresentam muitos problemas, como informações vagas e imprecisas sobre a primeira ocorrência registrada das palavras na língua, o *terminus a quo*, o que impede um maior aprofundamento tanto nas diferentes sincronias pretéritas quanto nos inúmeros sistemas a que pertence o étimo.

Podem-se citar outras deficiências: a) a confusão entre derivação prefixal e sufixal com etimologia; b) a confusão entre o étimo da palavra e sua origem remota; c) a falta de cuidado com étimos de línguas ágrafas (ameríndias e africanas); d) o desconhecimento sobre a influência árabe (CORRIENTE, 2008, 2013); e) a abundância de étimos fantasiosos que descaracterizam o estudo etimológico como um trabalho científico. Além disso, ainda não há até hoje uma metodologia para o trabalho sobre o *terminus ad quem*, e a língua portuguesa está longe de ter hipóteses etimológicas e *termini a quo* para acepções, pois o que existe são majoritariamente datações de lemas.

Na melhor das hipóteses, algumas obras reúnem o conhecimento de vários etimólogos e confiam em suas supostas especializações. Cunha, por exemplo, com suas pesquisas de português medieval

---

<sup>2</sup> Disponível em: <[www.usp.br/nehilp](http://www.usp.br/nehilp)>. Acesso em: 21 fev. 2018.

(CUNHA, 2006) e de palavras de origem tupi (CUNHA, 1989), chegou ao estabelecimento de bons *termini a quo*. Já Machado (1952-1977) foi um dos únicos que se preocuparam com palavras de origem árabe e com palavras portuguesas ocultas em textos do latim medieval. E os dados levantados por ambos são repetidos pelos dicionários que se sucedem de forma assistemática.

Também há grandes lacunas na pesquisa de diversos séculos e de sistemas linguísticos a que se vinculam os étimos. Os citados trabalhos de Cunha se concentram nas etapas mais antigas do português, mas há poucos estudos sobre o século XX. Os séculos XVII e XVIII são extremamente produtivos, no entanto, ainda bastante desconhecidos.

No tocante a sincronias antigas, toda a questão referente à língua portuguesa se vincula e se alia à pesquisa atualmente desenvolvida não só para a língua galega, mas também para o asturo-leonês e para os demais idiomas ibero-românicos. Em alguns casos, para a argumentação etimológica e a defesa de proposições etimológicas que se preocupem com a origem remota de uma palavra, até mesmo toda a pesquisa em Filologia Românica, em Língua Latina e em Indo-Europeu é importante.

O NEHiLP pretende fornecer dados para uma compreensão mais abrangente do fenômeno diacrônico e para a reconstrução de sincronias pretéritas não só da língua portuguesa mas também de sistemas direta ou indiretamente vinculados ao léxico português (substratos germânico, árabe, tupi, quimbundo e outros).

### **Processamento automático de textos e retrodatação**

Como uma boa explicação etimológica depende de coleta exaustiva de dados associados a datações seguras, para sanar a deficiência etimológica dos estudos históricos/diacrônicos em língua portuguesa é preciso reunir especialistas nas línguas que participam dos étimos e pesquisadores capacitados para realizar a busca de dados confiáveis em manuscritos e impressos. Neste segundo ponto, os recursos atuais da informática podem agilizar a coleta de material.

Uma das tarefas do etimólogo é coletar contextos em determinadas obras e associá-los à data de publicação da mesma. Por meio dessas datações, determina-se o *terminus a quo* da palavra, e estabelecem-se etimologias. Este trabalho, entretanto, está sujeito a erros motivados pelo manuseio de um grande número de informações ou mesmo pelo cansaço e falta de atenção. Acredita-se que as possíveis falhas humanas podem ser evitadas com o uso de um programa computacional que faça a comparação automática de um determinado *terminus a quo*, conhecido até o momento, com a data de uma nova obra a ser analisada.

A partir dessa ideia, em 2012, deu-se início à criação de um programa que extraia todas as palavras de um texto e que as compare com informações de *termini a quo* reunidas num banco de dados, de modo que, a cada processamento de um novo texto, o programa alerte sobre a existência, para toda palavra, de data anterior ou igual à do banco de dados. No momento, criou-se o nome *Moedor de Textos* para esse programa, que, nas discussões sucessivas, tornou-se apenas *Moedor*.

Simultaneamente, desenvolveu-se um sofisticado banco de dados, que partiu das informações contidas na primeira edição do *Dictionarium ex Lusitanico in latinum sermonem*, de Jerônimo Cardoso de Lamego (1508-1569)<sup>3</sup>, de 1562. Portanto, o ano de 1562 era o *terminus a quo* das primeiras palavras do banco de dados do DELPo, base de comparação de todas as obras submetidas pelo programa *Moedor*, denominada TAQ-zero.

## **Funcionamento do programa *Moedor***

Idealizado por Mário Eduardo Viaro, Nilsa Areán-García e Zwinglio de Oliveira Guimarães-Filho em 2012; o programa foi desenvolvido em linguagem de programação PERL<sup>4</sup>, por Marco

---

<sup>3</sup> Obra disponível no acervo da Biblioteca Nacional Digital da Biblioteca Nacional de Lisboa, que pode ser acessada por meio do *link*: <<http://purl.pt/15192/3/#/0>>. Acesso em: 16 fev. 2018.

<sup>4</sup> Trata-se de uma linguagem especialmente versátil no processamento de cadeias (*strings*), manipulação de texto e *pattern matching* (reconhecimento de padrões) implementado

Dimas Gubitoso e Gustavo Luiz Vieira<sup>5</sup>. O programa *Moedor* foi concebido para realizar automaticamente a datação e retrodatação dos *termini a quo* de palavras do português, identificadas por meio da sequência de caracteres, ou seja, seu significante ou lema. Dessa forma, o usuário submete ao programa um texto datado e recebe como resultado do processamento uma lista com sugestões de lemas de palavras candidatas a retrodatação.

O uso do programa é aberto a qualquer interessado, por meio do *link*<sup>6</sup>, mas o acesso ao banco de dados do DELPo é restrito a pesquisadores previamente cadastrados no *site* do NEHiLP. Por isso, foi criado um sistema de segurança que permite acesso às funcionalidades do programa de acordo com sete níveis de usuário: anônimo, registrado, autorizado, pesquisador, moderador, administrador e *root*.

Para começar a usar o programa *Moedor*, qualquer usuário tem de preparar o texto que pretende processar: ele deve estar em formato TXT, com codificação UTF-8 (e não ANSI ou ISO). Além disso, toda parcela de texto que não seja apropriada para o processamento via *Moedor* (como trechos em línguas estrangeiras), deve ser colocada entre colchetes duplos [...]. Há também formas de anotação específicas para a localização das ocorrências (páginas, capítulos, etc.) nos textos processados, obrigatórias apenas na utilização por usuários treinados, que têm acesso à versão fechada do programa. Estes também devem cadastrar os dados bibliográficos completos da obra em questão antes de dar início a seu processamento.

Assim que o texto é processado, o programa oferece uma lista de sugestões de lemas com *terminus a quo* inferior (data mais antiga) que o cadastrado no banco de dados, pré-classificadas graças a diferentes etapas de análise computacional. A partir dessa lista,

---

através de expressões regulares, com flexibilidade e capacidade de fazer códigos funcionais.

<sup>5</sup> Todas as etapas do desenvolvimento do programa foram acompanhadas e contaram com a consultoria dos pesquisadores Mário Eduardo Viaro, Aldo Luiz Bizzocchi e Mariana Giacomini Botta.

<sup>6</sup> Disponível em: < [http://www.nehilp.org/~nehilp/programas/moedor\\_aberto.php](http://www.nehilp.org/~nehilp/programas/moedor_aberto.php)>. Acesso em: 20. fev. 2018.

os usuários que têm autorização para cadastrar tais informações podem acessar as fichas “Inserir nova acepção” ou “Inserir nova abreviatura”.

## **Etapas do processamento computacional**

A primeira tarefa realizada pelo programa *Moedor* é a segmentação do texto processado em “contextos”, isto é, sequências de caracteres, incluindo espaços, vírgulas, ponto e vírgulas e dois pontos, definidas pela ocorrência de ponto final ou ponto de exclamação ou ponto de interrogação. Em seguida, são feitas as análises de tudo o que foi previamente anotado no texto: todo trecho de texto que foi marcado com colchetes duplos [...] é ignorado na análise; o mesmo ocorre com espaços em excesso, anotados com o símbolo \_; palavras com segmentação indevida, por conta de sua antiguidade ou edição, e que foram previamente tratadas com o símbolo =, são processadas como uma unidade. Por fim, o programa reconhece e associa as anotações referentes à localização dos contextos na obra (páginas, capítulos, etc.), salvando tais informações no banco de dados. Além disso, tudo o que tiver sido marcado dentro de chaves {...} passa a ser identificado como uma ocorrência, independentemente da existência de sinais de pontuação (artifício usado para casos de palavras grafadas com letras e números, como “7mbro” usado, em algumas obras, para referir-se ao mês setembro).

Depois dessas primeiras análises, o programa salva os “contextos” já tratados no banco de dados do DELPo. Esses “contextos” são arquivados de maneira desordenada, de modo a não permitir a recuperação integral dos textos processados, resguardando os direitos dos colaboradores do núcleo de pesquisa sobre seus *corpora*. Nesta etapa, o programa analisa as sequências de caracteres individualmente e desconsidera símbolos como apóstrofes e hífens, que poderiam comprometer os resultados.

A etapa seguinte do *Moedor* visa a identificar nos textos abreviaturas, que são cadastradas diferentemente no banco de dados do DELPo e, portanto, devem ser pré-classificadas pelo programa. Essa análise consiste na verificação da existência de uma sequên-

cia de letras maiúsculas cruzada com dados de uma lista prévia de abreviaturas, elaborada por Viaro em 2013. Quando identificada, ela passa a ser interpretada pelo programa, apenas para fins de pré-classificação, como uma sequência regular de caracteres (em minúsculas).

A seguir, todas as sequências de caracteres do texto que não foram excluídas ou que não se encontram nos casos previstos de abreviaturas são submetidas ao teste *desclitizador*, que consiste em regras simples de transformação. Por exemplo, a sequência “amá-lo” é interpretada em sua forma canônica “amar”. A partir de então, ela passa a ser considerada uma “ocorrência”, ou seja, está habilitada a figurar na lista de sugestões de lemas, propostas como resultado do processo de moagem. Embora esse teste ainda careça de certos ajustes, para que não seja descartado o clítico ligado ao verbo, sua ação não compromete o funcionamento inicial do programa.

A análise seguinte dá conta da identificação de unidades formadas por meio de sufixos superlativos e procede sua lematização (a palavra “digníssimas”, por exemplo, passa a ser lida pelo programa como “digníssimo”, masculino plural). O mesmo é feito com palavras formadas por sufixos diminutivos (-zinho, -ãozinho) e aumentativos (-zão, -ãozona).

Na sequência, o *Moedor* realiza diversas verificações para cada “ocorrência” salva no banco de dados, com base em listas elaboradas por Viaro entre os anos de 2013 e 2014: em primeiro lugar, analisa se se trata de palavra composta com hífen, comparando-a com uma lista de flexões, e realiza sua desflexão, necessária para a identificação e lematização das unidades em casos de plural; além disso, todas as outras unidades identificadas no plural ou conjugadas são desflexionadas. Em segundo lugar, o programa identifica se a “ocorrência” em questão é um número romano e, na etapa posterior, verifica se se trata de um caso de antropônimo ou topônimo, comparando-o com as informações existentes no banco de dados.

O passo seguinte é a identificação de possíveis homônimas. As “ocorrências” que ainda não foram pré-classificadas são submetidas à comparação a uma lista de homônimas conhecidas do português

e têm sua data confrontada novamente com o que consta no banco de dados. Nesta etapa, é feita também a verificação de possíveis homonímias entre palavras não lematizadas, como “canto” (verbo “cantar”) e “canto” (substantivo).

Todo este processamento aqui descrito ocorre internamente durante a moagem de um texto qualquer. É de acordo com essas análises que as “ocorrências” são pré-classificadas pelo *Moedor*, o que determina a forma como elas aparecem na lista de resultados que é apresentada ao usuário. Essa lista mostra, à esquerda, as sugestões de lema, acompanhadas dos contextos em que as mesmas aparecem no texto analisado. A lista é dividida em três cores, segundo a pré-classificação realizada pelo programa:

a) Sugestões de lema na cor vermelha – são as que apresentam no texto analisado data mais antiga que o *terminus a quo* cadastrado no banco de dados, ou seja, por meio desta análise, elas possivelmente retrodatam;

b) Sugestões de lema na cor amarela – são as que têm uma data a ser registrada pela primeira vez, ou seja, elas figuram no banco de dados, mas ainda sem uma data de sua primeira ocorrência na língua;

c) Sugestões de lema na cor azul – são as inexistentes no banco de dados do DELPo, sujeitas a ter sua primeira ocorrência registrada.

**Figura 1** – Exemplo de resultado oferecido pelo programa *Moedor*

| Sugestão de lema | Possível retrodatação desta palavra  |
|------------------|--|
| <u>acabar</u>    | Contexto 1: Mons. Le Blanc voltou de Paris a Doncker que provido de grossas sommas de dinheyro para <b>acabar</b> a obra do Canal de Mardick, e segundo as cartas daquella Praça se trabalha nella com todo o virgor.<br>Contexto 2: annos, havendo fallecido em Luneville de huma violenta doença de bexigas, que lhe <b>acabou</b> a vida dentro de 4.   |
| <u>aceitar</u>   | Contexto 1: dobroens ao Duque de Ossuna para a despeza da sua jornada; porém os inglezes não querem <b>aceitar</b> estas livranças, de que se entende que a guerra com Inglaterra está muy vizinha.<br>Contexto 2: O Principe de Palestrina <b>aceitou</b> o caracter de Embayxador, que S. Mag. Catholica lhe havia conferido, para em seu nome render as graças a S. Santidade, por haver mandado hum Legado à Latere à Rainha sua esposa. |
| Sugestão de lema | Possível nome próprio  |
| <u>Amazonas</u>  | Contexto 1: horas, passando mostra ao Regimento Real, assistindo tambem neste acto a Senhora Duqueza de Berry, e muytas Damas em traje de <b>Amazonas</b> .  |
| <u>Amsterdam</u> | Contexto 1: Os avisos de <b>Amsterdam</b> dizem haverem entrado naquelle porto 24.   |
| <u>Pedro</u>     | Contexto 1: hum filho, a quem deu o nome de <b>Pedro</b> relativo ao de ambos seus avós os Marquezes de Marialva, e de Angeja; e no dia seguinte bautizou o Conde de Villa Verde seu irmão huma filha, a quem puzerao o nome de Maria, e he a sua primogenita.   |
| Sugestão de lema | Palavra não reconhecida  |
| <u>abbade</u>    | Contexto 1: O <b>Abbade</b> del maro voltou de Turim com mais amplas instruccões para tratar o ajuste das differenças, que ha entre esta Curia, e o Tribunal da Monarquia no Reyno de Sicilia.<br>Contexto 2: Esta resolução faz persuadir que o Marquez del Borgo, e o <b>Abbade</b> Mauro Ministros de Sicilia, e Saboya partiaõ desta Corte para Turim descontentes do mau successo das suas negociaçoens.                                |
| <u>abonação</u>  | Contexto 1: Assegura se q. o Conde de Konigseck offerceeo a S. Mag. da parte do Imperador seu amo de lhe assistir cõ aquelle numero de tropas de q. elle necessitar: os Ministros dos Estados Geraes fizeraõ a mesma asseveração, limitandose ao numero estipulado no seu tratado de <b>abonação</b> .   |
| <u>abuso</u>     | Contexto 1: Sua Mag. tem determinado de convocar de sua authoridade hum Concilio nacional para reformação de alguns <b>abusos</b> .  |

Fonte: Programa *Moedor*.

## Inserção de informações no banco de dados do DELPO

O modo prioritário de cadastramento de informações no banco de dados do núcleo de pesquisa se dá a partir do processo de moagem, descrito no item anterior. A partir da lista de resultados do *Moedor*, ao clicar sobre uma das palavras da coluna da esquerda, abre-se uma pequena janela onde são mostradas as acepções já cadastradas daquela sequência de caracteres e duas opções para o pesquisador: “Inserir nova acepção” ou “Inserir nova abreviatura”.

Ao escolher “Inserir nova aceção”, será aberta uma janela com a ficha “Nova aceção”, cujos dados inseridos serão registrados no banco de dados. O mesmo ocorre quando a opção for “Inserir nova abreviatura”. A ficha “Nova aceção” é composta pelos campos:

- 1) Metalema
- 2) Flexão
- 3) Variante
- 4) Classificação morfológica
- 5) Classificação semântica
- 6) Abonação
- 7) Obra
- 8) Data
- 9) Origem
- 10) Etimologia
- 11) Etimologia da flexão
- 12) Imagem
- 13) Fonte da imagem

Cada um dos campos corresponde a um atributo ou tabela do banco de dados e responde às necessidades de interrogação das informações para a elaboração do DELPo. A função dos campos da ficha “Nova abreviatura” é a mesma. Eles são listados abaixo:

- 1) Abreviatura
- 2) Interpretação
- 3) Abonação
- 4) Obra
- 5) Data

Além desta forma de inserção de dados via resultados do *Moedor*, foram desenvolvidas outras maneiras de cadastramento de informações: os sistemas de “Inserção manual” e de “Inserção via concordanciador”. Essas possibilidades surgiram para resolver um dos principais problemas decorrentes da moagem dos textos: a não identificação de aceções. Isso se deve ao fato de o programa reali-

zar análises por meio da comparação de cadeias de caracteres, sem levar em consideração características semânticas, que demandariam uma anotação minuciosa dos *corpora*, fugindo aos objetivos das pesquisas do DELPo.

Dessa forma, ao se processar um texto do século XVIII em que ocorre a sequência de letras “brigadeiro”, ela será identificada como a primeira ocorrência da palavra *brigadeiro*, no sentido de “oficial comandante de uma brigada” que, segundo o *Dicionário Houaiss* (HOUAISS; VILLAR, 2009), deriva de “brigada”, no português desde meados do século XVII. Entretanto, quando o pesquisador moer um texto do século XX, em que aparece novamente a sequência “brigadeiro”, mesmo que corresponda a uma acepção distinta, no caso “doce feito com leite condensado e chocolate, apresentado sob a forma de bolinhas cobertas de chocolate granulado”, esta não será identificada como uma possível retrodatação, pois o programa vai descartá-la na primeira comparação com as datas que constam no banco de dados. Para realizar a datação da primeira ocorrência dessa sequência de caracteres com esta acepção, o pesquisador precisará recorrer ao programa *Concordanciador*, que também foi desenvolvido por Gubitoso e Vieira, em 2014 e se encontra acessível apenas para usuários cadastrados<sup>7</sup>. A partir do *Concordanciador*, a inserção de informações no banco de dados é feita exatamente da mesma forma que via *Moedor*, por meio das fichas “Inserção de nova acepção” e “Inserção de nova abreviatura”.

Também estão previstos os casos em que o colaborador não tenha a intenção de processar um texto completo, mas queira apenas contribuir com dados recolhidos manualmente, durante suas investigações pessoais. Ele poderá fazer o cadastramento das informações no banco de dados do DELPo por meio do *Sistema Papavero*, uma forma de inserção manual de dados, que foi nomeado em homenagem ao Prof. Dr. Nelson Papavero, integrante do núcleo, pesquisador de grande renome nas áreas bioló-

---

<sup>7</sup> Disponível em: <<http://www.nehilp.org/-nehilp/programas/concordanciador.php>>. Acesso em: 20. fev. 2018.

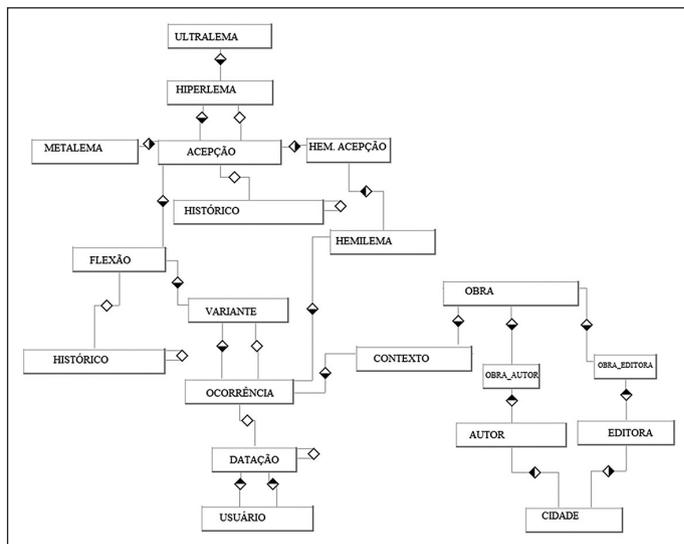
gicas e autor de muitas obras que são ricas fontes para o trabalho etimológico.

## **A estrutura do banco de dados do DELPo**

As informações armazenadas no banco de dados partem dos “contextos” e “ocorrências” obtidos pelo processamento de textos no programa *Moedor*, conforme descrito em 3.1, e também podem ser inseridas por meio do programa *Concordanciador* e do *Sistema Papavero*, citados na seção anterior. Elas são salvas de maneira estruturada, com o objetivo de facilitar sua recuperação com a ajuda de um sistema de busca.

Quando um texto é processado e nele são identificados “contextos” e “ocorrências”, também são salvas a sequência de caracteres, a flexão e a variante. Como já foi dito, o programa analisa apenas a sequência de caracteres (à qual se deu o nome de metalema, uma vez que não é possível diferenciar casos de homonímia dos de polissemia), criada com a ajuda do teste desflexionador e à qual se associam variantes ortográficas flexionadas, extraídas do texto que foi submetido. A atribuição de sentido (diferenciação das acepções, que devem ser datadas de maneira independente) é feita pelo pesquisador, por meio do campo “Classificação Semântica” da ficha “Nova acepção”. O conceito de metalema, inexistente até então na terminologia linguística, tem caráter puramente operacional no processo de moagem, visto que a decisão sobre a presença de homonímia e/ou de polissemia cabe ao pesquisador, com base em dados etimológicos e semânticos. O trabalho de vinculação hierárquica por meio da etimologia, que é estabelecido *a priori* na lexicografia tradicional (entradas diferentes para os casos de homonímia, entrada única para a polissemia), é uma tarefa *a posteriori* do núcleo de pesquisa. O esquema apresentado a seguir descreve a estrutura básica do banco de dados do DELPo:

**Figura 2** – Estrutura do banco de dados do DELPo



**Fonte:** Elaboração própria, com base no banco de dados do DELPo.

## Os casos das abreviaturas, alótipos e cognatos

No banco de dados do DELPo, esses elementos são organizados em estruturas independentes, com o objetivo de facilitar o trabalho do etimólogo no momento da elaboração de propostas, e também visa aumentar as possibilidades de pesquisa de informações pelos usuários.

As abreviaturas, nomeadas pelos pesquisadores do núcleo como hemilemas, são cadastradas independentemente das acepções, pois sua identificação pode, muitas vezes, mostrar-se problemática. Em textos antigos, as abreviaturas podem dar margem a muitas interpretações, e escolher uma delas para vinculá-la é bastante arriscado. Mesmo nos casos em que as abreviaturas são óbvias, convencionou-se tratá-las separadamente como um hemilema. A discussão etimológica fará menção às abreviaturas, no caso de retrodatações baseadas em interpretações das mesmas. A decifração de um hemilema pode conduzir a uma palavra flexionada, no entanto, no ban-

co de dados eles remetem às acepções e não às suas flexões: por exemplo, “mtos” ou “mta” é interpretado como uma das acepções do metalema “muito” e não como as flexões “muitos” ou “muita”, respectivamente. Da mesma forma, “afr” pode ser interpretado como “africano”, “africana”, “africanos” ou “africanas”.

No caso de alótipos, como “assobiar”/“assoviar”, não é possível falar de variação gráfica. A decisão de qual das duas formas é a principal e qual é a secundária deve ser estabelecida pela frequência de uso. No banco de dados do DELPo, apenas o alótipo principal terá o campo “Etimologia” completo por meio das fichas: os secundários farão uma remissão ao alótipo principal.

Por fim, para a discussão etimológica, também é importante reunir palavras simples e derivadas/compostas, bem como famílias de cognatos e constelações lexicais. Nesse caso, palavras como “amigo” e “amizade” ou “puro” e “apurar” podem ser ligadas umas às outras numa pesquisa, não por categorias morfológicas ou semânticas, nem estritamente pela forma atual. O modo de fazer isso será semelhante aos casos de alotopia: no banco de dados, serão associados por seus metalemas a uma tabela específica, a saber, a do ultralema.

## **Considerações finais**

A primeira etapa dos projetos do NEHiLP foi concluída em 2017 e hoje se encontra sob coordenação de Vanessa Martins do Monte. O projeto DELPo continua ainda no bojo desse núcleo e, em fevereiro de 2018, contava com 170 obras inseridas e mais de 11.000 metalemas, grande parte deles inserido por alunos de graduação bolsistas PUB/USP desde 2016 (Camila Santos Rocha, João Carlos Soares Mendes, Lucas de Sousa Gurgel, Pablo Sandino Ferreira Botelho, Gilvandro Mendes Monteiro, Rodrigo Moreira da Silva, Silmara Dias de Souza, Raphael Sponton Peres, Angélica Solânia de Abreu e Wesley de Souza Sobrinho). O trabalho computacional também se desenvolveu com a ajuda de alunos de iniciação científica do EACH-USP, da POLI-USP e sobretudo do IME-USP, sob supervisão de Marco

Dimas Gubitoso (Gustavo Luiz Vieira, Clayton Lopes Fonseca, Yuri Roberto Amaral Gaspar, Antônio Augusto Abello, Patrick Abrahão Menani, Hiroshi Hayashida, Fábio Brzostek Muller). Com a proposta de preencher uma lacuna nos estudos etimológicos da língua portuguesa, a aliança entre o uso de ferramentas computacionais e a reunião de especialistas de diversas áreas da linguística e da filologia permitirão que, mediante parcerias e financiamento, em breve estejam disponíveis a datação e a retrodatação das primeiras ocorrências do maior número possível de palavras do português, primeiro passo para a elaboração de um dicionário etimológico completo e fundamentado em bases científicas, algo inexistente até o momento.

## REFERÊNCIAS

CORRIENTE, F. Los arabismos y otras voces medio-orientales del dicionário Houaiss da língua portuguesa. **Filologia e Linguística Portuguesa**, São Paulo, v. 15, n. esp., p. 69-184, 2013. Disponível em: <<http://www.revistas.usp.br/flp/article/view/82819>>. Acesso em: 16 fev. 2018.

CORRIENTE, F. **Dictionary of Arabic and allied loanwords**. Leiden; Boston: Brill, 2008.

CUNHA, A. G. da. **Vocabulário histórico-cronológico do português medieval**. Rio de Janeiro: Casa de Rui Barbosa, 2006. CD-ROM.

CUNHA, A. G. da. **Os estrangeirismos da língua portuguesa: vocabulário histórico-etimológico**. São Paulo: Humanitas, 2003.

CUNHA, A. G. da. **Dicionário histórico das palavras portuguesas de origem tupi**. São Paulo: Melhoramentos: Edusp, 1989.

CUNHA, A. G. da. (Coord.). **Dicionário etimológico Nova Fronteira da língua portuguesa**. Rio de Janeiro: Nova Fronteira, 1982.

HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss Eletrônico da Língua Portuguesa**. Versão 3.0. Rio de Janeiro: Objetiva, 2009.

MACHADO, J. P. **Dicionário onomástico-etimológico da língua portuguesa**. Lisboa: Horizonte, 1981.

MACHADO, J. P. **Dicionário etimológico da língua portuguesa**. Lisboa: Confluência, 1952-1977. 5 v.

VIARO, M. E. **Etimologia**. São Paulo: Contexto, 2011.



# DICIONÁRIOS PARA LÍNGUAS INDÍGENAS DO BRASIL

**Jorge Domingues LOPES**  
**Ana Suelly Arruda Câmara CABRAL**

## **Introdução**

Os dicionários de língua são fontes de dados linguísticos e culturais, que têm como referência básica o léxico nas suas mais diversas contextualizações discursivas e que são concebidos, primordialmente, como banco de dados para os usuários falantes da língua dicionarizada.

No caso específico do Brasil, onde há grande diversidade linguística, e como a maior parte das línguas que representa esta diversidade é indígena, a produção de dicionários de língua se reveste de uma importância muito grande, principalmente com o início do processo de escrita dessas línguas, que desencadeia o surgimento da necessidade de materiais não somente de leitura, mas também que tratem da língua em si nos seus mais diferentes aspectos, como são os casos dos diferentes tipos de dicionários.

Contudo, no âmbito dos estudos linguísticos das línguas indígenas do Brasil (LIBs), os dicionários ainda são os instrumentos menos desenvolvidos, mesmo diante do avançado processo de enfraquecimento em que a maioria dessas línguas se encontra, e com significativa parte delas já dispendo de uma escrita consolidada.

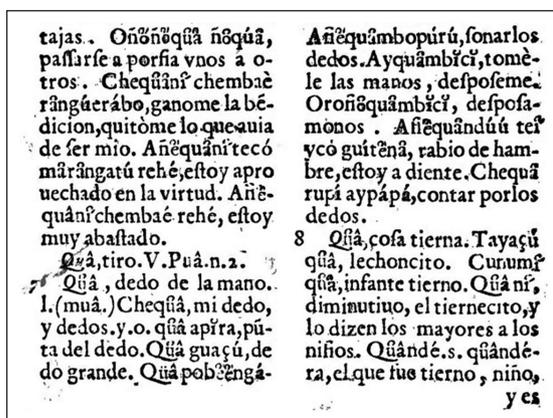
## Breve histórico dos dicionários das LIBs

Os primeiros materiais lexicográficos a serem produzidos efetivamente para LIBs datam do século XVII, como fontes linguísticas e culturais das mais ricas desde então. Num primeiro conjunto de materiais estão aqueles que receberam a denominação *vocabulário* e, somente no século XVIII, é que os materiais lexicográficos de LIBs passam a receber o título de *dicionário*. Porém, mesmo considerando a antiguidade e a quantidade toda a documentação lexicográfica já produzida dessas línguas, longe ainda está o cenário em que elas se encontrem documentadas lexicograficamente como deveriam.

Em se tratando dos principais tipos de dicionários de LIBs – dicionários de língua, dicionários etimológicos, dicionários escolares, dicionários terminológicos, dicionários de tradução –, o panorama é o seguinte:

- dicionários com características de dicionário de língua, de dicionário etimológico e de dicionário de tradução. Por exemplo: *Vocabulário e Tesoro do Guaraní* (RUIZ DE MONTOYA, 1639, 1640); *Dicionário Wayampí* (GRENAND, 1989); *Dicionário de Tupi Moderno* (BOUDIN, 1966)

Figura 1 – Fragmento do *Tesoro de la Lengva Gvarani*



Fonte: Ruiz de Montoya (1639, p. 325, verso).

Figura 2 – Fragmento do *Vocabulario de la Lengua Guarani*

| P. ante E.  | P. ante E.   | 139   |
|---|--|---|
| Pesadilla. Querãç. Che-<br>môque râç mbæ, causa-<br>me pesadilla.                             | Pescado. Pirá.   | Pescador. Yeporaca hára.<br>Pindapoi tára.  |
| Pelado en háblar. nãç pohñ<br>añbae. nãç ngãç i. l. ñãç ñãç<br>tã rerequã ra. Yũç ñãç<br>bac. | Pescador q me pesca. Chẽ<br>pirári poracá hára. Che-<br>pinda poltára. | Pescar, caçar. Poracá.  |
| Peladumbre. Pã ecotõç.<br>Pãpochĩ.  | Pescar cercãdo el río. Pã-<br>pã pipẽ amboá pirá.                      | Pescar con ançuelo. Apin-<br>dãpoi. Ta.   |
| Pesar del bien ageno. Am-<br>boaç abã recõ catú. A-<br>nẽmboçyai heco catú<br>rchẽ.           | Pescar có red. Apiramboá<br>pãçã pipẽ.                                 | Pescar para otro. Aypora-<br>Pescar para si. Ayporacá.<br>Pescõcon dar. Aytuarúpã.<br>Pescuçõ. Ayú. |
| Pesar en balança. Ahãã m-<br>bacããngã nãẽmbẽ yõ-<br>baũbae pipẽ. Pẽçu pipẽ<br>ahãã.           |  |   |

Fonte: Ruiz de Montoya (1640, p. 139).

Observando, sobretudo, o *Vocabulario*, apesar de apresentar como entrada do verbete a palavra em Espanhol, possui toda a informação interna (formas correspondentes e enunciados) do verbete escrito na língua Guaraní, sem tradução para o espanhol.

Figura 3 – Fragmento do *Dicionário de Tupi-Moderno*

|  |   |
|--|---|
| pado: impúbero(meninos) (guar.; <b>apia</b> ); <b>apiaw-ahĩ</b> (vide supra) = mção de uns dez anos.   | ferir ou tocar), tiro de espingarãa impacto.  |
| <b>Apiha</b> (Mont. I p. 392 — <b>apiçaquá</b> =), ouvido (parte interna) (guar.; <b>apisa</b> ); <b>apiha</b> = fecho de mola (vide <b>muka-apiha</b> ); <b>apiha(i)-im</b> = surdez (guar.; <b>apisa'im</b> ); <b>apiha(i)'ĩ-ma'ẽ</b> = surdo: <b>apihaka</b> (B.C. p. 41 — <b>apisacá</b> = mandar o ouvido) = estar ensurdecido: <b>hẽ-apihaka</b> = estou ensurdecido: <b>apihak-ahĩ</b> = ensurdecido (tiro etc.); <b>hẽ-apihak-ahĩ</b> = estou ensurdecido: <b>apiha-katu</b> (Restivo p. 407 — <b>cheapiça katu</b> = oir atentamente) = ouvir bem (contrário de surdo): <b>hẽ-apiha katu</b> = eu ouço bem (guar.; <b>apisa-katu</b> ); <b>apiha-kwa-ahĩ</b> = dor de ouvido: <b>apiha-kwar</b> (Mont. I p. 392 — <b>apiçaquá</b> = oido, el organo) = ouvido (canal auditivo): <b>a-zu-pĩk hẽ-apiha-kwar</b> = eu me tapo os ouvidos (guar.; <b>apisa-kwã</b> ); <b>apiha-</b> | <b>Api-hẽm</b> , vale, thalweg.<br><b>Apik</b> (reg.: Mont. I p. 338 — <b>cambi tipĩ ag</b> = leite quaxada): coalhar: <b>kamĩ u-apik</b> = o leite coalhou (guar.: <b>wapĩ</b> ).<br><b>Apik</b> (reg.: Mont. II p. 51 — <b>apig</b> = sentar-se: B.C. p. 42 — <b>apig</b> = assentar-se): sentar-se, assentar-se: <b>acocoor-se: u-apik</b> = ele sentou-se: <b>ẽ-apik!</b> = sente! — ( <b>wapĩ</b> ); <b>apikaw</b> (Mont. II p. 51 — <b>apicã</b> = banco: <b>apicãb</b> = banco, assento) = banco, assento (guar.; <b>apikã</b> ); <b>apik. .ho</b> (de <b>apik</b> = sentar-se: <b>ẽ-ho</b> = ir) = afundar: <b>u-apik o-ho</b> = ele afundou: <b>apik. .ho ã-pipẽ</b> (vide Mont. p. 291 — <b>ã-pipẽ</b> = no fundo da água) = ir a pique: <b>ãar-uhũ u-apik o-ho ã-pipẽ</b> = o batelão íc: a pique. |

Fonte: Boudin (1966, p. 30).

Já no dicionário de Boudin (1966), um dos mais completos em termos de registros de língua já produzidos no Brasil, além da informação de cunho etimológico (citando com bastante frequência em diferentes partes do verbete fontes mais antigas de outras línguas Tupí-Guaraní, como, por exemplo, Restivo e Montoya, os principais gramáticos do Guaraní Antigo), a ênfase está na língua indígena, para a qual há copiosa documentação.

Com isso, importante é destacar que dicionários, de um modo geral, podem apresentar diferentes perspectivas linguísticas:

- **Monolíngues:** Inexistente no caso das línguas indígenas (há projetos em andamento, mas nenhum ainda resultou em publicação impressa disponível à respectiva comunidade indígena e/ou ao público em geral).

- **Bilíngues:** A maioria dos dicionários de línguas indígenas são bilíngues em **Língua Indígena-Português**, mas poucos trazem a versão **Português-Língua Indígena**, e quando trazem uma versão assim, trata-se, em geral, de correspondência vocabular.

- **Multilíngues:** Há poucos dicionários multilíngues no Brasil, dentre os quais podemos citar o *Dicionário Cinta-larga-Português-Inglês* (SANDBERG, 1978) e o *Dicionário Canela* (POPJES, 2005).

A título de exemplo, apresentamos, inicialmente, dois dicionários bilíngues de LIB produzidos no Brasil.

**Figura 4** – Fragmento do *Dicionário da Língua Asuriní do Tocantins*

|  |
|--|
| <p><b>-hóp</b> n Ia ‘fuligem’<br/>tatahówa sé moasóm ‘a fuligem me fez espirrar’</p> <p><b>hyjrón</b> n III ‘esp. de papagaio (segundo os Asuriní, quando esse papagaio sonha à meia noite é porque alguém vai adoecer; as mães colocam, então, brasa embaixo da rede da criança para que ela não adoça.)’<br/>hyjróna ikérahým ‘hyjróna sonha’</p> <p><b>-hyké</b> n Ia ‘mãe (fala de mulher, afetivo)’<br/>peká sowé pesehýke pé ‘sejam bons para a mãe de vocês’</p> <p><b>hysohó</b> n III ‘esp. de papagaio que come ingá’<br/>hysohóa o’ó ingá ‘hysohóa come ingá’</p> |
|--|

Fonte: Rodrigues e Cabral (2003, p. 54).

**Figura 5** – Fragmento do *Dicionário Yuhup-Português*.

|  |
|--|
| <p>ã <i>pron. poss.</i> primeira pessoa do singular “meu, minha”: <b>ã móm yohoyi num níi m’ di’</b> <i>estou indo procurar meu machado um pouco mais para lá.</i> ◇ <b>ãh</b> eu. ◇ <b>ãnih</b> me, para mim.</p> <p><b>ag<sup>1</sup></b> <i>sub.</i> fruta.</p> <p><b>ag<sup>1</sup></b> <i>v.intr.</i> frutificar, dar frutos: <b>wah-naw ba’ã wap agi</b> o abiu está dando muito fruto.</p> <p><b>agi</b> <i>interj.</i> “ai!”</p> <p><b>ah<sup>1</sup></b> <i>loc.adv.</i> distância média: <b>pön ah</b> <i>meia distância rio acima;</i> <b>merah</b> <i>meia distância rio abaixo;</i> <b>põ ah</b> <i>meia altura acima (vertical);</i> <b>tu ah</b> <i>meia altura abaixo (vertical).</i></p> <p><b>ah<sup>2</sup></b> <i>posp.</i> relativizador: <b>ãh ah booro’ hámáp</b> <i>eu é que irei para a roça;</i> <b>ãm ín ah booro’ hámáp</b> <i>sua mãe é que irá para a roça;</i> <b>ãm ah háy hámáp</b> <i> você é que irá para o mato;</i> <b>ãm sât ah hóp day’ ayáp</b> <i>seu irmão é que irá pescar.</i> <b>Níih tóh wayap ín tön hū’i wayap yi’ ni këy hóp ah ín biih wene sorop</b> <i>Agora nós já temos tudo (produtos não indígenas de pescaria), por isso, comemos mais peixe do que caça.</i></p> |
|--|

Fonte: Silva e Silva (2012, p.115).

Dicionários como estes de Rodrigues e Cabral (2003) e de Silva e Silva (2012) apresentam na microestrutura informações adicionais, como notas linguísticas, informações culturais e remissivas a formas variantes.

Além disso, alguns dicionários incluem textos que completam a informação já dada no verbete correspondente, como é o caso da proposta do *Dicionário Bilingue Terena-Português*, de Silva (2013).

Figura 6 – Fragmento do *Dicionário Bilingue Terena-Português*

|  |   |
|--|---|
| <p>(→miyoti, <b>namukoti,</b><br/><b>nemukoti, sipokoti, sipone,</b><br/><b>veoponoti</b>)</p>   |    |
| <p><b>viajaxoti</b> v. viajar. <b>Viajaxotimo</b> undi iharotike. "Eu vou viajar amanhã"</p>   | <p>Fonte: <a href="http://suguedes.blogspot.com.br/2012/02/bem-te-vi.html">http://suguedes.blogspot.com.br/2012/02/bem-te-vi.html</a></p>   |
| <p><b>vihehe</b> n. caraguatá (planta, <i>Bromeliaceae</i>)</p>  |    |
| <p><b>vikovoku</b> n. lugar onde balançamos. Enepora pi'ati xuve tikoti <b>vivikovoku</b>. "Nós balançamos nesses dois pés de árvore".</p>       | <p>"Enepora ho'openo kuahati vituka pia hiha koahamaka peyo kamo vo'oku yomoti ivui kixoixoa ne kamo kuxeati hikena. Kuanemaka ahati eneikoponea ko'ovokuti ya hiuponi koeti kuane enepo noxo senohiko houxoko koyeti ahati eneoponea oukeke ne senohiko houxo koieti, koanemaka enepo eneo ape maka kixoekeaku vahere emouti,ou apemaka kixoekeaku apeyamo orixokoti ya hieuke ne kopenoti"</p>  |
| <p><b>visi visi</b> n. grilo. Enepone visi visi niko noneti. "O grilo come a plantação" (inseto, <i>Gryllidae</i>)</p>                           | <p>O bem-te-vi é um pássaro que fica em cima do cavalo comendo "bichinhos". Ele também é um pássaro que dá avisos, quando há um grupo de mulheres, ele vem bem devagar, pulando de galho em galho, ele fica no galho cantando, avisando que no meio delas uma engravidará. Quando ele canta em cima da casa está avisando que haverá algum desentendimento na casa ou ainda que tem alguém na comunidade falando mal deles. (Fonte: Anésio Alfredo Pinto, 2013)</p> |
|   |   |
| <p>Fonte: <a href="http://amamatureza.org.br/site/grilo-2.67601/">http://amamatureza.org.br/site/grilo-2.67601/</a></p>                          |   |
|   |   |
| <p>Encontrar um grilo dentro de casa é sinal de novidade ou de dinheiro.</p>   |   |
| <p><b>viteovoku</b> n. nossos ancestrais. Enepone <b>viteovoku</b> koati eneko'oviti. "Os nossos ancestrais davam muitos conselhos pra nós".</p> |   |

Fonte: Silva (2013, p. 256).

Outros dicionários podem, numa mesma publicação, apresentar as informações no sentido, por exemplo, LIB-Português e, em seguida, a correspondência vocabular Português-LIB, como o faz o *Dicionário Parintitin-Português Português-Parintitin*, de La Vera Betts (2005).

Nesse dicionário, percebemos que, enquanto na primeira parte há uma quantidade maior de informações em cada verbete (categoria gramatical, nota sobre cultura, formas remissivas), na segunda

parte o que há é, basicamente, uma lista de vocábulos que remete para a forma correspondente na LIB.

**Figura 7** – Fragmentos do *Dicionário Parintitin-Português Português-Parintintin*

|             |  |
|-------------|--|
| -akym       | d: molhado   |
| -akyr       | d: não-maduro, verde; o período da vida do homem antes da puberdade        |
| -'am        | vi: estar de pé; <u>veja</u> -po'am, -pyta'am (cf -hin, -mbohovivur, -vyr) |
| aman        | s: chuva; <u>veja</u> amandy (cf -ayviku' iuhu, -kyr)                      |
| amandiju    | s: algodão; <u>veja</u> kwaramandiju'i                                     |
| amandy      | s: água da chuva; <u>veja</u> aman   |
| chocar      | -mbukuv  |
| choque, dar | -mondyi  |
| chorar      | -jehe'o, -joporehyvytyg; parar de -- -japo'og                              |
| chorona     | -aãahỹ, -apiry   |
| choroso     | -maran   |
| chover      | -kypokaãuhũ, -kyr, -kyrapo'a   |

Fonte: La Vera Betts (2005).

Outro tipo de dicionário que podemos citar aqui é aquele que busca relacionar mais de duas línguas. Esse tipo de material multi-língue é, porém, bastante raro no conjunto de materiais lexicográficos das LIBs. A título de exemplo, tomemos o *Dicionário Canela*:

**Figura 8** – Fragmento do *Dicionário Canela*

|   |   |
|---|---|
| <b>acãnter</b> (de: cãnter) D.1.7. n. seu filho, um só. <i>your son, one only.</i>  | <i>it was him; you; I that... (emphasis on the subject). Ahhêr me atea ca me apãmjê krêr. Era você, você mesmo que comeu a carne do seu pai. It was you yourself that ate your father's flesh. atea carries emphatic meaning also. D.2.6.</i>           |
| <b>acatac</b> (de: catac) B.11.2. vt. morder forte. <i>bite hard. Rop te acatac. O cachorro te mordeu. The dog bit you. mordida de cachorro, papagaio, porco, jacaré. bite from dog, parrot, pig, alligator.</i>  | <b>ahhi</b> n. varredoras, coisas pequenas. <i>sweepings, small things. B.7.4.</i>  |
| <b>acaxwỹj</b> (de: caxwỹj) D.1.7. n. a sua filha, uma só, não é possível no plural. <i>your daughter, one only, not possible in plural.</i>  | <b>ah.hu.ti</b> (de: huti) A.8.5. n. thick all weeds. <i>SIN: ahkêt.</i>  |
| <b>actô</b> vt. muitos tipos de coisas. <i>many kinds of things. A.11.2.</i>  | <b>ahhyr</b> [ahta] <i>lay side by side. Wa ha paper to ajpên mã ahta. I will put the papers side by side. Hũmre pjê te ajpên mã ahhyr cajpuaw mã. The man walked along the boundary between to lands, not touching, leaving a small space between.</i> |
| <b>acxà quê</b> outra vez; depois de várias vêzes; regularmente. <i>again; after many other times: regularly. Wa ha acxà quê na Barma tũ. Eu irei a Barra de novo. (depois de muitas outras viagens regulares) I am going to Barra again. (after many other regular trips) B.9.</i> | <b>ah.kêt</b> (de: kêt) A.08.04. n. arbusto, mato, varredora ou lixo de chão. <i>bush, jungle, sweepings from floor.</i>  |

Fonte: Popjes (2005, p. 2).

Esse dicionário apresenta a entrada na LIB e formas correspondentes em duas línguas, Português e Inglês. Da mesma forma são apresentados os exemplos, sempre nas três línguas.

É possível afirmar que maioria dos dicionários construídos a partir da década de 1970 compartilham características de macroestrutura e de microestrutura.

Há o predomínio do uso da ordem alfabética na organização da macroestrutura desses dicionários – apesar de haver uma quantidade considerável de dicionários organizados por campos semânticos.

Com relação ao ordenamento das entradas por campos semânticos, é necessário destacar que os dicionários que apresentam essa perspectiva têm, em geral, uma grande dificuldade em estabelecer coerência para inserção de palavras em determinados campos semânticos, como, por exemplo, na decisão de inserir *awaxi'i* e *awaxi pororok* na mesma categoria, como se vê na Figura 9.

**Figura 9** – Fragmentos de dicionário organizado por campos semânticos

|  |  |
|--|--|
| <i>A.7.6 Insetos</i>                                     |  |
| <b>takangyr</b>  | a tocandira (a picada dela dói muito)  |
| <b>takangyruhu</b>                                       | a tocandira; a tocandera (tamanho maior; <i>uhu</i> , grande)                        |
| <b>tapia 'i</b>  | uma variedade de formiga   |
| <b>tapuru</b>  | o berne; a ura; o gorgulho; a lagarta  |
| <b>taraku'a</b>  | uma variedade de formiga   |
| <b>tarawe</b>  | a barata   |
| <b>tawarukwa howy</b>                                    | a borboleta azul e grande  |
| <b>tawarukwa pihun</b>                                   | a borboleta azul e grande com faixa preta nas asas                                   |
| <b>taxi</b>  | a formiga-cabaça   |
| <b>taxi pirã</b>   | a formiga-de-fogo  |
| <b>taxi tawa ra'yr</b>                                   | a formiga-doceira; formiga açucareira  |
| <b>to'ok</b>   | a taoca (formiga)  |
| <b>tukur</b>   | o gafanhoto  |
| <b>tukuxixik</b>   | uma variedade de gafanhoto   |
| <i>A.8.3 Verduras e Frutas Comestíveis (Veja A.8.4.)</i> |  |
| <b>awaxi 'i</b>  | o arroz (antigamente se falava <i>awaxi 'i</i> mas hoje em dia se fala <i>ahúi</i> ) |
| <b>awaxi pororok</b>                                     | a pipoca   |
| <b>irima</b>   | o limão  |
| <b>jaxi mani 'i</b>                                      | mandioca-inajá   |
| <b>jupukái</b>   | as castanhas da sapucaia   |
| <b>jurumu</b>  | o jerimum  |
| <b>iytyk</b>   | a batata-doce  |
| <b>kaka</b>  | o cacau cultivado  |
| <b>kakaran</b>   | o cacau do mato  |
| <b>kara</b>  | o cará   |
| <b>kantái</b>  | a castanha   |
| <b>katume</b>  | a banana-maçã  |
| <b>kawasu</b>  | a cabaça   |
| <b>kúí</b>   | a cuia   |

**Fonte:** Kakumasu e Kakumasu (2007, p. 30 e p. 32).

Vários dos dicionários que são apresentados como bilíngues, são, na realidade, parcialmente bilíngues. Um dos tantos problemas que subjazem à feitura dos dicionários é a ausência de *corpus*, bases fundamentais na sua elaboração, que tenham como referência o léxico da língua nativa e não o português, por exemplo.

Há ainda o fato de que a grande maioria dos dicionários de línguas indígenas está com edição esgotada e indisponível para as comunidades falantes das línguas.

Chegamos, então, à pergunta: para quem são os dicionários das LIBs? Para os linguistas, antropólogos, e outros estudiosos das línguas e culturas indígenas? Para um público erudito ou curioso? Para os falantes nativos das línguas?

### **Pensando em possibilidades de dicionarização das LIBs**

Sem dúvida, há grande necessidade de construção de projetos para documentação de línguas indígenas brasileiras, com ênfase na produção de materiais lexicográficos, mas há também de projetos que permitam a inserção dos falantes das línguas indígenas como pesquisadores de suas próprias línguas.

Dentre as várias barreiras encontradas pelos indígenas para participarem desse processo, está o próprio manuseio dos programas de computador que, em geral, são voltados ao uso por especialistas e exigem, muitas vezes, conhecimento avançado em informática.

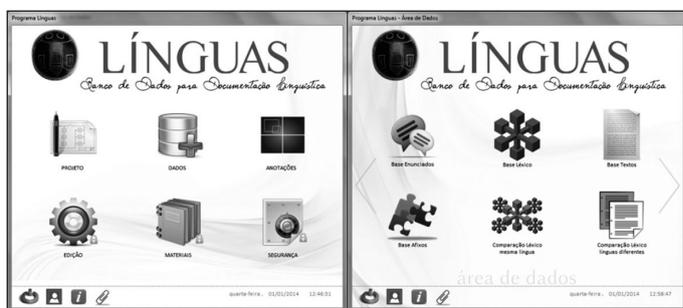
Considerando essa realidade, foi desenvolvida uma proposta de programa de computador, em português, que serve tanto para o armazenamento de dados linguísticos quanto para a produção de materiais lexicográficos (p. ex.), a partir desses dados. Trata-se do *Programa Línguas*.

### **Programa Línguas**

*Línguas – Banco de Dados para Documentação Linguística* é um programa de gerenciamento de banco de dados baseado em *Access* e *VBA*, cuja principal finalidade é auxiliar o trabalho de pesquisadores (sobretudo indígenas do Brasil, enquanto pesquisadores de suas próprias línguas) na documentação e análise de línguas.

Esse programa foi desenvolvido como parte da metodologia da tese sobre documentação e descrição da língua Suruí do Tocantins (família linguística Tupí-Guaraní), no âmbito do PPGL da UnB, sob a orientação da Profa. Dra. Ana Suelly Arruda Câmara Cabral, do Laboratório de Línguas e Literaturas Indígenas (LALLI/IL/UnB), e com o apoio do Centro de Estudos Lexicais e Terminológicos (Centro Lexterm), da UnB, coordenado pela Profa. Dra. Enilde Faulstich.

**Figura 10** – Janelas da Área de trabalho e da Área de dados



**Fonte:** Programa Línguas.

Mesmo em sua fase de testes, o programa já é capaz de armazenar e gerar materiais como dicionários monolíngues, bilíngues e multilíngues, além de permitir a pesquisa dos dados. Como exemplo, apresentamos a seguir dois fragmentos de materiais em fase de construção.

Figura 11 – Modelo 1

|   |   |
|---|---|
| <p><b>'og</b> <i>n.</i> uruhywykaj ywya urumumug ywykwara rupi urumuku'om sura ykupyjta urumusarukong uruupir urukarw; ti roga ti kehaw, uruapo pinowa ku urusehe urukehawamu urukupepyg urusehe; 'og. un ti ma'epotawamu aapo rako iako tarukeke taruke ti ma'etiru umuakym rapo amona n usapy'u rapo tata. Ajnon.: <i>asehu tese saupir sene rehe 'og; ti popytywo e apo 'og; ti rehe t aketehewehe ne wy wekow</i> (Wajwera umume'u). ■ Eisag → 'og.ete.</p> <p><b>'ogete.</b> <i>n.</i> uruapo ywya urumuseaupya, uruupir pinow'o ipyteripe ywa umuwyhog supatitawa; 'og. ipukuete muruwisaw erekehaw aschutese; upuga ruaw murerekotarete; upug ruaw hejkwehe n uhuwapyguwi namukuj hejkwehe upug. Ajnon.: <i>uruapo puta 'ogete semu'etawa ukaripe uruapo puta 'ogete t uesag wehe warasu tywo, 'og t uwawa</i> (lkatu umume'u). ■ Eisag → 'og.</p> | <p><b>'og manimea pukujtaw roga.</b> 'og. pykujpykujtaw upin mani'og, ukytyg, ukupepyg, umuhaw upykuj. Ajnon.: <i>ahaputa manime roga pupe wewahema</i> (Wajwera umume'u). ■ Eisag → 'og.</p> <p><b>tapyj.</b> <i>n.</i> ka'ape sepi uruapo tapyj urukehawamun amona wi uruapo tapyj amona neapy awi uruapo; urumonohog sepy ywya iapoa nune sepi a'eramu uruwerur taapone usaw ka'ape sepi; pinowaku sepi uruupir sepi tapyjromamu. ■ Ajnon.: <i>saapo sene tapyjromamu</i> (Wajwera umume'u).</p> |
|---|---|

Fonte: *Dicionário da Língua Aikewara* (Suruí do Tocantins) (LOPES, 2014, p. 6).

Figura 12 – Modelo 2

|   |  |
|---|--|
| <p><b>'ya</b> /ʔia/ <i>n.III</i> água Ex. 'y gygom água fria</p> <p><b>'ykwawa</b> /ʔikwawa/ <i>n.III</i> igarapé Ex. aha puta 'ype eu vou igarapé</p> <p><b>'ywa</b> /ʔiwa/ <i>n.III</i> árvore Ex. ywuhu árvore grande</p> <p><b>a'ete</b> /aʔe'te/ <i>n.III</i> jiboia. • Ver moj</p> <p><b>a'eteterona</b> (ipewara) /aʔetete'róna/ <i>n.III</i> sucuri. • Ver moj. (<i>Eumeces murinus</i>)</p> <p><b>akara</b> /aka'ra/ acará. (<i>Fam. Ciclideos</i>)</p> <p><b>akyky</b> /aki'ki/ <i>n.III</i> guariba. (<i>Alouatta guariba guariba</i>)</p> | <p><b>ka'i</b> /ka'ʔi/ <i>n.III</i> macaco-prego 1. ka'i iwete okowa'e; upyhyg iwyw ka'ia so uputiawarimu irahaj awpita so akakaw yware; ka'ia awiton iapina pisuna o macaco-prego vive nas copas das árvores; ele desce para pegar ouriço da castanha e sobe com ele na árvore e começa a bater no pau até o ouriço rachar; ele é marron com a cabeça preta Ex. ka'ia usemuaraj yware ywokomumepem o macaco-prego brinca no galho da árvore e quebra os galhos. (<i>Cebus nigrinus</i>)</p> <p><b>kahena'i</b> /kahena'ʔi/ <i>n.III</i> sagui. (<i>Fam. Cebidae</i>)</p> <p><b>katinkwera</b> /kati'jwera/ <i>n.III</i> catingueiro. • Ver misara. • Sin. pakwohow. (<i>Mazana gouazoubira</i>)</p> |
|---|--|

Fonte: *Dicionário da Língua Aikewara* (Suruí do Tocantins) (LOPES, 2014, p. 8).

A contribuição que se dá com esse programa e exemplificações de produtos não se limita à construção de ferramenta e de armarizamento e sistematização de dados linguísticos, mas contribui, sobretudo, para a padronização de dados para sua utilização imediata por quem os coletou e também para o próprio grupo indígena

falante de sua língua. Dessa forma, a nossa proposta materializa a ideia de Austin (2013) de uma documentação transparente, disseminada nas comunidades para uso em seu benefício:

A documentação linguística relaciona-se com os principais métodos para o registro e análise de línguas maternas e materiais culturais, e metadados sobre elas, em modos que são transparentes e reconhecíveis, e eles podem ser arquivados e disseminados junto às gerações presentes e futuras para uso. Muitos pesquisadores têm enfatizado a padronização de dados/metadados, análises e “boas práticas” (p.ex. E-MELD, OLAC), enquanto outros têm argumentado em favor de uma diversidade de abordagens que reconhecem o único contexto social, particular, cultural e linguístico no qual as línguas são usadas individualmente (AUSTIN, 2013, p. 4, tradução nossa).

### **Considerações finais**

Se o cenário de produção de dicionários no Brasil, no último quarto de século, melhorou consideravelmente, ainda estamos bem longe de um quadro em que todas as LIBs ainda vivas (mas também as línguas indígenas já extintas) tenham materiais lexicográficos produzidos com todo o rigor científico.

Para tanto, é necessário um avanço nos estudos lexicográficos dessas línguas, tanto para o aperfeiçoamento das técnicas de construção, quanto da própria crítica lexicográfica desses materiais.

Pensado, portanto, à luz de uma política voltada para o melhoramento das estratégias de fortalecimento das línguas e culturas indígenas do Brasil, o *Programa Línguas* pode contribuir como ferramenta lexicográfica e de documentação linguística para a otimização de trabalhos lexicográficos das LIBs, com a certeza de que isso só será possível com os falantes dessas línguas na qualidade de autores ou coautores e principais beneficiários.

## REFERÊNCIAS

AUSTIN, P. K. Language documentation and meta-documentation. In: OGILVIE, S.; JONES, M. (Ed.). **Keeping Languages Alive: Documentation, Pedagogy and Revitalization**. Cambridge: Cambridge University Press, 2013. p.3-15.

BETTS, L. V. **Dicionário Parintintin-Português, Português-Parintintin**. Brasília: SIL, 2005. Edição eletrônica. Versão original de 1981.

BOUDIN, M. H. **Dicionário de Tupi-Moderno (Dialeto tembété-nê-téhar do alto rio Gurupi)**. São Paulo: Governo do Estado, Faculdade de Filosofia, Ciências e Letras de Presidente Prudente, 1966.

GRENAND, F. **Dictionnaire Wayápi-Français, lexique Français-Wayápi (Guyane Française)**. Paris: Peeters: SELAF, 1989.

KAKUMASU J. Y.; KAKUMASU, K. I. **Dicionário por Tópicos Urubu-Kaapor-Português**. 2. ed. Cuiabá: Associação Internacional de Linguística, 2007.

LOPES, J. D. **Dicionário da Língua Aikewara (Suruí do Tocantins)**. Brasília, DF, 2014. Não publicado. Manuscrito disponível com o autor.

POPJES, J. D. **Dicionário Canela**. Brasília: SIL, 2005.

RODRIGUES, A. D.; CABRAL, A. S. A. C. **Dicionário da língua Asuriní do Tocantins**. Belém: UFPA: IFNOPAP; Brasília: UnB/IL/LALI, 2003.

RUIZ DE MONTOYA, A. Vocabulario de la lengua gvarani. In: \_\_\_\_\_. **Arte, y vocabulario de la lengua gvarani**. Madrid: Iuan Sanches, 1640. p. 103-234.

RUIZ DE MONTOYA, A. **Tesoro de la lengua gvarani**. Madrid: Iuan Sanches, 1639.

SANDBERG, C. D. **Dicionário Cinta-larga-Português-Inglês**. Posto Indígena Roosevelt. Rondônia: SIL, 1978. Mimeo.

SILVA, D. **Estudo lexicográfico da língua terena: proposta de um dicionário bilíngüe terena-português**. 2013. 292f. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 2013.

SILVA, C.; SILVA, E. **A língua dos Yuhupdeh**: Introdução etnolinguística, Dicionário Yuhup-Português e Glossário semântico-gramatical. São Gabriel da Cachoeira: Pró-Amazônia: AECIPY – Associação das Escolas e Comunidades Indígenas do Povo Yuhupdeh, 2012.

# DICIONÁRIO DO NORDESTE

Nelly CARVALHO

## Introdução

O Nordeste é a região brasileira com maior número de estados e o berço da colonização portuguesa no Brasil, com participação ativa em diversos episódios na história do país. Além disso, possui uma cultura popular diferenciada e considerável, a começar pelo seu modo de falar. Por isso, faz-se necessário armazenar essa cultura linguística por meio de uma coleta de dados que abrange o espaço geográfico da região. Estudos costumam focalizar a relação intrínseca existente entre a língua e a sociedade, na busca por explicações de ordem externa como condicionadoras da escolha de uma forma em detrimento de outra. Contudo, nos últimos anos, também têm sido realizados estudos que correlacionam a Língua Portuguesa falada no Brasil e o espaço geográfico, visando a uma futura sistematização. Tradições, costumes, histórias passadas de geração para geração. Tudo isso faz parte da história de uma comunidade. Por isso, é lícito respeitar e preservar esses elementos presentes na cultura, partilhados por grupos de pessoas que interagem diariamente, em especial, através da língua, uma vez que ela retrata a realidade de um povo, a partir do reconto de um fato, das lutas e desavenças de que participa e de seus atos, o que pode se manifestar através da pintura, do artesanato e da literatura.

Considerando, então, essa diversidade cultural, a averiguação de fatos linguísticos permite ver que traços da língua falada em certas

regiões têm se expandido pelos ambientes urbanos, especialmente por causa da migração, com que o Brasil sempre conviveu. O fato é que a Língua Portuguesa foi ancorada primeiramente no Nordeste, através dos donatários das capitânicas, na época em que a remota Terra de Vera Cruz era somente uma faixa estreita, demarcada pelo Tratado de Tordesilhas. Nas capitânicas mais ao Sul, como São Vicente e São Paulo de Piratininga, a presença dos jesuítas e sua forte doutrinação levou ao uso do *latim* em detrimento do *tupi*, para divulgar a fé cristã com maior propriedade. Na região mais ao norte, na capitania da Bahia e Pernambuco, a língua foi ajustada a novos hábitos fonéticos, recheando-se de termos indígenas e, mais adiante, os de origem africana. Essa modalidade de língua transplantada foi conservada quase sem ser modificada, ao contrário do que ocorrera no Sudeste, dada a amplitude de falantes de outras línguas. Como a língua e a cultura são indissociáveis, a cultura nordestina detém um traço eminentemente popular, distante dos veios escolares e do valor advindo da globalização.

### **Variante nordestina**

O Brasil, como resultado da colonização, tornou-se um arquipélago formado por linhas históricas refletidas no plano sociocultural e linguístico. Nesse aspecto, a língua portuguesa, trazida pelos colonizadores, foi-se propagando em ondas de ação lenta e eficiente sobre os falares indígenas, a partir de núcleos fundamentais como Pernambuco e Bahia. Foi essa língua transplantada, nos primórdios da colonização, a base do dialeto que leva seu povo a falar diferente do resto do país. Do Maranhão à Bahia, fala-se um dialeto cheio de arcaísmos e modismos variados no vocabulário, nos torneios sintáticos, na entonação e na prosódia. Não se fala no Recife e em Salvador como em Sorocaba, Bagé ou Manaus. No Nordeste, a língua portuguesa aportou primeiro. Aqui chegou com os donatários das capitânicas, quando o Brasil, ou melhor, a Terra de Vera Cruz era apenas uma faixa estreita, limitada pelo Tratado de Tordesilhas. Certo que São Vicente, no Sudeste, também foi uma próspera capitania, mas os jesuítas, que por lá viveram catequizando os gen-

tios, preferiam ensinar-lhes em latim e aprenderem a língua geral da costa –o tupi – para melhor divulgar a fé cristã.

Foram os nordestinos que deram início à saga da língua portuguesa no Brasil, adaptando-a a novos hábitos fonéticos, recheando-a de termos de origem indígena e, mais adiante, de origem africana, e guardando esta modalidade de língua transplantada, como um tesouro, sem quase modificá-la, porque a região não recebeu imigrantes, falantes de outras línguas. Partindo de *Assim falava Lampião* e da primeira versão do *Dicionário do Nordeste*, chega o Autor a esta obra de maior extensão e profundidade que, não sendo escrita por especialista (gramáticos, filólogos, lexicógrafos, doutores no assunto) é consequência de um trabalho que utiliza técnicas de jornalismo investigativo, fora dos padrões acadêmicos. Quem ler atentamente esse “labirinto da linguagem nordestina” verificará que este *Dicionário* foi mais longe do que prometiam suas palavras: oferece ao leitor conhecimento de um larguíssimo *corpus* de termos e expressões nordestinas exemplificado em textos representativos dos mais variados domínios da cultura, ao mesmo tempo que enriquece os especialistas com um manancial de pesquisas para surpreender as franjas semânticas das palavras carregadas de novos matizes de ideias. A indagação científica levou-o a estabelecer inteligente elo entre o significado de palavras usadas pelos falantes e sua natureza, o que motiva o nascimento de “termos fortes, rústicos, grosseiros, mas também daqueles dotados de lirismo, sensibilidade poética, bom humor e picardia” (CARVALHO; MOTA; BARRETO, 2000, p.65). A sensibilidade e a cultura deram a dimensão mais acertada de que, apesar da pujança e força desse labirinto riquíssimo que é a linguagem nordestina, isto não diminui os laços que presidem à herança recebida de Portugal. Estas suas palavras: “Se for correto afirmar que há uma ‘língua brasileira’ que se distancia daquela que lhe deu origem, também será válido dizer que esta língua resultou tão impregnada pelas características originais que continuará a se chamar ‘portuguesa’ por um tempo indeterminado, talvez infinito” (CARVALHO; MOTA; BARRETO, 2000, p.65).

Dentro do labirinto original da língua portuguesa reside, com conforto, a atual língua “brasileira”, com contornos próprios, com alma e perfil diferenciados da língua-mãe. No seu interior, da mesma forma, coabitam os subdialetos regionais, marcas registradas de áreas geograficamente extensas e culturalmente heterogêneas: o sertão cearense não é o pampa, o pantanal não é a falésia potiguar, e o coqueiral alagoano em nada se parece com a floresta tropical, a serra gaúcha ou a megalópole paulistana. As populações desses ecossistemas trazem consigo características herdadas do meio físico que costumam se apresentar ou se revelar através do sotaque, vestuário, culinária, costumes e outros aspectos da vida cotidiana que escolheram ou aceitaram. Acrianos e paraibanos falam a mesma língua, mas vivem em realidades tão distintas que apenas uma língua forte como a portuguesa, com capacidade de adaptação e de sobrevivência em ambientes extremos, poderia servir para a comunicação entre os povos da floresta e os moradores do sertão ou do litoral. O mesmo vale para paranaenses e goianos, baianos e catarinenses, cariocas e amazonenses, e tantas outras combinações possíveis num país como o nosso. Visto de perto, o labirinto da língua “brasileira” revela passagens e corredores ainda mais secretos. Os subdialetos no interior das enormes regiões são expressivos, e para encontrá-los basta cruzar a divisa entre os Estados. Incontáveis termos e expressões da fronteira oeste do Maranhão (na divisa “amazônica” com o Pará), por exemplo, nada têm de familiares para os moradores do brejo paraibano ou do sertão baiano. Às vezes, a diferença e desconhecimento ocorrem entre os moradores do litoral e os do sertão, dentro de um mesmo Estado, como é frequente observar em Pernambuco, no Piauí e no Ceará.

### **O papel do dicionário**

Hoje a pesquisa dessas relíquias vocabulares se nos apresenta menos rica em vista de que só bem depois do séc. XVI passa a despertar o interesse de estudiosos. O progresso desde o tempo da colônia e do Império e a abertura de estradas, favorecendo o trânsito de levas migratórias aos mais distantes rincões da nova terra,

já tinham contribuído para a feição multiforme que se apresenta a linguagem popular.

Além da fala popular, a atenção de Fred Navarro (2004), autor do *Dicionário do Nordeste*, que ora analisamos, não desprezou o registro da flora e fauna regionais, nem sempre bem estudadas pela tradição dicionarística luso-brasileira.

Coluna mestra da investigação lexicográfica, o testemunho da existência da palavra e de sua trajetória histórica vem atestado por uma copiosa exemplificação de textos procedentes dos mais variados quadrantes da cultura em língua portuguesa. O Autor estabeleceu diálogo com outros dicionários para obter melhor análise de palavras e locuções.

É que o dicionário não é só o celeiro do idioma, mas é o depositário da cultura. Nele estão contidas as interpretações da comunidade sobre fatos, objetos, pessoas. O dicionário é considerado um oráculo e um tira-teima, respeitado como obra imparcial que reproduz a língua real/social. Na micro-estrutura dos verbetes estão as acepções do termo-entrada bem como abonações que apresentam o funcionamento do termo no discurso nos verbetes. Neles, o sujeito da enunciação é apagado, transmitindo a impressão de neutralidade. Porém a marca do pensar da comunidade, da forma como é interpretada pelo autor está presente nas definições.

### ***Corpus* ampliado**

Só podemos apreciar algo quando nos afastamos, como nos ensina Paulo Freire com seu método. Um certo afastamento é que nos faz ver a realidade de forma objetiva sem o envolvimento que tolda a capacidade de observação e análise. Foi o que aconteceu com o autor do *Dicionário do Nordeste*, Fred Navarro, ao vir morar em São Paulo. A partir de um afastamento do objeto de análise, ele percebeu como a linguagem de seu lugar de origem era marcada por usos diferenciados. Lançou a primeira edição com o nome de *Assim Falava Lampião* e a seguir ampliou o *corpus*, lançando em 2004 a primeira versão do *Dicionário do Nordeste* com 5 mil termos. Agora, concluiu uma alentada segunda versão, ampliada com

uma quantidade enorme de termos entre os quais muitos neologismos. O registro é eclético com corte diacrônico: expressões arcaicas convivem com as atuais, gírias recentes com as antigas, retiradas de dicionários outros, músicas populares e autores regionais.

Na atual versão, com 10.500 verbetes, publicada em 2013, foram inúmeros os neologismos, pois os dialetos regionais não ficam parados no tempo; assimilam as novidades necessárias ao cotidiano dos falantes. Ao todo, foram encontrados no *corpus* que constitui o dicionário 55 neologismos com os mais variados tipos de formação e representando os estados da região: neologismos formais, neologismos semânticos, fraseologias neológicas, e alguns de origem não encontrada.

### Amostragem

A seguir, apresentaremos verbetes representativos de todos os estados do Nordeste.

**abaitolado** \ò\ • *s.m. e adj.* • CE • Efeminado, com jeito de *baitola* ou de viado. SDCE • [não consta dos dicionários consultados]

**abilocil** • *adj.* • PE • Neologismo a partir da aglutinação de ‘abilolado’ com ‘imbecil’, o que ressalta os sentidos pejorativos dos termos: seu irmão consegue ser mais abilocil do que o meu. • [origem de ‘imbecil’: latim ‘imbecillu’ (fraco de corpo, franzino)]

**agueiloado** \êi\ • *adj.* • PE • Neologismo para algo típico de ‘gays’: viu o corte de cabelo dele, todo agueiloado? • v. *aiailbrincar de pedro*

**beré doido** • *s.m.* • BA • Maluco, *zoropitó*, doido. • [‘beré’ não consta dos dicionários consultados, mas sim ‘beré-beré’ (joão-ninguém, zé-mané, indivíduo sem importância, destituído de poder econômico), no DHLP]

**bitelo** • *s.m.* • BA • Bebê forte e corado, ‘vendendo’ saúde. • Provável neologismo, registrado por Jorge Amado em *Dona Flor e seus dois maridos*: “Uma vez, ouvindo-o enaltecer o encanto de um corneta gordo e rosado, um bitelo, prêmio de robustez - infantil a exhibir-se no cromo de uma folhinha de ano, dispôs-se ela a

enfrentar o assunto perturbador: – Se você tem mesmo vontade de ter um filho, eu arrisco a operação.” • [não consta dos dicionários consultados]

**bizurunguinha** • *s.f.* • **chiquitinha** • **PE** • Graciosa, delicada, bonita. • “Isabel querida/– A menininha/mais bonitinha,/mais engraçadinha,/mais bizurunguinha/que eu já vi na minha vida,/amorável,/adorável,/adorável.” *Louvado e prece*, Manuel Bandeira, *Liliana*, Manuel Bandeira, em “Estrela da vida inteira” • “Capitão – Desce da gave gageiro/Que eu te quero abraçar/Essas 3 môrças que viste/Todas 3 hei de te dar/Uma para te coser/Outra para te calçar/A mais chiquitinha dela/Para contigo casar/Te darei tanto dinheiro/Que tu não sabes contar [...]” Citado em *Espetáculos populares do Nordeste*, Hermilo Borba Filho • [prováveis neologismos, não constam dos dicionários consultados]

**boyzinha** • *s.f.* • **PE** • Adolescente, *burrega*, garota. • [neologismo recifense, de origem não identificada]

**bufufa** • *s.f.* • **PE** • Pedaco esfarinhado, farelo. • [palavra não registrada pelos dicionários consultados; provável neologismo, registrado por Marilene Felinto duas vezes em “Postcard”, no conto *As horas abertas*: “Sentia que a olhavam como se de sua boca saíssem bufufas de farinha e ela engasgasse.” • “Silêncio completo, que ela não tinha entendido nada daquela massa de som informe que saía da boca dele como bufufas de farinha.”]

**calamengau** • *s.m.* • **BA** • Transa, sexo, trepada, de acordo com as segundas intenções de Xangai: “Marido se alevanta/e vem tomá um mingau/que é pra criá *sustança*/prá nós fazê um calamengau./Brincadera de manhã cedo/né minha véa/arrisca quebrá o pau/e ai d’eu sodade.” *Ai d’eu sodade*, Anônimo. • [neologismo não registrado nos dicionários e livros de referências consultados]

**catíngoria** • *s.f.* • **N.E.** • Neologismo para designar algo sem categoria ou de categoria inferior: pode parar, já deu pra ver a sua catíngoria.

**cavandante** • *s.m.* • **BA** • Cavaleiro andante, em neologismo registrado por Elomar: “Cavandante eu sou/por este reino sem-fim,/meu cavalo voou/procurando um lugar/que minha avó contava pra mim,/meu menino do São Joaquim,/cavaleiro do São Joaquim.” *Cavaleiro do São Joaquim*, Elomar

**contenteza** \ê\ • *s.f.* • **N.E.** • Estado de quem está alegre, contente. • Elba Ramalho e Chico César mencionam: “A contenteza do triste/*tristezura* do contente/vozes de faca cortando/como o riso da serpente [...]” *Béradêro*, Chico César • [não consta dos dicionários consultados] • [origem de ‘contente’: latim ‘contentu’ (limitado, represado, contido)]

**conversê** • *s.m.* • **BA** • // • **converseiro** • **PB / RN/ .MA** • Conversa mole, *bolodório*, enrolação. • A primeira é neologismo registrado por Jorge Amado em *Dona Flor e seus dois maridos*: “Primeiro o concerto, depois a pitaça. Por mais tentasse Édio, com um conversê tapeativo, adiar o momento do desastre, não o conseguiu, não obteve prazo nem apelação.” • [‘pitaça’, na citação, é qualquer tipo de comida, derivado de um dos sentidos originais do termo (refeição de excepcional qualidade, oferecida nos dias de festa)] • • “Um locutor tabacudo/de *converseiro* comprido/uns alto-falante rouco/que espalhe o alarido/microfone com flanela/ou vermelha e amarela/conforme a cor do partido.” *Comício de beco estreito*, Jessier Quirino.

**corniferação da humanidade** • *s.f.* • **CE** • Neologismo registrado pelo cantor cearense Falcão para nomear o processo (irreversível segundo ele) que transformará todos os homens e mulheres do planeta em chifrudos, cornos ou galhudos.

**derroteira** \ê\ • *s.f.* • **PE** • Desgraça, tragédia, trauma. • “A marca de Airine era no dedo, que tinha perdido na máquina de tecelagem, restando um toco de indicador. Foi em Paulista, dizia que foi. Na fábrica, em Paulista. – E não foi? Pois foi... Uma derroteira.” *O lago encantado de Grongonzo*, Marilene Felinto • [não consta dos dicionários consultados]

**desempacavirar** • *v.t.d.* • **SE** • Desempacotar, desembrulhar. • “Correram semanas. Adoei. A artrite amarrou-me à espreguiçadeira, o meu desgraçado corpo se cobriu de manchas. Capengando, abri a estante, exumei *O cortiço*, desempacavirei-o, restituí-o à convivência dos outros romances.” *Infância*, Graciliano Ramos • [não consta dos dicionários consultados]

**desenrolança** • *s.f.* • **PB** • Desenvoltura, autoconfiança, autonomia. • Neologismo registrado por Jessier Quirino: “E pegou o

microfone/com a maior desenrolança/falou de roubo, de jogo/de traficância e matança/partiu no mei três partido/partiu depois no comprido/dobrou e fez uma trança.” *Malandro na eleição*

**desvisível** • *adj.* • PE • Que não quer se ver, ‘invisível’, que se desconhece de propósito. • Neologismo registrado por Gilvan Lemos: “Todos como se não me conhecessem. Tão desvisível assim, de supetão, fiquei? Só você, Mardônio, inda me dá atenção.” *A lenda dos cem.*

**diferencita** • *s.f.* • PE • Neologismo para pequena diferença, registrado por Manuel Bandeira em “Estrela da vida inteira”, no poema *Joanita*: “Não é Joe, não é Joana,/Nem Juanita; é Joanita./A diferença é pequena,/Mas nessa diferencita,/Que em suma é tão pequenina,/Há a graça que não está dita,/Que é privilégio da dona,/Que já toda a gente cita/E assim talvez não reúna/Nenhuma moça bonita.”

**domingar** • *v.int.* • BA • Passear no dia de domingo. • “Já liquidei,/eu liquidei a prestação do paletó,/do meu sapato e da camisa/que eu comprei pra domingar com meu amor/lá no Cristo, lá no Cristo Redentor.” *É fim de mês*, Raul Seixas • “O jornal de manhã chega cedo/mas não traz o que eu quero saber./As notícias que leio conheço/já sabia antes mesmo de ler – ê, ê./Qual o filme que você quer ver – ê, ê./Que saudade, preciso esquecer – ê, ê./É domingo, ê, ê, domingou, meu amor.” *Domingou*, Gilberto Gil/Torquato Neto • [Torquato Neto atribui a Tom Zé a ‘invenção’ do verbo ao incluí-lo na letra da música *Dique do Tororó*: “Ê, ê, Tororó, ê, domingou/O dique está nascendo, ê, no domingá.” *Tropicalista lenta luta*, Tom Zé ] • [não consta dos dicionários consultados]

**emboloar** • *v.int. e v.pron.* • N.E. • Equivale na Região aos verbos ‘embolar’ ou ‘embolotar’ (encaroçar, encher-se de caroços). • [não consta dos dicionários consultados]

**embuchar** • N.E. • 1 • *v.t.i.* • Engravidar de alguém, *embarregar*, ficar grávida (barriguda, buchuda). • v. • 2 • *v.t.d.* • e *v.int.* • Engravidar alguém, provocar gravidez, *acertar na veia mestra*, *encher*, fertilizar. • [nessas acepções, não consta dos dicionários consultados]

**em situação desempregática • fraseol. • PE •**, Sem trabalho, desocupado, desempregado. Neologismo registrado pelo jornalista Homero Fonseca em *Roliúde*: “Estando eu em situação desempregática, seu Bartolomeu, aposentado de pouco, queria me ajudar. Me chamou pra uma conversa no quintal da casa dele, em Beberibe.” *Roliúde*, Homero Fonseca

**encardir o canelau • fraseol. • CE •** Zoar, *grear*, irritar, provocar. • [*canelau* é neologismo, não consta dos dicionários consultados]

**espoletado** *ve* ou *èl* • **adj. • N.E. • 1** • Enraivecido, encapetado, fulo da vida: ela já chegou espoletada da rua. • **2** • Atrevido, ousado, disposto a tudo. • “Não que eu goste de ficar ouvindo conversas dos outros, mas Tião Grande era homem suspeitoso, espoletado, cheio de manhas quando desejava levar vantagem no negócio.” *Caldeirão: a guerra dos beatos*, Cláudio Aguiar • [não consta dos dicionários consultados]

**esporreteado • adj. • BA •** Neologismo para ‘nervoso’, de cabeça quente, de pavio curto, mencionado por Jorge Amado: “Com diploma impresso e quadro de formatura exposto em loja da avenida Sete, desde uma turma antiga, à qual pertencera dona Oscarlinda, enfermeira de categoria, funcionária do Hospital Português, esbelta e esporreteada, doida por um enredo.” *Dona Flor e seus dois maridos*

**Funhenha.adj. PI:** arrepio e queimor.

**Piranhudo .adj. AL:** pessoa que gosta de frequentar prostíbulos

**Sanhuça .adj. AL:** bandido, capanga, desordeiro.

**T / NE:** peça dupla ou tripla em forma de T, que permite o múltiplo uso de uma mesma tomada, Benjamin

## Conclusão

Pelo exemplo dos verbetes apresentados, representando os vários estados do Nordeste brasileiro, podemos observar a presença de uma variedade linguística com características comuns, confirmando o que observou Antenor Nascentes, no pioneiro e já distanciado estudo sobre a dialetologia no Brasil. Apesar de características lexicais diferentes em cada estado, o vocabulário é intercompreen-

sível e tem, na maioria dos casos, origem episódica, mas guarda um traço de arcaísmo, próprio da primeira região a ser colonizada e que recebeu pouca influência de contingentes de imigrações posteriores, no século XIX, como é o caso das regiões do Sul, Sudeste e Centro Sul.

### **REFERÊNCIAS**

CARVALHO, N.; MOTA, S. K.; BARRETO, J. R. P. **Dicionário do Frevo**. Recife: Ed. da UFPE, 2000.

NAVARRO, F. **Dicionário do Nordeste**. São Paulo: Estação Liberdade, 2004.



# QUANDO A ELABORAÇÃO DE DICIONÁRIOS PODE AUXILIAR NA REAFIRMAÇÃO DE UMA IDENTIDADE ÉTNICA: O CASO DOS PROFESSORES MURA DE AUTAZES

**Cristina BORELLA**  
**Eneida Alice Gonzaga dos SANTOS**

## **Introdução**

O povo indígena Mura ocupa vastas áreas no complexo hídrico dos rios Madeira, Amazonas e Purus. Vivem tanto em Terras Indígenas, quanto nos centros urbanos regionais, como Autazes, Borba e Manaus. São descritos, desde o século XVII, como “um povo navegante, de ampla mobilidade territorial e de exímio conhecimento dos caminhos por entre igarapés, furos, ilhas e lagos” (AMOROSO, 2009).

A língua Mura, pertencente à família linguística Mura, era falada até o início do século XX. De fato, a única documentação que se tem desta língua se resume a uma lista de aproximadamente 15 palavras publicada em 1925, por Curt Nimuendajú. Segundo Amoroso (2009), já no final do século XIX, há relatos que demonstram que além de falarem sua língua nativa, os Mura também

dominavam a Língua Geral, ou Nheengatú, no intercâmbio com brancos, negros e demais populações indígenas:

Em 1826 um observador anônimo deixou registrado que os Mura da embocadura do Madeira falavam “a língua geral além das suas três gírias - a articular nasal, a gutural e a da gaita” (C. Moreira Neto, 1988: 358). Situação linguística semelhante foi descrita por Barbosa Rodrigues (1975) no rio Urubu; Tastevin (1923) nas proximidades de Manaus; e Nimuendajú, em relação aos Mura dos rios Madeira e Solimões. (AMOROSO, 2009).

Em seu longo histórico de contato, de guerra e de resistência à expansão portuguesa e depois brasileira na Amazônia, o povo Mura sofreu diversos estigmas, massacres e perdas demográficas, linguísticas e culturais, deste modo, deixaram de falar tanto a língua Mura como o Nheengatú e se tornaram, como tantos outros povos indígenas brasileiros, monolíngues em Português<sup>1</sup>.

Conscientes do complexo processo histórico vivido por seu povo e do questionamento ainda presente na sociedade regional sobre a veracidade de sua identidade étnica, os Mura têm buscado, cada vez mais, o reconhecimento de seus direitos e da sua identidade como povo indígena. Este processo levou os professores Mura, pertencentes ao município de Autazes, a buscar uma educação escolar diferenciada que atendesse aos interesses de seu povo. Deste modo, esses professores concluíram em 2013 o “Curso Formação de Professores Indígenas Mura”, oferecido pela Universidade do Amazonas (UFAM), vinculado à Faculdade de Educação (FACED).

Ainda na busca de um fortalecimento da identidade étnica, os professores Mura solicitaram que lhes fosse ensinado, na disciplina

---

<sup>1</sup> Há relatos que ainda existem alguns velhos que falam Nheengatú, entretanto, não sabemos qual o nível de proficiência da língua nestas pessoas, já que não há documentação, seja de cunho antropológico, seja de cunho linguístico, sobre a localização destes falantes. De qualquer modo, é patente que a língua Nheengatú não é mais a língua de uso do povo Mura. Este fato pode ser comprovado nas mais variadas faixas etárias das comunidades Mura.

“Língua Indígena”, a Língua Nheengatú. Por solicitação destes mesmos professores, o Nheengatú também fez parte dos conteúdos programáticos da referida Licenciatura. Não é de se estranhar que os professores Mura desejem aprender/dominar uma língua indígena, afinal, para uma grande parcela (se não maioria) da sociedade regional não-indígena e, por vezes, indígena, a identidade étnica ainda está atrelada ao domínio de determinada Língua Indígena. Assim poder aprender/dominar o Nheengatú, como segunda língua, nada mais é do que um instrumento simbólico que esses professores utilizam para asseveração de sua identidade étnica, perante a sociedade não-indígena e perante a outros indígenas<sup>2</sup>. Como aponta Moraes et al. (2007, p. 8):

Os(as) professores(as), porém, têm consciência que falar ou estudar uma língua indígena não os faz “mais” ou “menos” Mura e que a língua portuguesa é hoje a sua língua materna. Entretanto, entendem que o Nheengatu poderá servir futuramente como forma de defesa e valorização junto a outros povos e frente à sociedade envolvente.

A partir de agosto de 2010, começamos a participar como docentes do “Curso de Licenciatura Formação de Professores Indígenas Mura”, ficando sob nossa responsabilidade a orientação da “Pesquisa em Letras e Artes e Cotidiano do Professor Mura<sup>3</sup>” cujo produto final consistiu em Trabalhos de Conclusão de Curso (TCC).

Trabalhos anteriores ao início do “Curso Formação de Professores Indígenas Mura” já demonstravam que a escola se tornara um espaço considerado importante para o fortalecimento

---

<sup>2</sup> Ainda não é certo que a língua Nheengatú será utilizada, mesmo como instrumento simbólico, pela comunidade Mura. Alguns professores participantes da referida licenciatura já ensinam alguns vocábulos desta língua em sala de aula, outros acreditam que este conhecimento só fará parte do repertório dos professores indígenas.

<sup>3</sup> A partir do terceiro ano, a referida licenciatura é subdividida em três grandes áreas, a saber, Letras e Artes, Humanas e Sociais e Exatas e Biológicas. A pesquisa aqui descrita foi trabalhada com a turma pertencente a grande área “Letras e Artes”.

da identidade étnica deste grupo, conforme aponta Moraes et al. (2007, p. 7)

[...] a escola Mura tem sido vista pelos professores como instrumento que pode colaborar na afirmação cultural e na superação de preconceitos e no fortalecimento interno de solidariedade interétnica [...] a busca pela afirmação identitária implica em aperfeiçoar a escola enquanto espaço de luta, colocando esta instituição a serviço de projetos presentes e futuros que concretizem o reconhecimento de sua identidade Mura.

Deste modo, logo de início, era de suma importância que as pesquisas dos TCC(s) contemplassem a cultura e o cotidiano do povo Mura, envolvendo não só os graduandos, mas seus respectivos alunos, o espaço da escola e a comunidade de um modo geral.

### **A elaboração de dicionários dos saberes Mura – descrição do processo**

Sabemos que o nível lexical de uma língua armazena e acumula aquisições culturais representativas de seu grupo de fala. Isquerdo (2001, p. 91) aponta que o estudo de campos semânticos específicos, como dicionários regionais, pode fornecer dados que “deixam transparecer elementos significativos relacionados à história, ao sistema de vida, a visão de mundo de um determinado grupo”. Quando registrados, de modo sistematizado “itens lexicais de uma língua dão coesão às sociedades e projeção às culturas, porquanto definem a identidade linguística de um povo” (KRIEGER, 2007, p. 297).

Os Mura, como as demais culturas indígenas, conhecem um vasto repertório de designações de fauna e flora. Este conhecimento, até aquele momento, havia sido transmitido oralmente, assim, o processo de levantamento e sistematização deste léxico poderia vir a contribuir tanto para o fortalecimento da identidade étnica quanto da identidade linguística deste povo. Logo após sugerirmos que os TCC(s) poderiam consistir na elaboração de dicionários, foi

unânime a consideração que estes dicionários seriam elaborados pelos e para o povo Mura. Deste modo sugerimos que os TCC(s) consistissem na elaboração de dicionários que levariam em conta, além dos saberes Mura, um português falado/compreendido/utilizado por este povo.

A partir desta decisão e durante todo o processo de confecção dos dicionários, os professores reconheceram o português, não só mais como sua língua materna, mas também como única língua depositária de conhecimentos culturais do referido povo.

### **Dicionários dos saberes Mura e identidade étnica**

A pesquisa, em seu total, levou dois anos e quatro meses para ser realizada. Os projetos para serem desenvolvidos nas respectivas escolas e comunidades foram divididos de acordo com a escola/comunidade a qual pertencia o grupo de professores. Após muitos debates, os professores decidiram que os dicionários que contemplariam parte dos saberes Mura abordariam 03 grandes temas e/ou campos semânticos: peixes, árvores, frutas e plantas medicinais.

A partir do tema do dicionário e dos critérios escolhidos para elaboração dos verbetes, foi possível aos professores Mura envolver seus alunos, que tiveram ampla participação na coleta de dados lexicais com os mais velhos da comunidade. Conforme foi relatado pelos professores, foram feitas reuniões com os membros de suas respectivas comunidades a fim de que o projeto fosse explicado e para que estes recebessem os alunos que visitariam as casas em busca de informações.

O desenvolvimento dessa pesquisa ocorreu a partir de 2010. Quando retornamos do curso de licenciatura, em uma reunião de pais e mestres da escola, falamos que iríamos realizar a pesquisa acima citada. Estiveram presentes nesta reunião, as lideranças da aldeia, os pais, corpo docente e discente. Todos por unanimidade concordaram.

Assim, a pesquisa deu-se de janeiro a abril, na escola, no período das aulas letivas, na turma do 5º ano, para as turmas dos professores Aldimar Pereira e Gilsivaldo Santos. Em grupo de 04 alunos, a coleta dos dados foi feita de acordo com um roteiro pré-elaborado pelos professores. Esses dados somaram-se em um total de 180 espécies de plantas existentes em nossa aldeia. (RODRIGUES; HIAMUTH, 2013, p. 2).

Deste modo, a elaboração destes dicionários teve um papel muito além de simples TCC(s), movimentando toda a comunidade; o produto se tornou algo que todos esperavam ver concretizado.

Pesquisamos com as pessoas mais idosas, parteiras, pajé da aldeia, etc. Cada uma dessas pessoas informou quais plantas que serviam para fazer remédio e combater tipos específicos de doenças. As lideranças disseram que só assim poderíamos voltar ao costume de usar os remédios caseiros e tradicionais [...] (BRAGA FILHO; BRAGA, 2013, p. 2).

Vale salientar que alguns dicionários foram ilustrados pelos alunos Mura e/ou graduandos, outros contaram com fotografias para ilustrar as entradas. Ao final, tendo a coleta de dados em mão, ficou sob a responsabilidade do grupo de professores a sistematização dos dados (entradas e verbetes) e conclusão dos dicionários. Ao todo, foram elaborados nove dicionários, conforme mostra a Tabela 1:

**Tabela 1 – Dicionários**

| <b>Dicionários</b>   | <b>Autores</b>  |
|--|---|
| Dicionário Ilustrado das Plantas Medicinais da Aldeia Murutinga            | Arlindo Ruso Braga Filho<br>Waldenia Mota Braga                               |
| Dicionário dos Saberes Mura: Variedades de Árvores da Aldeia Trincheira    | Alderico Vieira Neto<br>Liene do Nascimento Pereira<br>Maria José M. da Silva |
| Dicionário Ilustrado dos Saberes Mura: Variedades de Peixes do Lago Josefa | Kleber de Almeida Prado<br>Rosa Coelho Martins<br>Valdijane Prado Batista     |

| <b>Dicionários</b>  | <b>Autores</b>  |
|---|---|
| Minidicionário das Plantas existentes na aldeia Iguapenu                          | Aldimar Pereira Rodrigues<br>Gilsivaldo Santos Hiamuth                                |
| Dicionário de Conhecimentos Tradicionais das Plantas Medicinais da Aldeia Gapenu. | Elizabeth dos Santos Ferreira<br>Jerson dos Santos Ferreira<br>Valdívia Prado Batista |
| Dicionário Ilustrado dos Saberes Mura: Variedade de Peixes                        | Benezuete Soares da Silva<br>Rosiane de Almeida Tavares                               |
| Dicionário dos Saberes Mura: Variedade de Árvores                                 | Fernanda Lima de Oliveira   |
| Dicionário das Receitas de Plantas Medicinais da Aldeia São Félix                 | Conceição Marques Lima  |
| Dicionários de Receitas das Plantas Medicinais da Aldeia São Félix                | Luzia Pacheco dos Santos  |

**Fonte:** Elaboração própria.

## **A elaboração das entradas**

Na confecção dos dicionários dos Saberes Mura considerou-se o termo “entrada” como as palavras arroladas alfabeticamente, de acordo com o português falado pelos Mura. Os autores do “Minidicionário das Plantas Existentes na aldeia Gapenu” decidiram estabelecer a separação de sílabas nas entradas, conforme pode ser visto na Tabela 4.

## **A confecção dos verbetes**

Segundo Zgusta (1971 apud FERREIRA, 2005), ao construir os verbetes, o lexicógrafo deve considerar não somente a estrutura da língua em questão, mas também os aspectos culturais específicos da respectiva comunidade linguística. Para exemplificar este fato, a autora supracitada utiliza como exemplos as palavras inglesas “*table*” e “*desk*” e a palavra checa “*stul*”. Para um falante de inglês, é irrelevante se a mobília é de madeira ou de metal, antiga ou moderna, grande ou pequena, pois essas qualidades ou proprieda-

des não são critérios relevantes para caracterizar uma dada mobília como “*table*”. Na verdade, para o inglês, são outros critérios que definem uma determinada mobília como “*table*”: a forma chata, uma superfície horizontal, que possa ser suportada por pernas, etc. Por outro lado, se a peça é usada primeiramente não para comer, mas para escrever e geralmente para trabalhar, ela será chamada/classificada como “*desk*”, em contraste com “*table*” que geralmente não é utilizada neste caso, ou seja, o propósito da mobília é criterial para o falante de inglês. Nem todas as línguas/culturas utilizam os mesmos traços de critérios, por exemplo, os traços criteriais do inglês “*table*” são idênticos com os do checo “*stul*”, porém, para a língua checa, o propósito da mobília não é criterial, ou seja, é irrelevante se ela é usada para comer, jogar, trabalhar ou escrever. Uma boa parte do trabalho do lexicógrafo é estabelecer a criterialidade, ou seja, o que é criterial/relevante e o que não é. Ao começarem a estabelecer campos semânticos específicos, os professores Mura também começaram a se perguntar qual criterialidade utilizariam nos verbetes de seus dicionários.

Os critérios escolhidos pelos professores Mura para construir os verbetes estão fundamentalmente ligados à sua cultura. Os verbetes são somente descritivos, se restringindo quanto à informação taxionômica ao gênero gramatical do substantivo constante na entrada. Como já mencionado neste trabalho, os verbetes foram escritos em um português falado/compreendido e utilizado pelo povo Mura, deste modo, foram utilizadas escolhas lexicais próprias desta variante do Português. Podemos contemplar nas tabelas a seguir alguns exemplos de entradas e verbetes dos dicionários realizados pelos professores Mura.

**Tabela 2** – Dicionário Ilustrado dos Saberes Mura: Variedades de Peixes do Lago Josefa<sup>4</sup>

| <b>Entrada</b> | <b>Verbetes</b>   |
|----------------|---|
| Acará Açai     | <i>s.m. do Tupi (akará waçai). Não reimoso; de pequeno porte, mede aproximadamente 20 cm de comprimento; de coloração lilás; forma meio arredondada; corpo coberto por escamas; come de tudo.</i>   |
| Apapá          | <i>s.m. Não reimoso; de pequeno e médio porte, mede aproximadamente 1 metro de comprimento, quando adulto; de cor amarela brilhante; corpo coberto por escamas; come insetos e carne; conhecido como Olho de Plástico, Boca Fede, Peidão, Sarda ou Sardinhão.</i>   |
| Aruanã         | <i>s.f. do Tupi (arawanã). Não reimoso; de médio porte, mede aproximadamente 1 metro de comprimento; de coloração amarelada; corpo coberto por escamas grandes; com barbas; a fumaça de suas escamas servem para espantar carapanã e, também servem para fazer chá contra hemorragia; come insetos, carne e minhocas; conhecido como Lebréia, Macaco D'água, Sulamba.</i> |
| Gatinha        | <i>s.f. do Tupi (pirá pixana). Reimosa; de pequeno porte, mede aproximadamente 30 cm de comprimento; de coloração prateada; corpo coberto de escamas; come carne; perigoso por ter dentes afiados; conhecido como Peixe Gato.</i>   |
| Jaú            | <i>s.m. do Tupi (yau). Reimoso; de grande porte, mede aproximadamente 2 metros de comprimento; sua coloração é marrom escura, cabeça grande e boca enorme, ventre esbranquiçado; de couro liso; tem barbas; come carne; para alguns pode-se comer.</i>  |
| Pacu Murutinga | <i>s.m. do Tupi (paku murutinga). Reimoso; de pequeno porte, mede aproximadamente entre 10 a 20 cm de comprimento quando adulto; coloração branquicenta; tem forma arredondada; de escamas; come de tudo.</i>   |

**Fonte:** Prado, Martins e Batista (2013).

<sup>4</sup> O *Dicionário Ilustrado dos Saberes Mura: Variedades de Peixes do Lago Josefa* e o *Dicionário de Conhecimentos Tradicionais das Plantas Mediciniais da Aldeia Gapenu* foram os únicos dicionários que utilizaram em seus verbetes a informação etimológica das entradas, em grande parte, provenientes do Tupi. Para tanto, os grupos em questão contaram com a ajuda do professor da disciplina “Língua Indígena”.

Conforme podemos contemplar na Tabela 2, as seguintes propriedades/qualidades/características criteriais se mantiveram constantes, demonstrando assim o que a cultura Mura considera importante ao caracterizar os variados tipos de peixe. Podemos observar que a primeira informação consiste se o peixe em questão é reimoso ou não reimoso. Por outro lado, a definição do que se considera um peixe pequeno (medido geralmente em centímetros), médio (cerca um metro de comprimento) e grande (cerca de dois metros de comprimento). A coloração, também é criterial, já que a descrição é por vezes pormenorizada, a forma (arredondado/meio arredondado), tipo de derme (carapaça/liso/escamas).

Por sua vez, os autores em questão, ao sistematizar os verbetes, preferiram utilizar palavras de comum acesso aos consulentes, portanto, informações específicas como “herbívoros”, “carnívoros”, “insetívoros” ou “onívoros”, vocábulos estes que foram inicialmente utilizados na elaboração deste dicionário, foram substituídos por frases como “come insetos e carne”, “come insetos, carne e minhocas”, “come carne”, “come frutas, insetos” ou “come de tudo”. Ainda se valendo da variante do português utilizada pelos Mura, foram utilizadas variações morfológicas (esbranquiçada/branquicenta) e como sinônimos da entrada “Boca Fedê”, “Peidão”. Também vem expresso no verbete se o peixe oferece algum perigo ao consulente, como em “perigoso por ter dentes afiados”; se tem mais de uma utilidade, como “espantar carapanã” ou algum fim medicinal “para fazer chá contra hemorragia”.

Por se tratar de um tipo de alimentação amplamente utilizada pelos povos amazônicos, os autores decidiram marcar apenas as exceções, ou seja, peixes não comestíveis/comestíveis para algumas comunidades Mura. Para este fim, utilizaram as frases “não se come”/“alguns comem”.

**Tabela 3** – Dicionário Ilustrado das Plantas Mediciniais da Aldeia Murutinga

| <b>Entrada</b>    | <b>Verbetes</b>   |
|-------------------|---|
| Alfavaca          | <i>s.f. pequena, cultivada em horta. As folhas servem para banho de cabeça e as sementes servem para diminuir a carne crescida no olho.</i>   |
| Capituzieiro      | <i>s.m. de médio porte, silvestre, de 8 a 10 metros de altura. Suas folhas servem para fazer xarope para curar a tosse e também serve para fazer banho contra o aborrecimento de criança.</i>             |
| Cuia Mansa        | <i>s.f. rasteira cultivada em quintal, não frutífera. Suas folhas servem para fazer banho em criança com aborrecimento.</i>   |
| Castanha Da Índia | <i>s.f. de pequeno porte, de 5 a 6 metros, cultivada, não comestível. As folhas e as sementes servem para a mordida de cobra e de outros insetos venenosos.</i>   |
| Catinga De Mulata | <i>s.f. de pequeno porte, não frutífera, cultivada em quintal. O sumo das folhas é indicado para o tratamento de cólica umbilical da criança. A água da sua flor serve para pressão alta.</i>             |
| Corama.           | <i>s.f. de pequeno porte, cultivada em quintal. Suas flores servem para fazer chá, para mulheres em ameaças da perda de bebê, a folha serve para o tratamento de tosse junto com outros ingredientes.</i> |

**Fonte:** Braga Filho e Braga (2013).

Por se tratar de um dicionário de plantas medicinais foram consideradas como criterialidade: “porte” (pequeno/médio/rasteira); se a planta pode ser cultivada, indicação do local de cultivo, como “horta/quintal” ou se esta é silvestre. Há indicação também das partes da planta (raiz/folha/semente/flores) e como estas partes devem ser utilizadas (chá/sumo/etc.); por fim a explicação de seu fim medicinal. Escolhas lexicais como “carne crescida no olho” ou “para fazer banho contra o aborrecimento de criança”, “mulheres em ameaças da perda de bebê”, “cólica umbilical da criança” deixam transparecer, como Isquierdo (2001) aponta, elementos significativos da visão de mundo deste grupo linguístico, que até o

momento, não havia sido nem sistematizada, nem estudada. Por uma questão de tempo, já que havia data certa para entrega do TCC, não foi possível aos autores realizarem um dicionário com busca reversa, para que o consulente examinasse o dicionário procurando tanto as plantas medicinais como as “enfermidades” por elas curadas.

**Tabela 4** – Minidicionário das Plantas existentes na aldeia Iguapenu

| <b>Entrada</b>     | <b>Verbetes</b>  |
|--------------------|--|
| Bre-u-ei-ro.       | <i>s. m. De grande porte, silvestre, nativo do solo de terra firme, não venenoso, se propaga por semente, não possui leite, resinoso, madeira de lei.</i>                            |
| Ca-pi-ta-ri-zei-ro | <i>s. m. De médio porte, silvestre, nativo do solo de várzea, não venenoso, se propaga por semente, não possui leite.</i>  |
| Ca-fe-ra-na        | <i>s. f. De médio porte, silvestre, nativa do solo de várzea, venenosa, se propaga por semente, não possui leite.</i>  |
| Ca-nei-ra          | <i>s. f. De pequeno porte, cultivada, nativa do solo de terra firme, não venenosa, se propaga por caule, não possui leite, parte comestível é o caule, da família das palmeiras.</i> |
| Ca-ca-u-ra-na      | <i>s. f. De médio porte, silvestre, nativa do solo de terra firme, não venenosa, se propaga por semente, não possui leite, parte comestível é o fruto cujo sabor é azedo.</i>        |
| Ca-ba-cei-ra       | <i>s. f. De pequeno porte, silvestre, trepadeira, rasteira, nativa do solo de várzea, venenosa, se propaga por semente, não possui leite.</i>  |

**Fonte:** Rodrigues e Hiamuth (2013).

Conforme podemos observar nos exemplos da Tabela 4, o “porte” (pequeno/médio/grande) foi considerado um critério importante, assim como se esta é “cultivada ou silvestre”; o lugar onde se pode encontrá-la (terra firme/várzea); se é “venenosa/não venenosa”; se possui “leite”; como se propaga, caso o consulente

queira plantá-la (caule/semente) e indicação da parte comestível (caule/fruto). A proposta deste dicionário foi, segundo os autores, demonstrar e levar as crianças a conhecerem os tipos de árvores da aldeia Igapenu, muitas delas já em franca extinção, principalmente com a criação de gado/búfalo pela sociedade circundante, conforme explicado na apresentação desse dicionário:

A população da aldeia Iguapenu, já faz algum tempo, vem desmatando a nossa floresta, causando, com isso, um transtorno a nossa fauna e flora [...] vale a pena ressaltar que essa pesquisa surtiu um grande efeito, pois teve uma grande influência na vida dos alunos e das pessoas que foram entrevistadas. A partir dela, os alunos e os entrevistados puderam entender a grande variedade de plantas existente na aldeia, sua classificação [...]. (RODRIGUES; HIAMUTH, 2013, p. 2).

## **Conclusão**

A experiência que os professores Mura tiveram ao longo de dois anos ao confeccionar seus dicionários e, ao envolver toda a comunidade e escola neste processo, promoveu resultados muito positivos. As escolas indígenas Mura, infelizmente, não contam com material didático especial que leve em conta a cultura desta etnia. O acervo das bibliotecas escolares é constituído de livros enviados pelo município e/ou estado. Deste modo, houve a preocupação de se fazer cópias dos nove dicionários para as escolas Mura de todos os professores participantes da Licenciatura, portanto, os professores das mais diversas áreas podem utilizar os dicionários em suas aulas. Infelizmente, os dicionários não foram ainda publicados, mas espera-se que futuramente, após algumas correções, isto possa ocorrer. Podemos notar que a grande maioria dos dicionários de línguas/povos indígenas não é realizada pelo próprio povo, mas por linguístas/lexicógrafos. A experiência de confecção dos dicionários para o povo Mura demonstrou que esta experiência pode e talvez deva ser feita pelo próprio povo, para que este possa decidir quais critérios utilizar em seus verbetes. Ao finalizar o trabalho, pode-

mos perceber que o processo de confecção dos dicionários envolvendo a coletividade pode promover uma experiência única e, no caso do povo Mura, um processo que, mais uma vez, reafirmou sua identidade étnica como povo indígena.

## REFERÊNCIAS

AMOROSO, M. **Mura**. 2009. Disponível em: <<http://pib.socioambiental.org/pt/povo/mura>>. Acesso em: 09 jun. 2011. Não paginado.

BRAGA FILHO, A.; BRAGA, V. **Dicionário Ilustrado das Plantas Medicinais da Aldeia Murutinga**. 2013. 46f. Trabalho de Conclusão de Curso (Licenciatura Específica para Formação de Professores Indígenas) - Faculdade de Educação, Universidade Federal do Amazonas, Autazes, 2013.

FERREIRA, V. R. S. **Estudo Lexical da Língua Matis**: Subsídios para um Dicionário Bilíngue. 2005. 211f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2005.

ISQUERDO, A. N. Vocabulário do Seringueiro: Campo Léxico da Seringa. In: OLIVEIRA, A. M. P. P. de; ISQUERDO, A. N. (Org.). **As Ciências do Léxico**: Lexicologia, Lexicografia, Terminologia. Campo Grande: Ed. da UFMS, 2001. p.89-98.

KRIEGER, M. G. O Dicionário de Língua como Potencial Instrumento didático. In: ISQUERDO, A. N.; ALVES, I. M. (Org.). **As Ciências do Léxico**: Lexicologia, Lexicografia, Terminologia. Campo Grande: Ed. da UFMS; São Paulo: Humanitas, 2007. v. 3. p. 295-311.

MORAES, A. A. A. et al. Vai Na Frente Que Tu É Peara: Estratégias do Movimento dos(as) Professores(as) Indígenas Mura, da Região de Autazes/AM, para quebrar as armadilhas da escola. In: CONGRESSO DE LEITURA DO BRASIL, 16., Campinas. **Anais...** Campinas, 2007. Disponível em: <[alb.org.br/arquivo](http://alb.org.br/arquivo)>

morto/edicoes\_antteriores/anais16/sem04pdf/sm04ss01\_02.pdf>.  
Acesso em: 9 mar. 2018.

PRADO, K. A; MARTINS, R. C; BATISTA, V. P. **Dicionário Ilustrado dos Saberes Mura:** Variedades de Peixes do Lago Josefa. 2013. 32f. Trabalho de Conclusão de Curso (Licenciatura Específica para Formação de Professores Indígenas) – Faculdade de Educação, Universidade Federal do Amazonas, Autazes, 2013.

RODRIGUES, A. P.; HIAMUTH, G. P. **Minidicionário das Plantas existentes na aldeia Iguapenu.** 2013. 50f. Trabalho de Conclusão de Curso (Licenciatura Específica para Formação de Professores Indígenas) - Faculdade de Educação, Universidade Federal do Amazonas, Autazes, 2013.



# A ELABORAÇÃO DO DICIONÁRIO CRÍTICO DA OBRA DE DOMINGOS VIEIRA FILHO<sup>1</sup>

**Conceição de Maria de Araujo RAMOS**

**José de Ribamar Mendes BEZERRA**

**Luís Henrique SERRA**

**Maria de Fátima Sopas ROCHA**

## **Introdução**

O *Dicionário Crítico da Obra de Domingos Vieira Filho* surgiu da necessidade de resgatar parte importante da produção científica sobre a linguagem e a cultura popular maranhense, produzida por um dos maiores estudiosos maranhenses do século XX, o professor Domingos Vieira Filho<sup>2</sup>, tornando assim mais acessível, à consulta de estudiosos, pesquisadores e do público em geral, uma obra fundamental para o patrimônio cultural do Estado.

---

<sup>1</sup> A pesquisa foi financiada pela Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA), processo nº 01046/2009, APP-Universal, Edital FAPEMA nº 10/2009.

<sup>2</sup> Domingos Vieira Filho nasceu em São Luís, em 1924 e faleceu, na mesma cidade, em 1981. Este estudioso do universo linguístico-cultural maranhense e membro do Instituto Histórico e Geográfico do Maranhão, da Comissão Maranhense de Folclore e da Academia Maranhense de Letras. Exerceu, dentre outros cargos/funções, os de professor da Universidade Federal do Maranhão, Diretor do Departamento de Cultura da Secretaria de Educação e Cultura do Estado do Maranhão e Presidente da Fundação Cultural do Maranhão. Escreveu mais de uma centena de obras.

Domingos Vieira Filho (D.V.F.)<sup>3</sup> é, reconhecidamente, referência indispensável para quem se proponha estudar a cultura brasileira e em especial a maranhense. Profundo conhecedor da história e do folclore maranhenses, sua obra é ampla, pela gama de conteúdos abordados, e minuciosa, pela riqueza de detalhes e profundidade da observação e da análise de temas de etnografia, folclore, literatura, história, linguagem.

Segundo Cordeiro (2003, p.16), é consenso entre aqueles que conheceram Domingos Vieira Filho que suas “[...] descobertas eram sempre examinadas e reexaminadas, testadas com maior rigor metodológico para a informação exata, abandonadas ou contraditadas aquelas que pareciam brotar da pura fantasia alheia.”

A importância de sua obra – que contribuiu para o delineamento do perfil social do brasileiro comum e em particular do maranhense, ao fotografar seu cotidiano no que concerne à linguagem e alimentação, a brinquedos, costumes, crenças, superstições, folguedos – no entanto, não pôde ainda ser adequadamente reconhecida em razão da dificuldade de acesso às obras por ele produzidas. Na realidade, o desconhecimento de seu trabalho deve-se a dois fatores fundamentais: (i) encontrar-se esparso em diversos jornais e revistas, na forma de artigos e (ii) constar de publicações (livros, plaquetas, folhetos, relatórios, catálogos) há muito tempo com edições esgotadas.

As dificuldades encontradas para consulta a essas fontes fundamentais para o estudo da cultura popular maranhense e a importância da obra de Domingos Vieira Filho, hoje rara não só pelos temas abordados como pela escassez de exemplares disponíveis para consulta, justificam a elaboração de um dicionário crítico da obra do autor.

Nos últimos anos, a Lexicografia vem despertando grande interesse entre os estudiosos da área da linguagem, mais especificamente da Linguística. No Brasil, segundo Isquierdo (2011), vive-se um momento muito propício para a Lexicografia, com

---

<sup>3</sup> Para este trabalho mantivemos a opção de mencionar o autor pelas iniciais D.V.F., usadas por ele próprio em seus textos.

uma efervescência, no seio das universidades, de pesquisas lexicográficas resultantes tanto de projetos individuais quanto de projetos coletivos.

É em meio a essa efervescência que estão surgindo dicionários dedicados a um só autor. Em se tratando do Brasil, destacam-se, dentre outras, as seguintes obras representativas desse gênero: *Gilberto Freyre de A a Z* (FONSECA, 2002) e *Dicionário Crítico Câmara Cascudo* (SILVA, M., 2003). Observamos, contudo, que grande parte dessas obras não é elaborada por pessoas com formação acadêmica específica em Lexicografia Teórica.

O projeto, cujo produto final da pesquisa foi o *Dicionário Crítico da Obra de Domingos Vieira Filho*<sup>4</sup>, teve como principais objetivos: levantar e organizar a produção científica escrita do pesquisador, com vistas à elaboração de um dicionário crítico sobre sua obra, e oferecer ao leitor um material de referência sobre o pesquisador e sobre a cultura maranhense, material esse elaborado com base nos princípios da Lexicografia.

Neste texto, além de enfocarmos as opções teórico-metodológicas que nortearam a elaboração do dicionário, descrevemos, em linhas gerais, a execução do projeto e apresentamos um exemplo de verbete.

### **Opções teórico-metodológicas norteadoras da pesquisa**

Os dicionários, objetos culturais por excelência, são uma forma de guardar não só a memória das línguas, mas também a da cultura das comunidades que falam a língua em questão, e que nela são refletidas e por meio dela são veiculadas. Assim se explica por que esse objeto cultural se faz sempre tão presente na vida de uma comunidade, sendo considerado “[...] uma das mais relevantes instituições da civilização moderna.” (BIDERMAN, 2001,

---

<sup>4</sup> O dicionário, no momento, passa por uma nova revisão, já que a reabertura da Biblioteca Pública Benedito Leite tornou possível a retomada da pesquisa em seus acervos e assim tivemos acesso a alguns artigos de jornal que não haviam sido localizados anteriormente. A obra será publicada pela EDUFMA, editora da Universidade Federal do Maranhão.

p.17), por ter entre outras funções, no âmbito das sociedades, a de informar.

É a Lexicografia a disciplina que se ocupa dos dicionários. Como estes, é também uma atividade antiga que tem acompanhando a própria história do homem. Vale lembrar que os dicionários encontram seus precursores nos glossários latinos, e que a **arte de fazer dicionários** foi, ao longo do tempo, distanciando-se do empirismo, refinando suas questões, redimensionando-se e tornando-se, assim, mais complexa.

Inserida, segundo alguns autores (cf. PONTES, 2009), no âmbito da Linguística Aplicada, a Lexicografia se volta para questões de natureza teórica e prática que subsidiam a produção de dicionários. Nessa perspectiva, a Lexicografia pode ser entendida tanto pelo viés da prática, o que leva à postulação de uma Lexicografia Prática, como pelo viés da teoria, o que dá origem ao aparecimento de uma Lexicografia Teórica ou Metalexigrafia. Prática e teoria levam a ocupações distintas no interior da Lexicografia: à Lexicografia Prática cabe a elaboração de dicionários, enquanto a Metalexigrafia se ocupa do estudo dos dicionários, isto é, de suas formas, estruturas e usos, de sua crítica, história e papel social, bem como das teorias e metodologias lexicográficas. Nesse sentido, a Metalexigrafia tem no próprio dicionário seu objeto de estudo.

No âmbito das obras lexicográficas, o estabelecimento de critérios para proposição de uma tipologia é um tema bastante discutido e complexo e bem pouco consensual (cf. SILVA, M. C., 2007). Quando a classificação das obras de referência se volta para a divisão **dicionários e enciclopédias**, ou, em outras palavras, **dicionários de língua e dicionários de coisas**, segundo Campos Souto e Pérez Pascual (2003), a fronteira entre essas duas obras se mostra mais tênue do que o desejado, já que toda enciclopédia contém um dicionário e este, por sua vez, contém elementos enciclopédicos. De acordo com Rey-Debove (1971 apud WELKER, 2004, p.45), “[...] o dicionário de língua diz o que significa o signo **leão**, ao passo que a enciclopédia diz e mostra o que é um leão.”

Tendo como pano de fundo essas ideias e considerando a finalidade da obra de referência que elaboramos – fornecer o maior número possível de informações sobre a **coisa nomeada**, que, em se tratando de nosso trabalho, são as obras de D.V.F. –, optamos pelo **dicionário enciclopédico** que, no dizer de Campos Souto e Pérez Pascual (2003, p.61, tradução nossa), “[...] é um exemplar híbrido, um gênero misto que reúne parcialmente as características do dicionário e da enciclopédia.”<sup>5</sup>

Assim, a obra elaborada apresenta a macroestrutura de um dicionário, mas também reúne elementos particulares da enciclopédia, como por exemplo, nomenclatura essencialmente nominal com a inclusão de nomes próprios e maiores informações sobre o assunto focado, dentro de um artigo enciclopédico que, segundo Correia (2009, p.27), apresenta “[...] estrutura própria dos textos *stricto sensu*, isto é, com partes que constituem a introdução, o desenvolvimento e a conclusão, e [que] faz uso dos mecanismos que lhe garantem coerência e coesão interna.”

## **Procedimentos metodológicos**

Uma investigação que tenciona levantar e organizar a vasta produção científica de um autor tem de utilizar uma metodologia que dê conta do que caracteriza sua fecundidade – a diversidade de temas. Para tanto, temos de começar por verificar quais são as grandes linhas temáticas de sua produção para que possamos coletar nas fontes impressas – jornais, revistas e livros – o *corpus* de referência. O marco temporal da pesquisa foi o século XX, entre as décadas de 40 e 80, período de produção e/ou publicação do autor investigado.

A pesquisa em arquivos e bibliotecas foi essencial para este projeto. Buscamos as fontes impressas em arquivos e acervos: (i) no âmbito do Maranhão, Biblioteca Pública Benedito Leite, Academia Maranhense de Letras, Instituto Histórico e Geográfico

---

<sup>5</sup> Original: “[...] es un ejemplar híbrido, un género mixto que reúne parcialmente las características del diccionario y de la enciclopedia.” (CAMPOS SOUTO; PÉREZ PASCUAL, 2003, p.61).

do Maranhão, Arquivo Público do Estado, Casa de Cultura Josué Montello, Centro de Cultura Popular Domingos Vieira Filho, Convento das Mercês, e (ii) fora do Estado, Biblioteca Nacional, Museu do Folclore, Biblioteca Mário de Andrade, Biblioteca da Universidade de Campinas (UNICAMP), Biblioteca da Faculdade de Letras da Universidade Federal do Rio de Janeiro (UFRJ), Biblioteca Pública do Estado do Ceará, Biblioteca Amadeu Amaral, Arquivo do Estado de São Paulo, entre outros.

A biblioteca particular do professor Domingos Vieira Filho, a que tivemos acesso graças ao apoio que obtivemos de sua família, em particular de sua filha, Dra. Flávia Teresa Viveiros Vieira, e de sua esposa, dona Ivete Viveiros Vieira, constituiu-se, também, em uma importante fonte de pesquisa.

Seguindo os procedimentos metodológicos, conseguimos formar um *corpus* significativo para constituição da macroestrutura do dicionário, como se pode observar nos quadros 1 e 2, a seguir. Esse material foi todo catalogado e encontra-se arquivado no Projeto Atlas Linguístico do Maranhão – ALiMA, constituindo um Banco de Dados. Vale destacar que, como uma parte significativa dos textos não está microfilmada nem pôde ser fotocopiada, em virtude de seu estado de conservação, optamos por fotografar as obras e convertê-las em texto de *word*, com o uso do programa ABBYY FineReader11.

**Quadro 1** – Banco de Dados

| BANCO DE DADOS DOMINGOS VIEIRA FILHO |                  |                |            |                    |  |   |
|--------------------------------------|------------------|----------------|------------|--------------------|--|---|
| TÍTULO                               | PERIÓDICO        | LOCAL          | DATA       | SUPORTE            | LOCALIZAÇÃO  | OBSERVAÇÕES   |
| MESTRE NINA RODRIGUES                | JORNAL DO BRASIL | RIO DE JANEIRO | 23/06/1957 | MICROFILME         | BN/MF<br>PRc-SPR 00009<br>Digitalizado no Temático do MF | Publicado em duas colunas na página de rosto do 5º caderno (ao final, 2 linhas ilegíveis). Conclui a matéria na 2ª página do 5º caderno com o mesmo título. |
| SOBRE OS PAUS DE UMA JANGADA         | CORREIO DO CEARÁ | FORTALEZA      | 31/08/1963 | REVISTA (ORIGINAL) | BN/MF<br>5-416,05,15<br>Digitalizado no Temático do MF   | Publicado no Caderno B, em 4 colunas, inicia na 2ª segunda página e conclui na 6ª. Ótimo estado de conservação.   |

| BANCO DE DADOS DOMINGOS VIEIRA FILHO |                    |                |            |            |   |  |
|--------------------------------------|--------------------|----------------|------------|------------|---|--|
| TÍTULO                               | PERIÓDICO          | LOCAL          | DATA       | SUORTE     | LOCALIZAÇÃO   | OBSERVAÇÕES  |
| O NEGRO NO FOLCLORE                  | CORREIO PAULISTANO | SÃO PAULO      | 08/06/1951 | MICROFILME | BN PR SPR 00140<br>Encontrado o artigo datado de 21/12/1951,    | Verificado o microfilme do período e não encontrada a matéria.   |
| A FESTA DO DIVINO EM SÃO LUÍS        | CORREIO DA MANHÃ   | RIO DE JANEIRO | 02/11/1951 | MICROFILME | BN/AEL/MF<br>SPR 00130(1-522)<br>Digitalizado no Temático do MF | Publicado no 2º segundo caderno, dessa data, em duas colunas, mas estão ilegíveis. AEL – UNICAMP (fotografado – pendrive). |

**Fonte:** Elaboração própria.

### Quadro 2 – Suporte Material e Modos de Difusão da Obra

| LIVROS | JORNAIS | REVISTAS | OUTROS<br>(CATÁLOGOS, PLAQUETAS) |
|--------|---------|----------|----------------------------------|
| 12     | 80      | 31       | 09                               |

**Fonte:** Elaboração própria.

Depois de realizada a recolha dos textos de D.F.V. foi feito o registro dos dados em fichas eletrônicas, com as seguintes informações sobre a obra: título, suporte material, referência completa, instituição em que foi localizada, pesquisador responsável, data da coleta e observações de caráter geral.

O passo seguinte consistiu em definir a macro e a microestrutura do dicionário. A macroestrutura está composta pelos seguintes elementos: sumário, prefácio, dois dedos de prosa com o leitor (seção em que explicamos em que consiste o dicionário e o porquê de sua elaboração, nossas opções teórico-metodológicas e agradecimentos), guia do usuário, lista de abreviaturas e siglas, nomenclatura, referências e apresentação da equipe de elaboradores do dicionário.

Com relação à microestrutura, o verbete está assim constituído: entrada+ numeração que indica a referência da obra-entrada + informações sobre a fonte da obra-entrada + resenha crítica da obra objeto do verbete + nota enciclopédica + remissiva.

A entrada corresponde ao título da obra de D.V.F. em análise. Está registrada com todas as letras maiúsculas e com o determinante posposto e separado por vírgula e vem grafada em negrito.

Após a entrada, encontra-se a numeração que indica a referência da obra-entrada, isto é, o número, entre colchetes e em negrito, remete o consulente à seção de referências da obra de D.V.F., para lá obter o conjunto de elementos descritivos da obra, que permite sua identificação individual.

Na sequência, encontram-se as informações sobre a fonte da obra-entrada: toda obra é seguida da indicação da data de sua publicação e de seu suporte material.

A resenha crítica da obra objeto do verbete é o elemento fundamental para o conhecimento e a compreensão da obra-entrada. Os comentários dos redatores dos verbetes são ilustrados com trechos extraídos da obra-entrada ou de outra obra de D.V.F. ou ainda de um outro autor que igualmente exemplifique o que foi dito. Quando a citação é extraída da própria obra resenhada, indicamos, entre parênteses, apenas o número da página; quando se tratar de outra obra do autor, informamos o ano de publicação e a página; para as citações de outros autores, utilizamos o sistema autor-data.

A nota enciclopédica contém informações que completam a resenha da obra e ajudam o consulente a ampliar seu conhecimento/compreensão acerca do assunto focado no verbete. Essas notas são apresentadas na forma de boxes.

A remissiva envia o consulente a outro verbete para lá conferir informações complementares, análogas ou contrastivas. A remissão de um verbete a outro é feita por meio da indicação da obra-entrada acompanhada do número correspondente a sua referência.

O passo seguinte correspondeu à etapa de elaboração dos verbetes.

Para efeito de ilustração, apresentamos parte de um dos verbetes do dicionário – **ADVINHAS POPULARES**<sup>6</sup>.

**ADVINHAS POPULARES** [2]. Este artigo compõe o conjunto de trabalhos dos números 9 e 10 (setembro e dezembro) do

---

<sup>6</sup> Optamos por conservar a grafia do original.

*Boletim Trimestral da Comissão Catarinense de Folclore\**, publicado em 1951.

Nele, D.V.F. arrolou elementos da cultura popular que, segundo ele próprio afirma ao rememorar a infância, fazem parte do universo de coisas que mais nos prendem a atenção, quando criança – as adivinhas populares:

De permeio com estórias de Tia Onça, Compadre Macaco e Mestre Coelho, lá vinham as advinhas, que chamávamos advinhações, umas simples, a solução aflorando à primeira vista, outras difíceis, complicadas, verdadeiros enigmas a espera de Édipo para resolvê-los. (p.187).

Compõem esse estudo, de **persistente traço folklórico**, “[...] adivinhações nascidas em terras maranhenses, da imaginação popular e expressando idéias ligadas a traços de cultura nativos.” (p.188). As advinhas são resultado de um paciente trabalho de recolha que, segundo explica o próprio D.V.F., foi realizado durante visitas a pontos distantes de São Luís, durante seus passeios dominicais, quando então colhia adivinhações como se estivesse **num rico jardim**.

Esse trabalho de recolha surge do engajamento do autor no movimento em defesa das tradições populares nacionais e em especial das maranhenses. Ele tem consciência de que “As advinhas, bem ou mal feitas, em prosa ou em verso, simples ou complexas [...]” (p.187), são universais, e de que no Brasil, apesar do interesse de Alcides Bezerra e de Amadeu Amaral\*\* pelo tema,

Nunca tivemos, entretanto, um estudioso que se detivesse em pesquisas profundas, que rastreasse as origens no labirinto das vias universais, acompanhando **pari-passu** o itinerário desta ou daquela, numa coleção de milhares, com fizeram Fernand Caballero e Machado y Alvarez na Espanha e Eisen na Estônia. (p.187-188, grifo do autor).

Na amostra que apresenta no artigo, DVF reúne: 76 adivinhas maranhenses, incluindo-se neste total uma variante; 11 adivinhas variantes das catalogadas no Maranhão, sendo algumas delas registradas na Paraíba e outras em Natal e outras; 2 colhidas em outros países e 2 extraídas de obras literárias.

\* A Comissão Catarinense de Folclore existe há mais de 60 anos, com sede em Florianópolis, em Santa Catarina. Regularmente publica um boletim que tem como objetivo difundir e preservar o folclore brasileiro.

\*\* DVF faz referência aos trabalhos “Demopsicologia: adivinhas”, de Alcides Bezerra, publicado em 1913, no volume 4 da *Revista do Instituto Histórico e Geográfico Paraibano*, e “Tradições populares: adivinhas e outros problemas”, de Amadeu Amaral, publicado em 1931, no número 120 da *Revista da Academia Brasileira de Letras*.

### **Considerações finais**

O *Dicionário Crítico da Obra de Domingos Vieira Filho*, com mais de cento e trinta verbetes, promoverá o acesso de estudiosos e do público em geral à ampla literatura sobre a cultura popular brasileira, com ênfase na cultura maranhense. Além disso, proporcionará às gerações mais jovens o contato com a obra de um pesquisador que teve como preocupação primeira a preservação e o registro da memória cultural.

Como uma obra de referência, o Dicionário será um guia, um mapa que possibilitará um conhecimento mais amplo do conjunto de uma obra fundamental para todos que desejem conhecer, entender, estudar, examinar o patrimônio linguístico-cultural maranhense e por ele navegar.

### **REFERÊNCIAS**

BIDERMAN, M. T. C. As ciências do léxico. In: OLIVEIRA, A. M. P. P. de; ISQUERDO, A. N. (Org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. 2.ed. Campo Grande: Ed. da UFMS, 2001. v.1, p.13-22.

CAMPOS SOUTO, M.; PÉREZ PASCUAL, J. I. El diccionario y otros productos lexicográficos. In: MEDINA GUERRA, A. M. (Coord.). **Lexicografía española**. Barcelona: Ariel, 2003. p.53-78.

CORDEIRO, J. M. Domingos Vieira Filho: um amante da cultura popular maranhense. **Boletim de Folclore**, São Luís, n.25, p.16-17, jun. 2003.

CORREIA, M. **Os dicionários portugueses**. Lisboa: Editorial Caminho, 2009.

FONSECA, E. N. da. **Gilberto Freyre de A a Z**: referências essenciais à sua vida e obra. Rio de Janeiro: Zé Mario Editor, 2002.

ISQUERDO, M. A. N. Os estudos lexicográficos no Brasil: um percurso histórico. In: CARDOSO, S.; MEJRI, S.; MOTA, J. (Org.). **Os dicionários**: fontes, métodos e novas tecnologias. Salvador: Vento Leste, 2011, p.113-144.

PONTES, A. L. **Dicionário para uso escolar**: o que é como se lê. Fortaleza: EdUECE, 2009.

SILVA, M. (Org.). **Dicionário crítico Câmara Cascudo**. São Paulo: Perspectiva; FFLCH/USP: Fapesp; Natal: EDUFRN: Fundação José Augusto, 2003.

SILVA, M. C. P. da. Para uma tipologia geral de obras lexicográficas. In: ISQUERDO, A. N.; ALVES, I. M. (Org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. Campo Grande: Ed. da UFMS; São Paulo: Humanistas, 2007. p.283-293.

WELKER, H. A. **Dicionários**: uma pequena introdução à lexicografia. Brasília: Thesaurus, 2004.



# BRINCADEIRAS INFANTIS E O LÉXICO REGIONAL: UM ESTUDO DE DADOS GEOLINGUÍSTICOS

Laísa Bauermeister STELO  
Aparecida Negri ISQUERDO

## Introdução

A nomeação da realidade pode ser considerada o primeiro passo do homem no conhecimento do universo e, para nomear, o indivíduo se vale do léxico, nível linguístico que melhor evidencia a realidade do falante, visto que é a partir dele que os indivíduos registram as situações cotidianas e revelam sua identidade, tanto social como culturalmente. A Lexicologia, área da ciência que estuda o léxico, possui como objetos de investigação a análise da palavra, a categorização e a estruturação lexical (BIDERMAN, 2001).

A partir das escolhas lexicais dos falantes é delineada uma norma local, isto é, a “[...] variedade linguística que determinados grupos sociais adotam como referência em seus usos da língua.” (CASTILHO, 2010, p.686) e, quando a norma é de um grupo específico, é denominada norma regional:

O falar regional é o meio de expressão oral de um grupo humano, de uma aldeia, de um município, que tem consciência de certas particularidades linguísticas que o distinguem de

um grupo vizinho. O que caracteriza o falar é, de um lado, as suas discrepâncias da língua comum e, de outro, essa consciência que une os membros do grupo e os distingue dos grupos vizinhos. Podemos assim conceituá-lo como um feixe dos traços distintivos. (SILVA NETO, 1957).

Cada falante nomeia objetos que estejam de acordo com o seu universo. O morador de um país com diversas árvores e plantas e que conviva com elas diariamente, por exemplo, saberá não só dar nome a elas, mas também a sua utilidade. Um visitante que esteja nesse mesmo lugar terá que se adaptar a essa norma regional, para a sua própria sobrevivência social e física.

Ao deparar-se com uma língua diferente da sua, o ser humano sabe que não se trata da realidade linguística a qual pertence, pois conhece instintivamente os fonemas que compõe sua própria linguagem. Este trabalho almeja observar como se configura o léxico e suas múltiplas facetas no português brasileiro, bem como as diferentes designações para **bola de gude** na região Centro-Oeste.

## Metodologia

Este trabalho vincula-se ao projeto *Tesouro do Léxico Patrimonial Galego e Português: Brasil*, de natureza internacional, sediado na Universidade de Santiago de Compostela que reúne três comitês de pesquisa: o galego, o português e o brasileiro. No Brasil é coordenado pela Universidade Federal do Rio de Janeiro, sob a responsabilidade da Profa. Dra. Sílvia F. Brandão. A UFMS participa desse projeto tendo sob sua responsabilidade os trabalhos relativos aos documentos da região Centro-Oeste. O projeto *Tesouro Léxico* está reunindo, num *corpus* único, o léxico das línguas galegas e portuguesa, registrado em obras lexicográficas e em atlas linguísticos produzidos nas diferentes fases da história do galego e do português que, por se encontrarem dispersos e, de um modo geral, inéditos, se tornam de difícil acesso. No caso da UFMS, a equipe de pesquisa está reunindo obras sobre o léxico recolhido da língua oral, nas diferentes fases da história da

região Centro-Oeste, em áreas claramente delimitadas pela pesquisa (localidades dos Estados de Mato Grosso do Sul, de Mato Grosso e de Goiás) que contenham algum produto lexicográfico (glossários, vocabulários, dicionários) ou um atlas linguístico com dados lexicais mapeados. Tendo em vista a riqueza e a diversidade do *corpus*, a base de dados do *Tesouro Léxico* permite diferentes perspectivas de estudo, dentre outras, o estudo comparativo de dados a partir de diferentes obras que integram a base de dados do projeto.

Para este estudo, foram selecionadas as seguintes obras que fazem parte do acervo desse projeto: *ALMS – Atlas Linguístico de Mato Grosso do Sul* (OLIVEIRA, 2007); *ALiPP – Atlas Linguístico do município de Ponta Porã – MS* (REIS, 2006); *ALMESEMT – Atlas Linguísticos da Mesorregião Sudeste de Mato Grosso* (CUBA, 2009) e *ALiCoLa – Atlas linguístico de Corumbá e Ladário* (ALENCAR, 2013). Considerando o fator domínio de abrangência (CARDOSO, 2010), esses atlas assim se classificam: *ALMS* (atlas estadual); *ALMESEMT* (atlas regional); *ALiPP* (atlas municipal) e *ALiCoLa* (atlas municipal).

O *ALMS – Atlas Linguístico de Mato Grosso do Sul* – foi organizado por Dercir Pedro de Oliveira (2007) e teve como principal objetivo conhecer as características da língua falada em Mato Grosso do Sul. Para isso, o atlas contou com uma extensa rede de pontos com trinta e duas localidades.

O *ALMESEMT – Atlas Linguístico da Mesorregião Sudeste de Mato Grosso* – foi produto da dissertação de Mestrado de Marigilda Antonio Cuba, defendida em 2009, no Programa Estudos de Linguagens da UFMS/Campo Grande. O trabalho teve como objetivo mapear fatos linguísticos da região estudada. A rede de pontos do *ALMESEMT* abrangeu oito localidades.

O *ALiPP – Atlas Linguístico de Ponta Porã* – foi produzido por Regiane Coelho Pereira Reis, também como dissertação de Mestrado, defendida em 2006, no Programa de Pós-Graduação em Letras da UFMS/Campus de Três Lagoas. O atlas apresenta um diferencial em relação aos demais, pois mapeou as variantes lexicais documentadas por três línguas faladas na faixa de fronteira do

Brasil com o Paraguai: português, espanhol e guarani. O ALiPP contemplou uma rede de oito pontos.

O *ALiCoLa – Atlas Linguístico de Corumbá e Ladário* – foi desenvolvido por Beatriz Alencar, como dissertação do Mestrado que foi defendida em 2013 no Programa de Estudos de Linguagens da UFMS. O ALiCoLa coletou dados de cinco localidades, duas sedes de municípios (Corumbá e Ladário) e três distritos de Corumbá (Coimbra, Porto Esperança e Albuquerque).

Inicialmente foram catalogadas as cartas lexicais relativas à área semântica de **brincadeiras infantis** dos quatro atlas selecionados para o estudo. Foram elencadas as seguintes cartas comuns nos atlas selecionados: **Bola de gude**: 4 atlas (ALMS, ALMESEMT, ALiPP e ALiCoLa); **Pipa**: 4 atlas (ALMS, ALMESEMT, ALiPP e ALiCoLa); *Estilingue*: 3 atlas (ALMS, ALMESEMT e ALiCoLa); **Cobra-cega**: 3 atlas (ALMS, ALMESEMT e ALiCoLa); **Amarelinha**: 2 atlas (ALMESEMT e ALiPP); **Peteca**: 2 atlas (ALiPP e ALiCoLa).

Para este trabalho foram tomados como *corpus* os dados registrados em quatro cartas que mapeiam designações para o conceito “coisinhas redondas de vidro com que os meninos gostam de brincar”: ALMS (QSL 0461.a); ALMESEMT (QSL 109); ALiPP(QSL 236) e ALiCoLa (QSL 506).

### **Análise dos dados**

A consulta às cartas selecionadas resultou no levantamento de seis variantes lexicais (bolita, birola/biroca/biloca, bolinha de gude, bolinha de vidro, burca e bulica), assim dispostas nos quatro atlas consultados:

**Quadro 1** – Designações para bola de gude nos atlas pesquisados

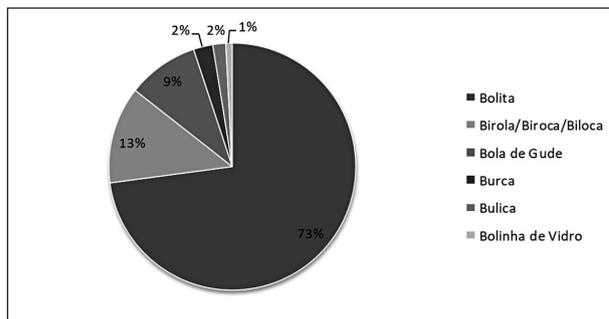
| <b>Variantes lexicais</b> | <b>ALMS</b> | <b>ALMESEMT</b> | <b>ALiPP</b> | <b>ALiCoLa</b> |
|---------------------------|-------------|-----------------|--------------|----------------|
| bolita                    | X           | X               | X            | X              |
| birola/biroca/biloca      | X           |                 |              |                |
| bolinha de gude           | X           |                 | X            | X              |
| bolinha de vidro          | X           |                 |              |                |
| burca                     | X           |                 |              |                |
| bulica                    | X           |                 |              |                |

**Fonte:** Elaboração própria.

Nota-se pelos dados registrados no quadro que **bolita** foi a única variante comum nos quatro atlas, evidenciando, assim, que essa unidade lexical já está arraigada na fala dos habitantes dos dois Estados da região Centro-Oeste, cuja língua foi documentada pelos atlas pesquisados. O designativo **bolinha de gude** foi o segundo mais produtivo, estando registrado no ALMS, no ALiPP e no ALiCoLA.

Com o objetivo de demonstrar a distribuição das variantes lexicais catalogadas e a respectiva produtividade de cada uma delas, na sequência, os gráficos 1, 2, 3 e 4 demonstram o comportamento dessas variantes no âmbito de cada atlas consultado, iniciando pelo atlas regional *Atlas Linguístico de Mato Grosso do Sul*.

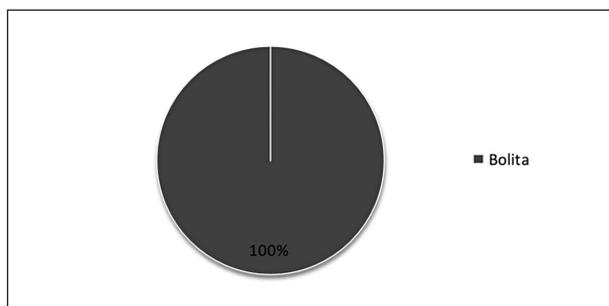
**Gráfico 1** – Designações para “bola de gude” no ALMS



**Fonte:** Dados da pesquisa.

Nota-se que, no Estado de Mato Grosso do Sul, a variante léxica **bolita** lidera com 73% de ocorrência, seguida dos itens léxicos **birola/biroca/biloca** que figuram apenas do ALMS, computando 13% das respostas. Na sequência situam-se **bola de gude**, com 9%, **burca** e **bulica** com 2% cada e **bolinha de vidro** com 1% de produtividade. A seguir o gráfico explicita a produtividade das variantes apuradas no *Atlas Linguístico da Mesorregião Sudeste de Mato Grosso*.

**Gráfico 2** – Designações para “bola de gude” no ALMESEMT

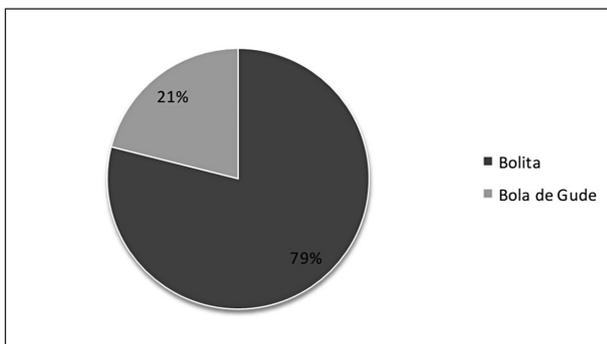


**Fonte:** Dados da pesquisa.

O Gráfico 2 atesta a força da variante lexical **bolita** na Mesorregião Sudeste de Mato Grosso, que é dominante no voca-

bulário dos moradores dessa região. Também o *Atlas Linguístico de Ponta Porã* confirma a supremacia dessa variante na fronteira do Mato Grosso do Sul com o Paraguai (Gráfico 3).

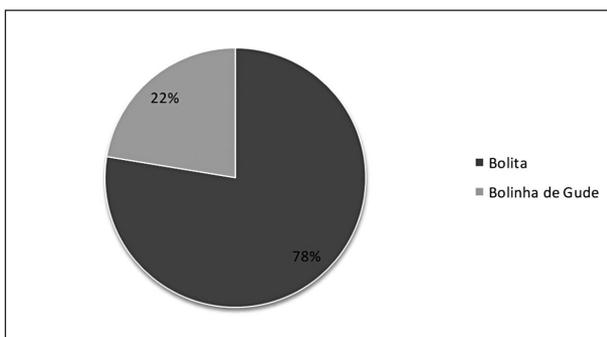
**Gráfico 3** – Designações para “bola de gude” no ALiPP



**Fonte:** Dados da pesquisa.

Esse gráfico demonstra que a maior quantidade dos entrevistados do município de Ponta Porã faz uso dos itens lexicais **bolita** (79%) e **bola de gude** (21%) para nomear o brinquedo em questão. Realidade similar foi evidenciada pelo *Atlas Linguístico de Corumbá e Ladário* (Gráfico 4).

**Gráfico 4** – Designações para “bola de gude” no ALiCoLa



**Fonte:** Dados da pesquisa.

O Gráfico 4 demonstra que na fronteira de Mato Grosso do Sul com a Bolívia ocorre o mesmo fenômeno observado na fronteira com o Paraguai, à medida que evidencia o domínio do item lexical **bolita** (78%), seguido de **bolinha de gude** (22%) na fala dos habitantes de Corumbá e Ladário.

Na verdade, a situação de fronteira dos Estados de Mato Grosso (com a Bolívia) e de Mato Grosso do Sul (com a Bolívia ao norte e com o Paraguai) favorece a interinfluência cultural e linguística, em especial nas cidades localizadas na fronteira seca do Brasil com esses países.

O léxico como sistema aberto e em expansão sempre está suscetível a abrigar novas criações, sejam elas próprias da língua, sejam elas decorrentes da situação de contato, como é o caso da variante lexical **bolita**, um empréstimo ao espanhol assimilado pela população. Há também a variação desse item léxico no espanhol latino-americano, *balita*, muito utilizado no Paraguai e na Argentina, para referir-se ao mesmo jogo. Na Argentina há, inclusive, campeonatos de *balita* e outros jogos em praças que chegam a reunir 120 meninos.

Na sequência, o Quadro 3 registra a dicionarização das variantes em estudo, cotejando definições de três dicionários, um do espanhol - *Diccionario de la lengua española* – RAE e dois do português - *Novo Dicionário Aurélio da Língua Portuguesa* (FERREIRA, 2004) e *Dicionário Eletrônico Houaiss da Língua Portuguesa* (HOUAISS, 2001).

**Quadro 3** – Dicionarização das variantes para **bola de gude**

| Variantes lexicais           | RAE   | Houaiss (2001)   | Ferreira (2004)  |
|------------------------------|---|--|--|
| Bolita                       | <p>1. f. Juego de niños que se hace con bolas pequeñas de barro, vidrio u otra materia dura. U. m. en pl.</p> <p>2. f. Cada una de estas bolas.</p> | <p>Rubrica: ludologia.<br/>Regionalismo: Rio Grande do Sul.<br/>m.q. <b>gude</b></p>   | <p>[Do esp. plat. <i>bolita</i>.]<br/>Substantivo feminino.<br/>1. Bras. RS V. <i>gude</i>.</p>  |
| Birola/<br>Biroca/<br>Biloca |   | <p><b>Biloca</b> –<br/>Rubrica: ludologia.<br/>Regionalismo: Goiás.<br/><b>1</b> gude jogado ger. com cinco buracos<br/><b>2</b> cada um desses buracos</p>  | <p><b>Biloca</b> –<br/>1. Bras. GO. Jogo em que os contendores, usando o polegar e o indicador, procuram, como num piparote, atirar botões num pequeno buraco.</p>   |
| Bolinha de Gude              |   | <p><b>Gude</b> –<br/>Rubrica: ludologia.<br/>Regionalismo: Brasil.<br/><b>1</b> jogo infantil com bolinhas de vidro que, num percurso de ida e volta, devem entrar em três buracos dispostos em linha reta, saindo vencedora a criança que chegar primeiro ao buraco inicial<br/><b>2</b> bolinha us. nesse jogo<br/><b>3</b> qualquer outro jogo infantil com bolinhas de vidro</p> | <p><b>Gude</b><br/>1. Jogo infantil em que se procura fazer entrar em três buracos bolinhas de vidro, ou os carocinhos pretos do fruto do saboeiro, ganhando o jogador que chega primeiramente de volta ao primeiro buraco: “Raul brincava sossegado com as bolas de <b>gude</b>” (Lia Correia Dutra, <i>Navio sem Porto</i>, p. 171).<br/>2. Bolinhas de vidro, etc., usadas nesse jogo. [Sin. ger.: <i>baleba, bilosca, biosca, bolita, búraca, búrica, cabiçulinha, fro, peteca, piroasca, ximbra</i> e (lus.) <i>berlinde e bute</i>.]</p> |

**Fonte:** Elaboração própria.

Pode-se observar que o único item lexical registrado nos três dicionários foi **bolita**. Já as variantes lexicais **bolinha de gude**, **birola/biroca/biloca** e **bolinha de gude** foram identificados nos dois dicionários de língua portuguesa. Já os itens lexicais **burca** (variante de búrica) e **bulica** (variante de bolita) não estão dicionarizados.

## Considerações finais

Como já apontado, a unidade lexical **bolita**, pela força como é usada nos Estados de Mato Grosso do Sul e Mato Grosso, pode ser considerada um regionalismo sul-mato-grossense e mato-grossense, ao contrário de **bola de gude** que representa a norma nacional, difundida pela escola. Pode-se considerar **bolita** como um caso de empréstimo do espanhol em decorrência da fronteira desses Estados com países de língua espanhola (Paraguai e a Bolívia). No dicionário da RAE – Real Academia Espanhola (20-), há o registro de *bolita* com remissão para *canita*, unidade léxica definida como “1. juego de niños que se hace con bolas pequeñas de barro, vidrio u otra materia dura. U. m. en pl.” 2. f. Cada una de estas bolas” (BOLITA, [20-]).

Uma das possibilidades que podem ter levado os falantes dessa região a utilizarem essa unidade lexical é a busca pela expressividade da língua, uma vez que a situação de contato linguístico favorece o uso frequente do item léxico **bolita** nos dois lados da fronteira, o que contribuiu para que ela se arraigasse na norma linguística e norma local e, por isso, se sobrepujasse, em termos de uso, ao designativo **bola de gude**.

É preciso considerar também que os Estados de Mato Grosso e Mato Grosso do Sul foram palco de muitas guerras, e desde a sua conquista, nos séculos XVII, XIX e XX, houve conflitos principalmente entre os indígenas, castelhanos, portugueses, brasileiros, paraguaios e bolivianos (QUEIROZ, 2003). Essa disputa, além de causar morte e destruição, também contribuiu para a diversificação linguística e cultural de todos os povos envolvidos. Enfim, a amostra de dados lexicais, aqui analisada, confirma que o léxico da língua movimenta-se entre os eixos da conservação e da mudança (ISQUERDO, 2003), interferindo fortemente nesse processo fatores extralinguísticos, dentre os quais, os apontados ao longo deste trabalho.

## REFERÊNCIAS

- ALENCAR, B. A. **Atlas linguístico de Corumbá e Ladário**: uma descrição da língua portuguesa falada no extremo oeste de Mato Grosso do Sul. 2013. 439f. Dissertação (Mestrado em Estudos de Linguagens) – Centro de Ciências Humanas e Sociais, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2013.
- BIDERMAN, M. T. C. **Teoria linguística**: teoria lexical e linguística computacional. São Paulo: Martins Fontes, 2001.
- BOLITA. In: Real Academia Española [RAE]. **Diccionario de Lengua Española**. Madrid: RAE, [20-]. Disponível em: <<http://lema.rae.es/drae/?val=bolita>>. Acesso em: 07 mar. 2014.
- CARDOSO, S. A. **Geolinguística**: tradição e modernidade. São Paulo: Parábola, 2010.
- CASTILHO, A. T. de. **Gramática do Português Brasileiro**. São Paulo: Contexto, 2010.
- CUBA, M. A. **Esboço de um atlas lingüístico de Mato Grosso**: a língua falada na Mesorregião Sudeste. 2009. 402f. Dissertação (Mestrado em Estudos de Linguagens) – Centro de Ciências Humanas e Sociais, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2009.
- FERREIRA, A. B. de H. **Novo Dicionário Aurélio da Língua Portuguesa**. Versão 5.0. Curitiba: Positivo, 2004.
- HOUAISS, A. **Dicionário Eletrônico Houaiss da Língua Portuguesa**. Versão 1.0. Rio de Janeiro: Objetiva, 2001.
- ISQUERDO, A. N. Léxico em tempo e espaço: a questão dos regionalismos. In: MARIN, J. R.; VASCONCELOS, C. A. de. (Org.). **História, região e identidades**. Campo Grande: Ed. da UFMS, 2003. p.165-181.
- OLIVEIRA, D. P. de (Org.). **Atlas linguístico de Mato Grosso do Sul**. Campo Grande: Ed. da UFMS, 2007.

QUEIROZ, P. R. C. Temores e esperanças: o antigo sul de Mato Grosso e o Estado Nacional Brasileiro. In: MARIN, J. R.; VASCONCELOS, C. A. de. (Org.). **História, região e identidades**. Campo Grande: Ed. da UFMS, 2003. p.19-46.

REAL ACADEMIA ESPAÑOLA [RAE]. **Diccionario de Lengua Española**. Madrid, [20-]. Disponível em: <<http://dle.rae.es/?w=diccionario>>. Acesso em: 07 mar. 2014.

REIS, R. C. P. **Atlas linguístico do município de Ponta Porá-MS: um registro das línguas em contato na fronteira do Brasil com o Paraguai**. 2006. 469f. Dissertação (Mestrado em Letras) – Departamento de Educação, Universidade Federal de Mato Grosso do Sul, Três Lagoas, 2006.

SILVA NETO, S. **Guia para estudos dialectológicos**. 2.ed. Belém: Instituto Nacional de Pesquisas da Amazônia, 1957.

# **O VOCABULÁRIO RURAL DA SERRA DA CANASTRA/MG: UM ESTUDO LINGUÍSTICO NA NASCENTE DO RIO SÃO FRANCISCO**

**Gisele Aparecida RIBEIRO**

## **Introdução**

Concebemos a linguagem como uma realidade cultural completíssima, polifacetada – dada a sua diversidade, e organizada. Graças a esse sistema organizado de signos, conseguimos nos comunicar: informando, expressando nossos sentimentos, influenciando outras pessoas e entendendo o que nos dizem. Em sua plenitude funcional, a linguagem utiliza-se de palavras, constituindo a língua.

Falar uma língua significa compartilhar com outros, com povos do presente, com povos do passado e com povos do futuro aquilo que temos e o que não temos em comum. Como pessoas que pertencem a uma determinada comunidade, ou como alguém que assume a tradição idiomática da comunidade em que vive, damos, em nossa vida, continuidade a um processo sócio-histórico-cultural estabelecido por nossos antepassados. As línguas são, pois, conforme aponta Coseriu (1991, p.16, tradução nossa): “[...] técnicas históricas da linguagem e, na medida em que se encontram

estabelecidas como tradições firmes e peculiares, reconhecidas por seus próprios falantes e por falantes de outras línguas.”<sup>1</sup>

Como o homem vive em sociedade, língua e sociedade relacionam-se intimamente: uma não existe sem a outra. Se por um lado uma língua só existe se há uma comunidade que a utiliza, por outro lado um agrupamento de pessoas só será uma comunidade se tiver uma língua que possibilite a comunicação desse grupo.

A esse acúmulo de conhecimentos que advém das realizações do homem em seu grupo e que é transmitido de geração a geração dá-se o nome de **cultura**. A língua é, pois, condição fundamental para a existência de cultura e, ao mesmo tempo, recebe influência dessa cultura, uma vez que se correlaciona com as atividades sociais do homem, podendo, por isso, variar e mudar. Nas palavras de Coseriu (1991, p.32, tradução nossa), “o homem vive em um mundo linguístico que cria o mesmo como ser histórico”<sup>2</sup>. A língua é, pois, um fato social, histórico e, portanto, cultural.

Levando-se em consideração o estudo da língua coligado ao estudo da cultura, analisar o léxico de uma comunidade é revelar as práticas sociais em seu acervo de palavras; é compreender a história, as manifestações artísticas, as religiões, as atividades econômicas, os valores, etc. como sendo importantes elementos constitutivos de um grupo. É dar-se a conhecer todo o seu patrimônio sociocultural; tudo aquilo que construiu, constrói e que deixará para a posterioridade.

Esta pesquisa está inserida no GruMEL (Grupo Mineiro de Estudos do Léxico – coordenado pela Profa. Dra. Maria Cândida Trindade Costa de Seabra) que tem como objetivo desenvolver atividades de pesquisa, transferência de conhecimento e difusão de informações acerca dos estudos lexicológicos, lexicográficos, toponímicos e terminológicos. Contempla, pois, o estudo do léxico sob

---

<sup>1</sup> Original: *(Las lenguas son, en efecto,) técnicas históricas del lenguaje y, en la medida en que se hallan establecidas como tradiciones firmes y peculiares, reconocidas por sus propios hablantes y por los hablantes de otras lenguas.* (COSERIU, 1991, p.16).

<sup>2</sup> Original: *“El hombre vive en un mundo linguístico que crea el mismo como ser histórico.”* (COSERIU, 1991, p.32).

diferentes perspectivas teóricas, constituindo um espaço estimulante ao diálogo entre acadêmicos de diferentes instituições.

Em um primeiro momento apresentaremos algumas questões teóricas, após, um breve histórico do Parque Nacional da Serra da Canastra. Logo depois, será exposta uma análise parcial dos dados encontrados em algumas entrevistas já transcritas e em seguida serão feitas as considerações finais.

## **O homem, a língua e a cultura**

Como produto e instrumento da cultura, a língua é ferramenta da sociedade. Através dela os falantes de determinada comunidade se expressam e expressam seus valores, suas ideias, suas experiências e, assim, vão construindo sua sociedade, sua história.

Devido a esse caráter histórico, uma mesma língua possui divergências, entretanto essas divergências estão conectadas: no espaço (diatópica), no tempo (diacrônica), socioculturalmente (diastrática), no discurso (diafásica) – um entremeado de dialetos, níveis e estilos, como bem mostra Coseriu (1992, p.37, tradução nossa), quando afirma que “[...] uma língua histórica nunca é um único sistema, mas sim um emaranhado de – em parte – diferentes sistemas. Há diferenças desde o ponto de vista fonético, gramatical e léxico.”<sup>3</sup>

Para esse autor (1992, p.38), “[...] uma língua que não apresenta nenhuma diversidade no espaço, no tempo, na estratificação sociocultural ou no estilo é uma língua que não está viva.”

## **O homem, a língua e a sociedade**

Desde os primeiros estudos feitos acerca da linguagem, procuraram-se encontrar respostas para entender a relação língua e sociedade, haja vista que esses elementos estão intimamente ligados.

---

<sup>3</sup> Original: “Una lengua histórica nunca es un único sistema, sino un entramado de – en parte – diferentes sistemas. Hay diferencias desde el punto de vista fonético, gramatical y léxico.” (COSERIU, 1992, p.37).

Saussure, prenunciador da corrente estruturalista, reconhecia a importância de natureza etnológica, histórica e política da língua, apesar de seus estudos terem se voltado fundamentalmente para o organismo linguístico interno. Para Saussure (1969, p. 23), “[...] a língua é um fato social, no sentido de que é um sistema convencional adquirido pelos indivíduos no convívio social.”

No início do século XX, Meillet (apud ALKMIN, 2001, p.24) já demonstrava interesse pela relação língua e sociedade tratando a língua como um fato social e, em consequência disso, ampliando os estudos saussureanos, quando afirmou que

[...] a linguagem é, eminentemente, um fato social. Tem-se, frequentemente, repetido que as línguas não existem fora dos sujeitos que as falam, e, em consequência disto, não há razões para lhes atribuir uma existência autônoma, um ser particular. Esta é uma constatação óbvia, mas sem força, como a maior parte das proposições evidentes. Pois, se a realidade de uma língua não é algo de substancial, isto não significa que não seja real. Esta realidade é, ao mesmo tempo, linguística e social.

É pelo exercício da linguagem, pela utilização da língua, que o homem constrói sua relação com o outro. Benveniste (1989, p.27) diz que “é dentro da, e pela língua, que indivíduo e sociedade se determinam mutuamente.” Esse mesmo autor (BENVENISTE, 1989, p.101-102) afirma, ainda, que o homem “[...] se situa necessariamente em uma classe, seja uma classe de autoridade ou classe da produção.” e, conseqüentemente, a língua, enquanto prática humana, “[...] revela o uso particular que grupos ou classes de homens fazem dela, como, também, as diferenciações que daí resultam no interior de uma língua comum.”

### **A pesquisa léxico-histórico-cultural**

Sabe-se que o homem expõe suas ideias e seu modo de ver o mundo, principalmente, através das palavras. Assim, palavras são criadas porque o homem é um ser de novos pensamentos e palavras

também são norteadoras, pois lhe informam sobre as coisas do mundo. Por ser uma das criações humanas, a língua é parte de sua cultura; por acumular simbolicamente e transmitir o que o homem já construiu em termos mentais e materiais, a língua é, também, o grande veículo das realizações humanas. Cardoso (1988, p.231) afirma: “[...] instrumento social de comunicação, a língua existe intimamente ligada à cultura de um povo. É ao mesmo tempo elemento da cultura.”

O léxico é a área dos estudos da linguagem que melhor aborda a realidade linguística cultural e social de uma comunidade. É através da palavra que todo conhecimento adquirido em várias áreas do saber e acumulado ao longo da história de um povo é transmitido às gerações seguintes. Sendo assim, quando uma comunidade se serve de vocábulos que expressam ou tentam expressar sua maneira de ver e sentir o mundo, ela passa a constituir uma espécie de documento vivo de sua própria história como, também, de toda a sociedade e cultura que a regem.

Os estudos sobre o léxico têm se mostrado altamente importantes, pois, por meio dele conhecemos os fatores sócio-econômicos, culturais, políticos e religiosos de uma determinada sociedade.

A pesquisa lexical está inserida na Lexicologia que é um ramo da linguística que tem por objetivo o estudo científico do léxico sob diversos aspectos. Para isso, ela procura determinar a origem, a forma e o significado das palavras que constituem o léxico de uma língua bem como o seu uso na comunidade dos falantes. Assim, por meio da Lexicologia torna-se possível observar e descrever cientificamente as unidades léxicas de uma comunidade linguística.

Segundo Isquierdo (2001, p.91)

[...] o estudo do léxico regional pode fornecer, ao estudioso, dados que deixam transparecer elementos significativos relacionados à história, ao sistema de vida, à visão de mundo de um determinado grupo. Deste modo, no exame de um léxico regional, analisa-se e caracteriza-se não apenas a língua, mas também o fato cultural que nela deixa transparecer.

Já Biderman (2001, p.132) afirma que “[...] o léxico pode ser considerado como o tesouro vocabular de determinada língua. Ele inclui a nomenclatura de todos os conhecimentos linguísticos e não-linguísticos e de todos os referentes do mundo físico e do universo cultural.”

Portanto, para se compreender a importância do estudo do léxico é preciso reconhecer a relevância do tripé léxico, cultura e sociedade. Segundo Duranti (2000, p.127):

Se quisermos compreender o papel da língua na vida das pessoas, precisamos ir além do estudo de sua gramática e entrar no mundo da ação social, onde as palavras são encaixadas e constitutivas de atividades culturais específicas, tais como contar história, pedir um favor, mostrar respeito, insultar [...].

### **Parque Nacional da Serra da Canastra: uma breve história**

O Parque Nacional da Serra da Canastra (PNSC), que compreende uma área de aproximadamente 200 mil hectares, foi criado em 3 de abril de 1972, por meio do Decreto nº 70.355 e está situado na região sudoeste do Estado de Minas Gerais, abrangendo os municípios de São Roque de Minas, Sacramento, Delfinópolis, São João Batista do Glória, Capitólio e Vargem Bonita.

Quanto ao sistema hidrográfico, a região abrange as porções de cabeceiras das bacias hidrográficas dos rios São Francisco e Paraná, estando esta representada pelas bacias do rio Grande, ao sul, e a do rio Paranaíba, ao norte, a qual recebe os aportes das cabeceiras do rio Araguari. As características mais marcantes da região são a densa rede de drenagem com inúmeros tributários e centenas de nascentes que alimentam os diversos cursos de água.

Para a região, foram distinguidas seis bacias hidrográficas mais importantes: rio Grande e ribeirão Santo Antônio, ao sul; e ribeirão Grande, rio São Francisco, rio Araguari e rio Santo Antônio, ao norte.

Os primeiros momentos históricos relacionados com a região onde o PNSC está inserido vinculam-se à chegada da expedição comandada por Américo Vespúcio à foz do rio São Francisco, em 1501, no dia de São Francisco, motivo pelo qual o rio recebeu esse nome. Entretanto, a exploração deu-se apenas com a vinda da comitiva de Tomé de Souza, que trouxe o português Garcia d'Ávila, um dos primeiros desbravadores a penetrar o interior do continente a partir do "grande rio".

Uma vasta região do interior seria então povoada pelo colono, tendo o vale do rio São Francisco, denominado o rio dos currais, como condutor natural dessa ocupação essencialmente fundada na atividade pecuária, que abastecia de carne todo o litoral canavieiro.

Mas foi o sonho de riqueza e as crenças do homem europeu na existência de um paraíso material no novo continente que impulsionaram diversas expedições a buscarem cada vez mais longe as terras desconhecidas. Diferentes roteiros e relatos de viagens quinhentistas oriundos das diversas expedições que penetraram o sertão falavam de serras resplandcentes de ouro e prata e uma grande lagoa, "fabulosamente rica", existente no interior do continente.

Entre as muitas expedições que, nos séculos XVI e XVII, buscaram as cabeceiras do rio São Francisco, destaca-se a de Francisco Bruza de Espinosa, que, guiado por índios "mansos", avançou pelo rio São Francisco até as proximidades do rio das Velhas, de onde retornou devido à resistência dos índios hostis. Em 1581, Gabriel Soares de Sousa tentou itinerário semelhante, vindo a falecer durante a expedição. Outras expedições se sucederam partindo de diferentes pontos da costa e de São Paulo, como a empreendida por André de Leão, em 1601.

No que se refere às populações indígenas presentes na região do alto rio São Francisco, poucos são os estudos existentes, mas pode-se afirmar que, antes do processo de colonização portuguesa, a região foi habitada por diversas nações, podendo-se citar os acorá, araxá, araxaué, bororo, cataguase, estes predominantes no sudeste do Estado, e os caiapó, que se destacam entre os grupos que ocuparam o noroeste e o oeste mineiro.

O crescente trânsito na região, decorrente da abertura dos caminhos do ciclo minerador de Goiás e Paracatu, favorecia a fuga dos escravos trazidos pelos colonizadores que aqui chegavam, os quais sumiam das comitivas nas viagens e iam juntar-se aos quilombos, que ocupavam a margem esquerda do rio São Francisco e suas nascentes.

Outros povoados quilombolas também se estabeleceram ao sul e ao norte das serras da Canastra e Marcela e nas regiões mais próximas dos municípios de Piumhi e Formiga, sendo dizimados pelas expedições de extermínio a mando das autoridades coloniais.

No século XIX, a região foi percorrida pelo mineralogista alemão W. Von Eschwege e pelo botânico francês Auguste Saint-Hilaire, que registraram informações sobre os locais por onde passaram. Seus relatos tiveram grande importância no conhecimento da história da região.

A maior parte dos núcleos de povoamento que hoje compõem a rede urbana em torno da serra da Canastra, embora tenha origem no período setecentista, consolidou-se no século XIX. Essa rede surgiu de uma origem comum, assentada na aniquilação ou expulsão dos índios e quilombolas e na penetração gradativa dos colonos brancos e mestiços, que, devido à crise da mineração nos grandes centros auríferos, buscaram novas terras e atividades econômicas. A princípio, empreendiam atividades de garimpo e pequenas faisqueiras e, após esgotados os recursos auríferos, dedicaram-se à agricultura e especialmente à criação de gado.

Ao longo do século XIX, podem ser apontados alguns eixos de ocupação e desenvolvimento que levaram à diferenciação das localidades: o Triângulo Mineiro, o rio São Francisco e o rio Grande.

No que se refere aos aspectos culturais, pode-se afirmar que a região é bastante rica, tendo sua cultura baseada no tear artesanal, na produção de farinha, doce caseiro e queijos, além de festas tradicionais e populares e do patrimônio edificado.

## **Metodologia**

Para fundamentação teórica deste trabalho, procedemos à leitura de vários textos de diversos autores que tratam do léxico e da sua relação com a cultura e a sociedade. Nossa pesquisa se apoia, portanto, nos fundamentos da Sociolinguística e da Lexicologia.

Como se trata de um trabalho de cunho lexicológico, envolvendo, portanto, a sociedade e a cultura locais, propusemo-nos a focalizar aspectos históricos da região pesquisada, conhecimentos necessários, segundo Seabra (2004), para um estudo bem fundamentado do léxico regional.

Os dados aqui apresentados foram retirados de entrevistas já realizadas na zona rural dos municípios pertencentes ao PNSC. Em um primeiro momento, lemos as entrevistas realizadas com pessoas idosas (acima de 75 anos) que variavam a escolaridade entre não alfabetizados e alfabetizados. Em seguida, retiramos aquelas lexias que, a nosso ver, representavam o dialeto rural. Em um terceiro momento, procedemos à consulta dos dicionários Houaiss e Villar (2001) e Aurélio (FERREIRA, 1999).

## **Apresentação dos resultados parciais**

Tendo em vista o objetivo deste artigo, procuraremos descrever e analisar os dados hauridos do *corpus*, selecionados a partir de algumas entrevistas realizadas na comunidade rural da Serra da Canastra, região Sul/Sudoeste de Minas Gerais. A análise aqui abordada será apenas quantitativa. São 81 lexias apresentadas em ordem alfabética e transcritas conforme o modelo adotado por Amaral (2003) e Seabra (2004) que, por sua vez, seguem, em sua maioria, normas adequadas pela equipe de trabalho do projeto *Filologia Bandeirante* (COHEN, 1997). As lexias foram analisadas através de dois dicionários, sendo eles Houaiss e Villar (2001) e Aurélio (FERREIRA, 1999). Passemos à análise quantitativa dos nossos dados.

## **As Lexias**

Agoro, alquer, ajitório, amarelão, aperto, assistente de criança, bamo, bana, bassora, bichareda, bera, berá, berganha, bote, brabeza, buli, cado, caducano, caduco, coro, criação, cumpadi, cumê, cutelo, das vez, dereitinho, desmantelano, destampeí, dibuiano, donde, doraiada, estundá, fuçá, gatinhá, gente, graúdo, imbuchá, impariado, impreitada, inda, ingastaio, intimão, intremeio, invocado, iscumungado, jeriza, male, mandiocal, neveo, paió, panha, pelejei, pinga, posei, premero, querdito, rabicho, raliá, rama, resguardo, riba, rombo, ruindade, sacudida, sapé, sole, sorterão, taio, tale, tapo, terra, traveiz, trem, trucia, trupiquei, tulerá, turá, tustão, uai, zangá, zelá.

## **Análise quantitativa**

Considerando os objetivos traçados para a presente pesquisa – levantamento e análise do léxico da Serra da Canastra – apresentaremos esses resultados parciais com o intuito de melhor operacionalizarmos o processo de sistematização dessas unidades linguísticas e verificarmos a incidência de fatores socioculturais na configuração da construção do léxico do português dessa região no decorrer da pesquisa.

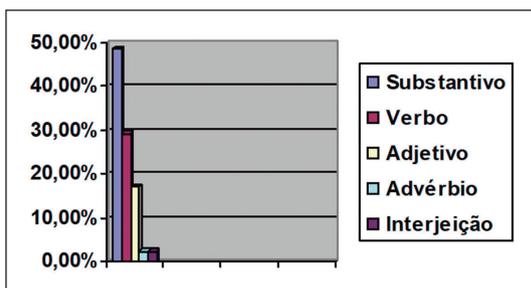
## **Quanto ao número de lexias presentes em cada dicionário**

Das 81 lexias coletadas até o momento verificamos que 41 são dicionarizadas em pelo menos um dos dois dicionários escolhidos.

## **Dicionarização segundo a classificação gramatical**

Neste item abordamos o número de lexias dicionarizadas ou não-dicionarizadas de acordo com as classes gramaticais. O gráfico a seguir mostra o percentual de unidades léxicas por função gramatical entre aquelas dicionarizadas:

**Gráfico 1** – Percentual de lexias dicionarizadas por função gramatical



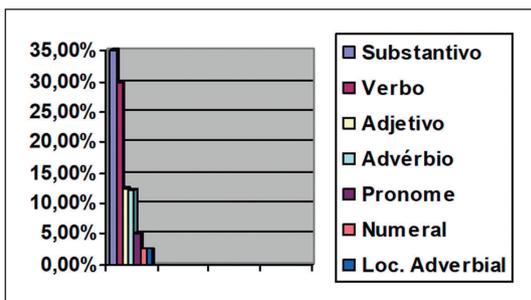
Fonte: Dados de pesquisa.

Após análise das 41 lexias dicionarizadas constatamos que 20 delas desempenhavam nas frases o papel de substantivos, o que equivale aos 48,78% mostrados no Gráfico 1. Os verbos somaram 12 lexias o que equivale aos 29,26%. Os adjetivos aparecem em terceiro lugar com 7 unidades lexicais, o que representa 17,07%. Os advérbios e as interjeições somaram 2 unidades lexicais, contribuindo com 4,86% do total de vocábulos dicionarizados.

Passemos agora a quantificação das lexias não-dicionarizadas por função gramatical.

Veja a seguir o Gráfico 2:

**Gráfico 2** – Percentual de lexias não-dicionarizadas por função gramatical



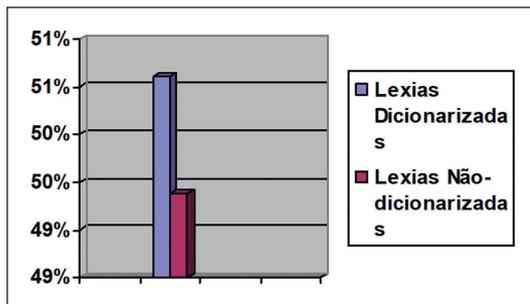
Fonte: Dados da pesquisa.

Quanto aos 40 vocábulos não-dicionarizados, mais uma vez os substantivos se destacaram com 14 casos ou 35%. A seguir aparecem os verbos com 12 unidades, o que representa 30% dos casos. Os adjetivos e os advérbios vêm em terceiro lugar com 5 lexias cada ou 25% do total não-dicionarizado. Os pronomes não-dicionarizados aparecem em quarto lugar com 2 casos, o que representa 5% do total não-dicionarizado. Os numerais e as locuções adverbiais aparecem em último lugar somando 5%.

### Quanto às lexias dicionarizadas e não-dicionarizadas

Após analisar as 81 lexias, identificamos 40 lexias que não foram encontradas em nenhum dos dicionários consultados, ao passo que as outras 41 foram encontradas em pelo menos um desses dicionários. Cabe ainda salientar: a) aquelas lexias que no contexto das entrevistas apresentaram sentido incompatível às acepções dicionarizadas foram computadas como não-dicionarizadas; b) vocábulos que apresentavam alguma variação resultante de processos fonológicos ocorridos em relação às formas das entradas nos dicionários também foram computados como não-dicionarizados; c) diferenças ortográficas como consoantes duplas ou terminações em *-am*, bem como diferenças de gênero, número e grau não foram consideradas como diferentes das formas encontradas nas entradas dos dicionários consultados.

**Gráfico 3** – Distribuição percentual das lexias dicionarizadas e não-dicionarizadas



Fonte: Dados da pesquisa

As 81 lexias utilizadas para essa análise foram divididas em dois grupos: lexias dicionarizadas, representadas pela barra azul do gráfico, e as lexias não-dicionarizadas, representadas pela barra vermelha. Entre os dois grupos predominaram as primeiras com 50,61% ou 41 lexias dicionarizadas, contra 49,39% ou 40 lexias não-dicionarizadas.

### **Considerações finais**

Os resultados parciais aqui apresentados constituem uma pequena amostra da pesquisa que ainda está em andamento no Programa de Pós-Graduação de Estudos Linguísticos da Universidade Federal de Minas Gerais. As transcrições das entrevistas já estão concluídas e estamos na etapa de seleção do *corpus* para posterior análise linguística.

Pelo fato de estar em uma região onde o povoamento aconteceu principalmente por causa das terras férteis e pela presença de riqueza material, esperamos encontrar uma certa homogeneidade linguística na região, ou seja, esperamos encontrar redes semânticas que privilegiem o povoamento do território e também a vida do homem nesse seu espaço de vivência e convivência. Baseando-nos no que observamos nos dados aqui apresentados, podemos dizer que essa região parece ter criado um vocabulário específico; já que o índice de lexias não-dicionarizadas quase alcança o índice das lexias dicionarizadas.

Vale a pena ressaltar que esses são os primeiros resultados, e essa é uma análise inicial do trabalho em questão. É preciso enfatizar que ainda há muito a ser pesquisado.

### **REFERÊNCIAS**

ALKMIM, T. M. Sociolinguística. In: MUSSALIM, F.; BENTES, A. C. (Org.). **Introdução à linguística**. São Paulo: Cortez, 2001. p.21-47.

AMARAL, E. T. R. **A ausência/presença de artigo definido diante de antropônimo em três localidades de Minas Gerais:**

Campanha, Minas Novas e Paracatu. 2001. 139f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

BENVENISTE, E. **Problemas de linguística geral II**. São Paulo: Cia. Ed. Nacional: EDUSP, 1989.

BIDERMAN, M. T. C. O dicionário na contemporaneidade: arquitetura, método e técnicas. In: ISQUERDO, A. N.; OLIVEIRA, A. M. P. P. de (Org.). **As ciências do léxico: lexicologia, lexicografia e terminologia**. Campo Grande: Ed. da UFMS, 2001. p.131-144.

CARDOSO, S. A. M. Língua: meio de opressão ou socialização? In: FERREIRA, C. et al. **Diversidade do Português do Brasil: estudos de dialectologia rural e outros**. Salvador: Centro Editorial e Didático da UFBA, 1988. p.231-235.

COHEN, M. A. A. de M. et al. Filologia Bandeirante. **Filologia e Lingüística Portuguesa**, São Paulo, n.1, p.79-94, 1997. Disponível em: <<http://www.revistas.usp.br/flp/article/view/59647/62743>>. Acesso em: 16 fev. 2018.

COSERIU, E. **Competência linguística**. Madrid: Gredos, 1992.

COSERIU, E. **El hombre y su lenguaje**. 2.ed. Madrid: Gredos, 1991.

DURANTI, A. **Antropologia lingüística**. Tradução espanhola de Pedro Tena. Madrid: Cambridge University Press, 2000.

FERREIRA, A. B. de H. **Novo Aurélio Século XXI: o dicionário da Língua Portuguesa**. 3.ed. Rio de Janeiro: Nova Fronteira, 1999.

HOUAISS, A.; VILLAR, M. de S. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2001.

ISQUERDO, A. N. Vocabulário do seringueiro: campo léxico da seringa. In: ISQUERDO, A. N.; OLIVEIRA, A. M. P. P. de (Org.). **As ciências do léxico: lexicologia, lexicografia e terminologia**. Campo Grande: Ed. da UFMS, 2001. p.91-100.

SAUSSURE, F. **Curso de Linguística Geral**. São Paulo: Cultrix, 1969.

SEABRA, M. C. T. C. de. **A formação e a fixação da língua portuguesa em Minas Gerais**: a toponímia da região do Carmo. 2004. 368f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.



# **O REGIONAL E O GLOBAL NOS NOMES PRÓPRIOS DE PESSOAS NO BRASIL**

**Jéssica Paula VESCOVI**  
**Márcia Sipavicius SEIDE**

## **Introdução**

Uma marca de identificação levada para a vida toda é o nome próprio, que pode ser revelador de idade, de cultura e de nacionalidade. A grande importância de um nome próprio de pessoa já pode ser percebida no ato do nascimento de um bebê, pois além de obrigatório, o nome do bebê é algo que causa muita preocupação nos pais, tendo em vista que o nome escolhido pode atribuir ao filho uma identificação simbólica que pode influenciar-lhe para a vida toda.

Tendo em vista que além de ser motivo de preocupação para os pais, o nome próprio de pessoa pode ser estudado numa tentativa de compreender as realidades de diferentes comunidades em dados períodos, este estudo procura analisar três comunidades diferentes investigando quais são os prenomes presentes em todas e quais são os prenomes presentes em algumas, chamando-os de nomes globais e de nomes regionais, respectivamente. Serão investigados, neste trabalho, o País Basco (2014a, 2014b), na Espanha, observando, a partir de um levantamento publicado pelo governo de tal estado,

os nomes preferidos dos bascos nos anos de 2008 a 2010, levando em conta tanto os nomes masculinos quanto os nomes femininos; o Brasil, de forma geral, a partir de uma pesquisa encomendada pela revista *Superinteressante* que demonstra quais são os nomes preferidos dos brasileiros em 2012 e de um levantamento feito pelo site *Terra* (LEVANTAMENTO, 2014), que apresenta os nomes mais usados desde o século XIX; e uma amostra de dados feita nos municípios de Palotina e Maripá, no oeste do Paraná, que engloba os nomes próprios de registrados dos anos de 1957 a 2007, em Palotina, e de 1966 a 2006, em Maripá, na qual foram recolhidos os nomes dos cem primeiros registrados em cada década.

Para tanto, em um primeiro momento, são apresentados alguns estudos acerca do sistema antroponímico de diferentes países, associando-os à questão da identidade proposta por Hall (2006). Na sequência, será apresentada uma análise dos nomes globais encontrados nas comunidades investigadas e, por último, será apresentada uma análise dos nomes regionais encontrados.

## **Os locais estudados**

Apresenta-se, a seguir, um pequeno percurso histórico sobre as comunidades que serão analisadas, uma vez que a apresentação da história destes locais contribui para um melhor entendimento das formas de nomeação e, portanto, das ideologias de cada localidade.

## **O País Basco**

O País Basco é um estado espanhol que conserva boa parte de sua cultura, principalmente através do nacionalismo e da língua falada, o *euskera*/basco que se difere do espanhol e do francês em muitos aspectos. Além disso, de acordo com vários sites, como o site do Estado Basco, trata-se de um estado autônomo que tem sua própria tradição e sua própria cultura. O estado tem aproximadamente 3 milhões de habitantes e encontra-se na região da Aquitânia.

## **O município de Palotina**

A cidade de Palotina, de acordo com informações apresentadas no livro de Reginato (1979) e de Barros (2002) é conhecida por ser uma cidade colonizada por gaúchos e catarinenses vindos ao Oeste do Paraná em meados dos anos 50. Assim como diz o próprio nome da cidade Palotina, em homenagem aos padres palotinos, a religião católica sempre foi a predominante da cidade, que tem como principal fonte de renda o agronegócio. No início de sua colonização, Palotina tinha cerca de 5000 habitantes e hoje se estipula, de acordo com o senso do IBGE de 2010, que a cidade tenha cerca de 28.000 moradores.

## **O município de Maripá**

A cidade de Maripá foi distrito de Palotina até o início dos anos 90. Conhecida, de acordo com informações colhidas no *site* da prefeitura desta cidade, pelas belas ruas enfeitadas de orquídeas e pela tradicional festa do peixe, a cidade traz consigo uma tradição germânica muito forte e tem em sua cultura a forte predominância do luteranismo. Além disso, o município maripaense tem cerca de 5000 habitantes e tem como base de renda o agronegócio, assim como a cidade de Palotina. Cumpre informar, contudo, que mesmo sendo distrito de Palotina, os registros começaram a ser feitos no cartório do distrito no ano de 1966 e é por isso que se tem essa data como a inicial na coleta dos dados desta cidade.

## **Os nomes globais**

Há vários nomes que podem ser encontrados em diferentes países, mas que podem ter uma significação similar quando vistas as possíveis motivações para as nomeações.

Nesta seção do artigo, intenta-se apresentar quais são os nomes encontrados em todas as fontes do *corpus*: no país Basco, nos nomes preferidos dos brasileiros, no *corpus* de Palotina e no *corpus* de Maripá. Nomes presentes em todos os *corpora* indicam que

foram utilizados em todas as localidades citadas; tais nomes foram considerados como nomes globais.

Conforme se pode observar na tabela de prenomes globais masculinos, todos os cinco prenomes globais são nomes religiosos que podem ou não ter sido empregados com motivação religiosa. Interessante ressaltar que esses prenomes podem ser encontrados em todos os locais do mundo, pois são vistos como “nomes fortes”, sendo também muito frequentes na mídia, por serem nomes banalizados.

Dadas suas características, é uma difícil tarefa apresentar uma possível motivação para o emprego destes nomes em todos os lugares estudados, porém, acredita-se que haja uma variação nas possíveis motivações, sendo todas possivelmente relacionadas com questões ideológicas e culturais.

Com relação aos nomes globais femininos, há, no topo, dois nomes muito disseminados e que também são nomes de santas: Ana e Maria. É necessário informar, contudo, que o prenome Ana, registrado no país Basco como **Anne** ou **Anna**, e o prenome Maria já eram encontrados, no Brasil, de acordo com o levantamento do *site Terra*, no século XIX, o que demonstra que há forte tradição destes no mundo. Quanto aos prenomes **Julia** e **Adriana**, presentes em todos os *corpora*, vê-se que são nomes relativamente novos e que surgiram e foram se espalhando pelo mundo principalmente depois da grande presença da mídia em todos os locais. No Brasil, de acordo com o levantamento feito pelo *site Terra*, os nomes surgiram principalmente depois das décadas de 60 e 70.

**Tabela 1** – Nomes masculinos e femininos

| MASCULINOS | FEMININOS |
|------------|-----------|
| JOÃO       | MARIA     |
| PAULO      | ANA       |
| LUCAS      | JULIA     |
| GABRIEL    | ADRIANA   |
| MARCOS     |           |

**Fonte:** Elaboração própria.

Vê-se, portanto, que a maioria dos nomes considerados globais são prenomes que têm toda uma trajetória histórica nos processos de nomeação das mais variadas comunidades e que não seria nada surpreendente encontrar um nome como Maria ou como João em países de colonizações diferentes, o que não pode ser dito dos prenomes considerados regionais.

## **Os nomes regionais**

Os nomes globais apresentam uma característica peculiar: representam o sistema antroponímico mundial de forma geral, ou seja, demonstram que determinadas crenças podem estar presentes em todos os lugares.

Por outro lado, os nomes regionais, assim classificados os nomes presentes em apenas algumas localidades, apresentam as crenças e as ideologias peculiares de determinado local. Em geral, conforme apresenta Langendonck (2007, p.306, tradução nossa)<sup>1</sup>, “[...] os nomes próprios são signos linguísticos ancorados socialmente e são grande parte e parcela do inventário linguístico de uma comunidade.” ou seja, em paráfrase ao dito pelo autor sueco, os nomes próprios de pessoas podem representar o pensamento de determinada população e é por isso que investiga-se nesta seção do artigo quais são os nomes locais das comunidades investigadas.

O estudioso onomástico também menciona que os nomes próprios estão sujeitos a alterações devido aos lugares em que são empregados, não tendo somente uma função formal, mas também uma posição social e ideológica, o que é contrário ao apresentado pelos filósofos da linguística, que dizem que a única função do nome próprio é designar e referir. O posicionamento do pesquisador sueco, contudo, vai ao encontro do ponto de vista defendido pela pesquisadora brasileira Maria Vicentina Dick do Amaral. Para Dick (1992), o nome próprio precisa ser estudado em seu aspecto social e cultural, tendo em vista que à onomástica interessa não o

---

<sup>1</sup> Original: “*Proper names are socially anchored linguistic signs and as such part and parcel of the linguistic inventory of a society.*” (LANGENDONCK, 2007, p.306).

nome próprio em si mesmo, mas sim a relação entre o nomeador, o objeto nomeado e o receptor, em outros termos, entre o nome, o portador do nome e o designador.

Tendo em vista que não se pode separar o estudo dos nomes próprios dos estudos etimológicos, não se pode deixar de mencionar Mansur Rosário Guérios (1981), professor da Universidade Federal do Paraná (UFPR) e responsável pela elaboração de um dos primeiros dicionários etimológicos de nomes próprios de pessoas no Brasil. Numa tentativa de associar os estudos nacionais sobre a onomástica com estudos internacionais, como o de Langedonck (2007), tenta-se relacionar os estudos de Dick e de Guérios à sócio-onomástica, tendo em vista que essa

[...] se vale do princípio de que a onomástica é constantemente renovada: novos nomes se originam com novas pessoas. Muito mais do que a sistematicidade lexical, as estruturas e os inventários dos nomes próprios de pessoas podem refletir os novos desenvolvimentos de uma sociedade, como por exemplo, a relação entre as diferentes classes sociais, entre pessoas jovens e velhas ou entre ideologias. (LANGEDONCK, 2007, p.308, tradução nossa).<sup>2</sup>

O autor sueco considera que os nomes próprios, principalmente de pessoas, são flexíveis e adaptáveis, tendo em vista que é esperado que eles sejam reflexos das crenças dos moradores de determinada comunidade.

A relação entre mudanças observadas no repertório dos nomes próprios de pessoas e aquelas ocorridas no âmbito social pode ser melhor entendida quando se leva em consideração a questão de identidade da era moderna proposta por Hall (2006). Para este

---

<sup>2</sup> Original: “*The socio-onomastic value of personal names also lies in the fact that the onomasticon is constantly renewed: new names originate with new individuals. Hence, much more than the appellative lexicon and system, the personal name inventories and structures can and do reflect new developments in society fairly quickly, e.g., regarding the man-woman relation, the differences between social classes, between old and young people, etc.*” (LANGEDONCK, 2007, p.308).

antropólogo, há inúmeros reflexos de uma questão que atinge a sociedade moderna: a crise de identidade dos migrantes às regiões a que emigra, que pode representar uma tentativa de manter a tradição, assim como pode representar uma tentativa de adaptação à sociedade a qual se está fazendo parte no momento.

A questão da identidade de uma comunidade é muito bem estudada por Stuart Hall, que aponta as implicações, as consequências e as qualificações da fragmentação da identidade. Sempre associadas, então, ao tempo e ao espaço ocupado pelo sujeito e, nas palavras de Hall

Todas as identidades estão localizadas no espaço e no tempo simbólicos. Elas têm aquilo que Edward Said chama de suas “geografias imaginárias” (Said, 1990): suas “paisagens” características, seu senso de “lugar”, de “casa/lar”, ou *heimat*, bem como suas localizações no tempo – nas tradições inventadas que ligam passado e presente, em mitos de origem que projetam o presente de volta ao passado, em narrativas de nação que conectam o indivíduo a eventos históricos nacionais mais amplos, mais importantes. (HALL, 2006, p.19).

Em suma, o que se pode dizer é que as identidades podem ser construídas a partir das nomeações e cada qual é constituída dentro de sua comunidade, comportando as ideologias, os costumes e as tradições desta e prova disso são os nomes regionais.

Conforme análise feita acerca dos nomes regionais das comunidades estudadas e retomando o apresentado no início do texto quando se menciona o contexto social-histórico de cada localidade, sugere-se, então, que, nos locais estudados, houve/há a intenção da manutenção da tradição cultural a partir do processo de nomeação, tendência que coexiste com o uso de nomes globais que se influenciam mutuamente.

## Os nomes regionais do País Basco

O país Basco, na Espanha, possui um sistema de nomeação muito singular, principalmente no que diz respeito aos primeiros nomes masculinos. A partir do levantamento feito acerca dos nomes próprios preferidos dos moradores daquela localidade de 2008 a 2010, foram selecionados os cinco primeiros colocados no *ranking* que por não aparecerem no *corpus* brasileiro, tanto no levantamento da *Superinteressante* de 2013 quanto no levantamento do *site Terra*, foram classificados como os nomes regionais masculinos preferidos dos moradores do País Basco nesta época, quais sejam:

**Tabela 2** – Nomes Regionais masculinos do País Basco

| NOMES REGIONAIS MASCULINOS – PAÍS BASCO |        |
|---|--------|
| 1º lugar                                | IKER   |
| 2º lugar                                | MARKEL |
| 3º lugar                                | OIER   |
| 4º lugar                                | UNAI   |
| 5º lugar                                | ANDER  |

**Fonte:** Elaboração própria

Conforme se observa na tabela, os nomes regionais de meninos do país estudado não são nomes comuns no Brasil, o que confirma a ideia de que cada comunidade constitui uma regra de nomeação, obedecendo a um padrão que, geralmente, segue as ideologias e as crenças deste local.

Com o intuito de investigar a presença destes nomes em outros momentos na Espanha, utilizou-se da ferramenta *Google* para identificar o uso em outros anos. Pesquisando nomeações anteriores utilizando estes nomes, constatou-se que entre os cinco primeiros, dois, *Iker* e *Oier*, eram jogadores de futebol, aquele defendendo a seleção espanhola e este defendendo a seleção espanhola e o time do Barcelona, um dos maiores times de futebol da Espanha; um

dos nomes; *Unai*, é de um ex-jogador da seleção espanhola e atual técnico de futebol daquela localidade.

Estes resultados indicam que seria comum aos meninos receberem nomes típicos espanhóis numa tentativa de manter a tradição e, acima de tudo, da tentativa dos pais em prestar uma homenagem a seus ídolos, no caso, os jogadores de futebol. A presença de nomes escolhidos por influência de pessoas celebradas pelos meios massivos de televisão aponta para uma motivação relativamente moderna. A percepção de novas motivações em determinadas comunidades demonstra que a língua está em constante movimento e esta dinamicidade dá origem a motivações novas. Neste sentido, as motivações apresentadas por Seide (2013) invocam situações não previstas por Dick (1992) e por Guérios (1981), tais como a causalidade, a busca de univocidade, as crenças do designador, as idiossincrasias do designador, além da influência midiática citada brevemente por Dick (1992) e por Guérios (1981).

Em contrapartida, na nomeação das meninas do país Basco, há uma diferença acerca das possíveis motivações na escolha destes prenomes. Conforme se observa no quadro abaixo, os cinco primeiros nomes regionais preferidos dos bascos foram *June*, *Uxue*, *Irati*, *Nahia* e *Izaro*. O primeiro colocado, *June*, de acordo com a etimologia e o significado apresentados por Guérios (1981) faz referência a uma deusa da mitologia grega, a deusa de Juno, ou a deusa da juvenilidade e é de origem francesa ou inglesa. Se este nome foi escolhido em função de seu significado etimológico, seu emprego vem ao encontro do já apresentado por Guérios (1981) quando diz que ao escolher os nomes dos filhos, os pais podem estar projetando o futuro deles; no caso de *June*, haveria a intenção de que a bondade prevaleça na personalidade da criança. Os outros nomes encontrados como regionais no estado do País Basco não são comuns no Brasil e devem estar atrelados à cultura daquela localidade.

**Tabela 3** – Nomes regionais femininos do País Basco

| NOMES REGIONAIS FEMININOS – PAÍS BASCO |       |
|--|-------|
| 1º lugar                               | JUNE  |
| 2º lugar                               | UXUE  |
| 3º lugar                               | IRATI |
| 4º lugar                               | NAHIA |
| 5º lugar                               | IZARO |

**Fonte:** Elaboração própria

### **Os nomes regionais de Palotina e Maripá**

Com a intenção de comparar os nomes regionais do País Basco com os nomes regionais do Brasil, mais especificamente da região Oeste do Paraná, foi feito um levantamento dos nomes regionais das cidades de Palotina e Maripá, observando década por década desde 1957 em Palotina e desde 1966 em Maripá, elegendos os mais excêntricos e os com maior frequência. Constatou-se, contudo, que, na antroponímia destas localidades, há uma maior presença de nomes regionais masculinos, o que é contrário ao visto no País Basco, no qual há maior variedade e maior diversidade na antroponímia feminina.

O perceptível na análise dos municípios paranaenses é que houve, por parte dos pais registradores em Maripá (uma cidade, como já apresentado) a intenção de manter a tradição germânica nos nomes dos filhos registrados naquela cidade, principalmente no início da colonização daquele município, nos anos 60 e 70. Acontece, então, o dito por Hall (2006) que afirmava haver a intenção de manter a identidade das comunidades em algumas situações.

**Tabela 3** – Relação dos nomes regionais masculinos de Maripá – PR

| NOMES REGIONAIS MASCULINOS – MARIPÁ |                           |
|-------------------------------------|---------------------------|
| 1966                                | WILFRIED, ROLF            |
| 1976                                | DANNY, FRIDOLINO, LILIVAL |
| 1986                                | DIETER                    |
| 2006                                | RYAN, LEANDER             |

Investigou-se, também, a etimologia dos nomes regionais masculinos da cidade de Maripá apresentada na tabela abaixo que intenta apresentar a origem de cada um a partir de pesquisas em Guérios (1981) e no *site* “Dicionários de nomes próprios” (2014).

**Tabela 4** – Origem etimológica dos prenomes regionais registrados em Maripá

| PRENOME                 | ETIMOLOGIA            |
|-------------------------|-----------------------|
| Wilfried (Winfried)     | Alemã                 |
| Rolf (abreviação Rolfo) | Alemã                 |
| Danny                   | Origem não encontrada |
| Fridolino               | Alemã                 |
| Dieter                  | Origem não encontrada |
| Ryan                    | Irlandês              |
| Leander                 | Grego                 |

**Fonte:** Elaboração própria

De acordo com o observado na tabela acima, dos 7 nomes classificados como regionais, a etimologia de 3 é alemã, o que aponta que, de certa forma, pode ter havido influência da ascendência na nomeação destes registrados. Também é possível observar que o nome *Dieter* não tem origem definida, mas pelas iniciais e pela origem da família (germânica), infere-se que pode ter havido a intenção de nomear o filho de acordo com a ascendência familiar.

Observa-se, contudo, que, na cidade de Palotina, não há tantos nomes regionais na antropônimo masculina e os nomes considerados regionais também apresentam uma característica peculiar: na maioria das vezes, remetem à origem da família e podem ter vindo com os colonizadores desta região, uma vez que a grande maioria dos nomes masculinos regionais foi encontrada na primeira ou na segunda década de colonização.

**Tabela 5** – Nomes regionais masculinos em Palotina – PR

| <b>NOMES REGIONAIS MASCULINOS – PALOTINA</b> |  |
|--|--|
| 1957   | LILI   |
| 1967   | NEORI, VANOR, INGO, IVANOR,<br>HERCILIO, ERY, MATINO |
| 1977   | SILVANOR, GELSEY, INGLEMAR, GILDOMAR                 |

**Fonte:** Elaboração própria

**Tabela 6** – Origem etimológica dos prenomes regionais de Palotina – PR

| <b>PRENOME</b>     | <b>ETIMOLOGIA</b>   |
|--------------------|---------------------|
| Lili               | Inglês              |
| Neori              | Origem desconhecida |
| Vanor              | Origem desconhecida |
| Ingo               | Alemã               |
| Ivanor             | Origem desconhecida |
| Hercílio           | Latim               |
| Ery                | Origem desconhecida |
| Matino – Martino   | Bíblico             |
| Silvanor           | Origem desconhecida |
| Gelsey             | Origem desconhecida |
| Inglemar (Ingomar) | Alemã               |
| Gildomar           | Origem desconhecida |

**Fonte:** Elaboração própria

É, também, importante observar que na comunidade de Palotina houve/há a presença de algumas famílias germânicas e os nomes regionais encontrados são, na maioria, oriundos destas famílias. Entre os nomes acima, são de família germânica **Lili, Neori, Ingo, Ivanor**. Em contrapartida, o nome **Matino** é de família de origem italiana e os demais nomes são de famílias de origem portuguesa ou brasileira. Conforme observado tanto no *corpus* de Maripá quanto no *corpus* de Palotina, vê-se que as famílias alemãs tinham maior tradição ao nomear seus filhos e esta tradição pode estar relacionada à intenção de manter a tradição germânica na família, comprovando a ideia de que os descendentes de alemães são mais tradicionais que os descendentes de italianos (ambos provenientes do Rio Grande do Sul ou de Santa Catarina) no oeste do Paraná.

### **Considerações finais**

As análises apresentadas acima mostram que, na antroponímia, o valor do signo está relacionado ao *corpus* em questão, ou seja, observando-se o nome em confronto aos outros nomes do mesmo *corpus*, logo

Seu valor não estará então fixado, enquanto nos limitarmos a comprovar que pode ser “troçada” por este ou aquele conceito, isto é, que tem esta ou aquela significação; falta ainda compará-la com os valores semelhantes, com as palavras que se lhe podem opor. Seu conteúdo só é verdadeiramente determinado pelo concurso do que existe fora dela. Fazendo parte de um sistema, está revestida não só de uma significação, como também, e sobretudo, de um valor, e isso é coisa muito diferente. (SAUSSURE, 2012, p.162).

Porém, mesmo utilizando a noção saussuriana para explicar o valor do nome próprio em cada comunidade, não se desconsidera o social de cada localidade, pelo contrário, levam-se em consideração as ideologias, a cultura e a tradição de cada local e é por isso que

se acredita que cada comunidade tem um sistema de nomeação, que neste caso, chama-se de “norma antroponímica”, e que pode ser associada à noção de norma proposta por Coseriu (1979), para quem a língua deve ser entendida como função e depois como sistema, tendo em vista que, na visão do autor, se a língua funciona, não é por ser sistema, mas por ser um sistema constituído a partir do momento que cumpre uma função, ligada a fatores históricos que demonstram as marcas da organização sistêmica da língua.

A existência de nomes globais, contudo, vai além do signo e da norma linguística, sendo indícios do movimento cultural supranacional da globalização. O uso concomitante de nomes regionais e de nome globais, por sua vez, confirma as consequências da globalização para as identidades culturais previstas por Hall (2006). Por um lado, a atenuação de certas características regionais por força da homogeneização cultural possibilita a presença de um mesmo elenco de nomes em lugares diferentes e distantes; de outro, o esforço do regional como resistência à globalização explica a permanência de nomes regionais. Ambas as tendências resultam na criação de sociedade híbridas, logo, em todos os casos estudados, os nomes regionais convivem com os nomes globais.

## REFERÊNCIAS

BARROS, D. A. **Palotina em destaque**. Palotina: Gráfica Imprevalé, 2002.

BUSTAMANTE, F.; CASTRO, C.; VAN DEURSEN, F. Os nomes preferidos do Brasil: os nomes que os pais escolhem para os filhos entregam muita coisa sobre a história do país. **Revista Superinteressante**, São Paulo, n.325, nov. 2013.

COSERIU, E. **Lições de linguística geral**. Rio de Janeiro: Ao livro técnico, 1980.

DICIONÁRIO DE NOMES PRÓPRIOS. Disponível em: <<http://www.dicionariodenomesproprios.com.br/>>. Acesso em: 10 abr. 2014.

DICK, M. V. P. A. **Toponímia e antroponímia no Brasil**. 2.ed. São Paulo: FFLCH, 1992. (Coletânea de Estudos).

GUÉRIOS, R. F. M. **Dicionário etimológico de nomes e sobrenomes**. 3.ed. São Paulo: Ave Maria, 1981.

HALL, S. **A identidade cultural na pós-modernidade**. 11.ed. Rio de Janeiro: DP&A, 2006. Originalmente publicado em 1992. Disponível em: <[https://edisciplinas.usp.br/pluginfile.php/4135505/mod\\_resource/content/1/A%20Identidade%20Cultural%20na%20P%C3%B3s-Modernidade%20-%20Stuart%20Hall.pdf](https://edisciplinas.usp.br/pluginfile.php/4135505/mod_resource/content/1/A%20Identidade%20Cultural%20na%20P%C3%B3s-Modernidade%20-%20Stuart%20Hall.pdf)>. Acesso em: 16 fev. 2018.

LANGENDONK, W. Van. **Theory and typology of proper names**. Berlin; New York: Mouton de Gruyter, 2007.

LEVANTAMENTO mostra os nomes preferidos nos últimos 120 anos. **Site Terra**. Disponível em: <<http://noticias.terra.com.br/educacao/voce-sabia/nomes-120-anos/>>. Acesso em: 10 abr. 2014.

PAÍS BASCO. **Lista de los 100 nombres de niño más frecuentes en la C. A. de Ereskadi**: 2012-2014. 2014a. Disponível em: <[http://www.eustat.es/ele0005700/tbl0005716\\_c.html#axzz2qHC0gccU](http://www.eustat.es/ele0005700/tbl0005716_c.html#axzz2qHC0gccU)>. Acesso em: 10 abr. 2014.

PAÍS BASCO. **Lista de los 100 nombres de niña más frecuentes en la C. A. de Ereskadi**: 2012-2014. 2014b. Disponível em: <[http://www.eustat.es/ele0005700/tbl0005715\\_c.html#axzz2qHDLSJlp](http://www.eustat.es/ele0005700/tbl0005715_c.html#axzz2qHDLSJlp)>. Acesso em: 10 abr. 2014

REGINATO, P. **História de Palotina**. Santa Maria: Ed. Palloti, 1979.

SAUSSURE, F. **Curso de linguística geral**. Organizado por Charles Bally, Albert Sechahaye; Tradução de Antônio Cheline, José Paulo Paes, Izidoro Blikstein. 34.ed. São Paulo: Cultrix, 2012.

SEIDE, M. S. As motivações contemporâneas para a escolha dos antropônimos. **EntreLetras**, Araguaína, v.4, n.2, p.90-101, ago./dez. 2013. Disponível em: <<https://scholar.google.com.br/citations?user=ELAGLsUAAA&hl=pt-PT>>. Acesso em: 10 abr. 2014.



# A REGIONALIDADE NA SELEÇÃO LEXICAL DE INFORMANTES DO ATLAS LINGUÍSTICO DO ESTADO DE ALAGOAS (ALEAL)

Maranúbia Pereira Barbosa DOIRON

## ALEAL: como, onde, por quê?

Os atlas linguísticos regionais, como não poderia deixar de ser, são reconhecidos como ferramentas extremamente eficazes para os estudos que concernem à descrição do léxico regional em comunidade de falantes. Nesse sentido, o *Atlas Linguístico do Estado de Alagoas* – doravante, *ALEAL* – projeto em nível de doutorado desenvolvido na Universidade Estadual de Londrina (UEL), sob a orientação da Prof.<sup>a</sup> Dr.<sup>a</sup> Vanderci de Andrade Aguilera, vem ao encontro desse propósito, ao priorizar a descrição de fenômenos relativos à variação linguística referentes a informantes do Estado de Alagoas.

Embora quatro localidades de Alagoas integrem a rede de pontos do *Atlas Linguístico do Brasil (ALiB)*: Maceió (ponto 77), Arapiraca (76), Santana do Ipanema (75) e União dos Palmares (74), a consecução de um atlas que abranja toda a superfície do Estado, no caso, mais de 27 mil quilômetros quadrados, ao investigar pontos ainda não recobertos, só vem a ampliar o campo de conhecimento acerca da realidade dialetal em todo o Estado, e, por conseguinte, do português do Brasil.

O *ALEAL* tem como objetivo geral, descrever os fenômenos linguísticos no Estado de Alagoas junto a informantes estratificados e selecionados na capital e em localidades interioranas, considerando, prioritariamente, as diferenças diatópicas, seus níveis fônico, lexical e morfossintático. Para tanto, documentou a modalidade oral de falantes da zona urbana de Alagoas, com o intuito de identificar a existência de marcas de conservadorismo linguístico e traços de ruralidade. Os dados coletados permitem examinar as variantes linguísticas em interface com outros ramos do conhecimento, sejam eles: história, antropologia, sociologia, geografia humana, entre os mais pertinentes.

A exemplo dos atlas linguísticos contemporâneos da chamada 2ª geração, o projeto *ALEAL* busca respaldo nos pressupostos teóricos e metodológicos da Geolinguística (DAUZAT, 1922) e da Dialetoлогия Pluridimensional (THUN, 1998), ao investigar *in loco* a realidade linguística de toda uma comunidade em seu espaço areal, no intuito de reunir dados que apontem as diferenças diatópicas, diagenéricas e diageracionais no Estado de Alagoas. Os dados levantados por meio de aplicação de questionários específicos, de natureza sociolinguística, serão cartografados de modo a apresentar uma visão pluridimensional das variantes dialetais obtidas.

No que tange à rede de pontos, o *ALEAL* segue as orientações de Nascentes (1958), com 21 localidades contempladas. Excetuando-se as cidades de Canapi e Quebrangulo, que não entraram no rol daquele autor, e a exclusão do município de Junqueiro, onde encontrei dificuldades pontuais que me fizeram desistir das entrevistas no referido ponto, a rede mantém-se fiel às sugestões de Nascentes (1958). São essas as localidades:

**Figura 1** – Rede de pontos - ALEAL

| <b>REDE DE PONTOS – ALEAL</b>  |                                  |
|--------------------------------|----------------------------------|
| Delmiro Gouveia (Ponto 1)      | Quebrangulo (Ponto 12)           |
| Canapi (Ponto 2)               | Limoeiro de Anadia (Ponto 13)    |
| Santana do Ipanema (Ponto 3)   | São Miguel dos Santos (Ponto 14) |
| Piranhas (Ponto 4)             | Pilar (Ponto 15)                 |
| Pão de Açúcar (Ponto 5)        | Marechal Deodoro (Ponto 16)      |
| Traipu (Ponto 6)               | Maceió (Ponto 17)                |
| Penedo (Ponto 7)               | União dos Palmares (Ponto 18)    |
| Piaçabuçu (Ponto 8)            | São Luiz do Quitunde (Ponto 19)  |
| Coruripe (Ponto 9)             | Porto Calvo (Ponto 20)           |
| Arapiraca (Ponto 10)           | Maragogi (Ponto 21)              |
| Palmeira dos Índios (Ponto 11) |                                  |

**Fonte:** Banco de dados da autora.

Quanto aos informantes, eles são em número de **dois** por localidade, um homem e uma mulher, na faixa de 30 a 50 anos, com baixa escolaridade, sem a influência da educação formal, a fim de obter uma fala mais conservadora. Para verificar a influência da variável faixa etária, selecionei sete cidades dentre as mais antigas do Estado: Maceió, Marechal Deodoro, Penedo, Porto Calvo, Santana do Ipanema, São Miguel dos Campos, União dos Palmares e, nessas localidades, entrevistei **quatro** informantes distribuídos em duas faixas etárias, 30 a 50 anos e 55 a 75 anos. Os informantes do *ALEAL* são, portanto, em número de **56**, dos quais, **28** são homens e **28** são mulheres.

Quando da definição do perfil, levei em consideração o fato de que muitas das questões, notadamente as que tratam do mundo rural, poderiam ser bem mais produtivas se os informantes fossem dessa faixa etária. Numa faixa abaixo da delimitada, por exemplo, entre 18 e 30 anos, a produtividade, muito provavelmente, seria menor, visto que os informantes da zona urbana não têm ou nunca tiveram contato com alguns dos referentes pro-

postos pelos questionários. Ademais, o resgate de uma fala mais conservadora torna-se mais passível de ser obtido com informantes que já estão na fase da maturidade do que entre pessoas mais jovens, especialmente, se levarmos em conta o distanciamento dos informantes com menos de 30 anos de temas ligados à ruralidade, dado o apelo cada vez mais crescente de tudo o que se refere ao mundo urbano.

O *corpus* do *ALEAL* é formado pelas respostas obtidas por meio de questionários formulados a partir do Atlas Linguístico do Brasil (ALiB), com algumas adequações. São elas:

Questionário Fonético-Fonológico (QFF): 89 questões.

Questionário Semântico-Lexical (QSL): 199 questões

Questionário Morfossintático (QMS): 32 questões

Perguntas Metalinguísticas: 6 questões

São, portanto, 326 questões.

O Estado de Alagoas, segundo a divisão dialetal de Nascentes (1958), faz parte do “subfalar nordestino”. No mapa que se segue, disponível em <[www.alib.ufba.br](http://www.alib.ufba.br)>, acesso em 02 de maio de 2014, vê-se que Alagoas é o primeiro Estado que aparece na divisão entre os subfalares baianos e subfalares nordestinos.

Quando da finalização do *ALEAL*, os dados coletados e analisados poderão comprovar se a proximidade entre os dois subfalares concorre para separar de modo significativo as variantes lexicais e fonético-fonológicas dos informantes de Alagoas, das variantes registradas pelos demais informantes oriundos do que Nascentes (1958) chamou de subfalar baiano, e que podem ser consultadas em atlas regionais publicados, em teses de doutorados, entre outros trabalhos científicos.

**Figura 2** – Mapa



Fonte: Nascentes (1958).

### **A influência do espaço fisiográfico e de aspectos sócio-históricos na nomeação do mundo**

Dados preliminares do *corpus* do *ALEAL* revelam em primeiro lugar uma forte presença da regionalidade na escolha lexical dos informantes no Estado. De maneira ainda superficial, já que as transcrições fonéticas e grafemáticas estão em curso, posso adiantar que:

- (i) algumas das unidades lexicais registradas aparentam ser motivadas semanticamente por fatores de natureza geográfica, sociocultural ou histórica.
- (ii) as ocorrências linguísticas registradas, conforme postula Biderman (2001, p.13), podem ter sido geradas por meio de “[...] atos sucessivos de cognição da realidade e de categorização da experiência, cristalizada em signos linguísticos: as palavras.”

Algumas respostas ao Questionário Fonético-Fonológico (QFF), cujo objetivo é obter dos informantes termos específicos a cada questão inquirida, embora com variações de forma de ordem fônica, não atenderam ao propósito do QFF. Contudo, essas respostas não invalidam as questões: subjazem a elas “[...] elementos significativos relativos à história, ao sistema de vida à visão de mundo de um determinado grupo.” (ISQUERDO; OLIVEIRA, 2001, p.91).

A Questão 10 do QFF do *ALEAL*, que pretende verificar a nasalidade da vogal alta inicial seguida de consoante nasal, no caso, a unidade lexical [ímã], obteve resultados que corroboram a presença de fatores regionais entre os informantes de Alagoas.

Passo, portanto, à forma como a questão é colocada pelo inquiridor ao informante, e, em seguida, à análise das ocorrências lexicais encontradas.

### **ÍMÃ**

“... aquilo que atrai objetos pequenos de metal, como agulha, prego e alfinete?”

Dos **56 informantes** distribuídos entre as **21 localidades** da rede de pontos em Alagoas:

**36** responderam **ÍMÃ**, portanto, **64%**.

**20** responderam **AZOUGUE**, somando, **36%**.

No gráfico que se segue, as respostas aparecem distribuídas por localidade, sexo e faixa etária:

*A regionalidade na seleção lexical de informantes do  
Atlas Linguístico do Estado de Alagoas (ALEAL)*

| Rede de pontos<br>(aqui, em ordem alfabética) | Informante<br>Homem<br>30 a 50 | Informante<br>Mulher<br>30 a 50 | Informante<br>Homem<br>55 a 75 | Informante<br>Mulher<br>55 a 75 |
|---|--------------------------------|---------------------------------|--------------------------------|---------------------------------|
| Arapiraca (Ponto 10)                          | azougue                        | ímã                             |                                |                                 |
| Canapi (2)                                    | azougue                        | ímã                             |                                |                                 |
| Coruripe ( )                                  | ímã                            | azougue                         |                                |                                 |
| Delmiro Gouveia (1)                           | azougue                        | ímã                             |                                |                                 |
| Limoeiro de Anadia (13)                       | azougue                        | azougue                         |                                |                                 |
| Maceió (17)                                   | ímã                            | ímã                             | ímã                            | ímã                             |
| Maragogi (21)                                 | ímã                            | ímã                             |                                |                                 |
| Marechal Deodoro (16)                         | azougue                        | ímã                             | ímã                            | azougue                         |
| Palmeira dos Índios (11)                      | ímã                            | azougue                         |                                |                                 |
| Pão de Açúcar (5)                             | azougue                        | ímã                             |                                |                                 |
| Penedo (7)                                    | ímã                            | azougue                         | ímã                            | ímã                             |
| Piaçabuçu (8)                                 | azougue                        | ímã                             |                                |                                 |
| Pilar (15)                                    | ímã                            | ímã                             |                                |                                 |
| Piranhas (4)                                  | ímã                            | ímã                             |                                |                                 |
| Porto Calvo (20)                              | ímã                            | ímã                             | ímã                            | ímã                             |
| Quebrangulo (12)                              | azougue                        | ímã                             |                                |                                 |
| Santana do Ipanema (4)                        | ímã                            | azougue                         | azougue                        | ímã                             |
| S. Luís do Quitunde (19)                      | azougue                        | ímã                             |                                |                                 |
| S. Miguel Campos (14)                         | azougue                        | ímã                             | azougue                        | ímã                             |
| Traipu (6)                                    | azougue                        | azougue                         |                                |                                 |
| União dos Palmares                            | ímã                            | ímã                             | ímã                            | ímã                             |

---

**TOTAL DE RESPOSTAS À QUESTÃO 10 DO QFF DO ALEAL**

Informante Homem – 30 a 50 anos: 11 responderam **azougue**; 10 responderam **ímã**.

Informante Mulher – 30 a 50 anos: 06 responderam **azougue**; 15 responderam **ímã**.

-----

Informante Homem – 55 a 75 anos: 02 responderam **azougue**; 05 responderam **ímã**.

Informante Mulher – 55 a 75 anos: 01 respondeu **azougue**; 06 responderam **ímã**.

**TOTAL: 56 RESPOSTAS (100%)**

**ÍMÃ: 64%**

**AZOUGUE: 36%**

-----

Expostos os registros lexicais, passo à análise dos dois itens obtidos: **ímã** e **azougue**.

Cunha (2010, p.73), em seu *Dicionário Etimológico*, define assim as duas variantes lexicais:

- **Azougue** *sm.* ‘designação vulgar do mercúrio’ XIV; ‘fig. Pessoa muito viva e esperta’ 1813. Do ár. *az-zâ’uq*

- **Ímã** *sm.* ‘óxido magnético de ferro, que tem a propriedade de atrair o ferro’ *fig.* ‘Que atrai, atrativo’ | *iman* XVII | Do fr. ant. *aïmant*, deriv. do lat. vulgar *adimas -antis*, por *adamas -antis* ‘ferro duro, diamante’ (CUNHA, 2010, p.350).

O *Dicionário Eletrônico Houaiss de Língua Portuguesa* (2009) apresenta as seguintes acepções para o vocábulo **azougue**:

substantivo masculino (datação: 1344)

- 1 m.q. **mercúrio** (quím)
- 2 pessoa de muita vivacidade e inquietude
- 3 Regionalismo: Brasil. Uso: informal.  
aguardente de cana; cachaça
- 4 Rubrica: angiospermas.  
m.q. **mercurial** (‘designação comum’)

Quanto à etimologia da unidade lexical **azougue**, diz Houaiss (2009): “do árabe *az-azawq* ‘mercúrio’”.

Para o item lexical **ímã**, Houaiss (2009) traz essas acepções:

substantivo masculino (datação: 1665)

- 1 Rubrica: física.  
m.q. **magneto**
- 2 qualquer objeto (barra, ferradura etc.) imantado  
Ex.: *comprou vários í. para enfeitar sua geladeira*
- 3 Derivação: sentido figurado.  
aquilo que atrai

Houaiss (2009) descreve assim a etimologia do vocábulo: “fr. *aïmant* ‘diamante’ > *aimant* ‘mineral’, do lat. *adimas, antis*, prov. f. intermediária entre o lat. cl. *adâmas, antis* e o lat. vulg. *diamas, antis* (orig. do port. **diamante**), com recuo da tônica”.

Tendo, então, definidas as duas unidades em relação às respectivas etimologias e acepções, em primeiro lugar, chamo a atenção para o fato de que o vocábulo **azougue** aparece na língua portuguesa três séculos antes de **ímã**. Entretanto, a unidade **azougue** continua se mantendo com boa expressividade, e, sobretudo, num

Estado que se localiza no Nordeste, a primeira região a ser colonizada pelos europeus.

Em segundo lugar, aponto para o fato de ambas as variantes lexicais se referirem a metais. Diante dessas duas assertivas, ao menos uma questão se impõe: como se explica a produtividade da designação lexical **azougue**, três séculos mais antiga, registrada 20 vezes, em lugar de **ímã**, a resposta esperada?

Uma das respostas possíveis talvez seja: “[...] por analogia entre características atribuídas aos referentes nomeados.” (ISQUERDO, 2003, p.176): o mercúrio amalgama-se ao ouro, numa reação semelhante à atração que o ímã exerce sobre os metais.

Essa analogia torna-se viável ao se considerar alguns fatores históricos. O processo migratório no Brasil, especialmente, quando do período denominado “ciclo do ouro”, entre os séculos XVII e XVIII, aplica-se bem à analogia entre os dois vocábulos: as palavras viajam com os falantes. Segundo Bueno (1998a, 1998b), os colonizadores, desde o início, visavam a exploração do território brasileiro, sobretudo, alimentar suas Metrôpoles com metais preciosos da Colônia. Os aventureiros traziam com eles as técnicas para a extração de minerais e, atreladas às técnicas, suas respectivas denominações correntes à época. Assim, na bagagem dos exploradores, além de ferramentas e demais utensílios, estava o mercúrio, mais conhecido por **azougue**.

A descoberta de aluviões e minas de ouro e pedras preciosas, especialmente na Bahia e em Minas Gerais, fez com que afluísse para esses Estados um número incontável de aventureiros em busca de fortuna. No caso específico da extração do ouro, os garimpeiros valiam-se de métodos rudimentares (atualmente, ilegais): recolhido em bateias, o material era submetido ao calor do mercúrio derretido. Esse processo permitia que grãos de areia, pedregulhos e metais sem valor evaporassem junto ao mercúrio, restando o que era de interesse de todos: o precioso ouro, resistente às altas temperaturas.

Essa compatibilidade físico-química entre os dois metais – ouro e mercúrio – pode explicar a ocorrência significativa da variante regional **azougue**, encontrada não apenas em Alagoas, mas no Nordeste de modo geral. Esse fato linguístico é assinalado por

Aguilera e Yida (2008) e Mota (2008), em trabalhos que tratam do Atlas Linguístico do Brasil (ALiB), e encontrada agora, em 2014, nos inquiridos do *ALEAL*.

O Estado de Alagoas limita-se ao Sudoeste com a Bahia, Estado que produziu ouro e pedras preciosas na época do Brasil Colônia. A Chapada Diamantina era um corredor para onde acorriam garimpeiros de vários Estados, entre os quais, Alagoas. Finda a riqueza, restam resquícios do que parece ser o vocabulário do garimpo naquele período: **azougue**.

Em relação à ocorrência das duas variantes lexicais por localidade (vide a lista da rede de pontos), a unidade **ímã** foi registrada por todos os informantes da capital Maceió (04), em União dos Palmares (04 respostas), em Pilar (02), Maragogi (02), Piranhas (02) e Porto Calvo (04). Dessas 06 localidades, Piranhas é a mais distante de Maceió (cerca de 275 km a oeste da capital). Todas as outras cidades ficam no entorno.

Quanto à variante **azougue**, ela aparece ao menos uma vez nas 15 localidades restantes, sendo que em cidades mais recuadas de Maceió, com baixa população e economia do setor primário, por exemplo, Traipu (02 respostas) e Limoeiro do Anadia (02), além de Canapi (01 resposta) e Delmiro Gouveia (01), em pleno Sertão alagoano, a unidade lexical continua presente, sendo lembrada como segunda resposta mesmo pelos informantes que registraram a variante **ímã**.

Outra questão proposta aos informantes do *ALEAL* também suscita discussões acerca da regionalidade na seleção lexical dos falantes.

Questão 29 do QFF.

### **ÁRVORE**

“...o que é que dá sombra nas ruas, no campo/para o gado nos pastos?”

Um informante do sexo masculino, na faixa etária de 30 a 50 anos, em Delmiro Gouveia (Ponto 04), no Sertão alagoano, realizou:

### **Pé de pau.**

Uma das acepções para **pé** é: ‘cada exemplar de uma planta’ (FERREIRA, 1998); para **pau**: ‘pedaço de madeira’ (FERREIRA, 1998).

A influência do ambiente físico – o Sertão – é uma hipótese para explicar a variante lexical **pé de pau**, visto que essa região do bioma Caatinga sofre com prolongadas estiagens: numa árvore, sobressai-se muito mais o tronco do que a copa verde – na maior parte do tempo o informante vê um **pé de pau**.

Aguilera (1994), no *Atlas Linguístico do Paraná* (ALPR), também obteve a resposta **pé de pau** de ao menos um informante. Para a autora, a razão dessa designação pode advir de outra hipótese: o informante, sobretudo, aqueles que residem em localidades remotas (em relação à capital do Estado), sem influência da educação formal, vê no vegetal árvore, antes de tudo, a utilidade em seu dia a dia. Segundo Aguilera (1994), informantes com o perfil citado registram **pé de pau** porque, para eles, o que interessa na árvore é o tronco, a madeira: o **pau**.

Um informante do sexo masculino, na faixa etária de 30 a 50 anos, em Canapi (Ponto 02), também no Sertão, não produziu a variante lexical genérica **árvore**, e sim, nomes de duas variedades específicas do vegetal:

**Pé de figu** [ficus]

**Pé de algaroba.**

Outro informante do *ALEAL*, do sexo masculino, na faixa etária de 30 a 50 anos, em Quebrangulo (ponto 16), ao Norte de Alagoas, região do semiárido, também forneceu como resposta um nome específico de uma variedade de planta:

INF.- **Juá.**

INQ.- **Juá** é o quê?

INF.- Uma pranta.

((item não obtido))

**Pé de juá** [juá é o fruto do juazeiro].

Em ambas as localidades predominam, tanto na zona urbana quanto no meio rural, essas e outras espécies de árvores, próprias da região do Semiárido brasileiro e muito conhecidas dos nordestinos.

Com exceção da Zona da Mata, a vegetação de boa parte do Estado de Alagoas – Agreste e Sertão – é a Caatinga, onde predominam cactos e árvores espinhosas: a variante **pé de pau** está em conformidade ao meio físico do informante. O mesmo acontece em relação às designações **pé de ficus**, **pé de algaroba** e **pé de juá**: é a presença da regionalidade no léxico dos informantes.

### **Azougue, pé de pau: indicações da regionalidade no léxico**

As análises apresentadas corroboram as reflexões de Biderman (1998, p. 117, grifo do autor) sobre o processo de nomeação e da designação da realidade:

[...] o conceito (significado) é tributário de uma realidade que o antecede e precede, realidade essa que nossa percepção/cognição percebe e interpreta, criando o *objeto mental* ou *unidade cultural* ao qual atribuímos um nome, isto é, palavra ou significante.

Acerca do entorno dos informantes, Isquierdo (2003, p.165-166) observa que “[...] o vocabulário de uma comunidade linguística reflete as diferentes idiossincrasias que marcaram a formação étnica dessa comunidade.” Isquierdo (2003) entende por comunidade linguística quaisquer agrupamentos humanos que possuem um código verbal comum e, ainda que não seja de uso exclusivo de todos os membros do grupo, acaba por se impor aos demais, ou à maior parte deles, constituindo-se em elo: esse vínculo marca a solidariedade entre os falantes.

A dinâmica da história das comunidades linguísticas, ou seja, da humanidade em qualquer tempo e espaço, tem demonstrado que se reflete direta ou indiretamente na vitalidade de algumas unidades léxicas. Os fatores históricos, políticos e culturais cortejam de

perto a conservação ou a mudança da língua. Nesse sentido, vale citar Isquierdo (2003, p.180):

Assim, o novo e o antigo coexistem sem se sobreporem. Investigar uma língua é também investigar uma cultura, considerando-se que o sistema linguístico, nomeadamente o nível lexical, armazena e acumula as aquisições culturais representativas de uma sociedade.

Não se trata aqui de sugerir e, nem, tampouco, afirmar, que as unidades léxicas **azougue** e **pé de pau** são exclusivas do Estado de Alagoas, ou, ainda, da região Nordeste, pois o tema regionalismo, ou seja, o que é ou o que não é regional abrange estudos mais específicos, o que não é a proposta do presente trabalho.

Dessa feita, trata-se de constatar que alguns dos informantes do *ALEAL* apresentaram as referidas designações por não conhecerem outras variantes para as questões inquiridas. Sendo assim, infiro que as duas variantes são regionais, mas também, e por que não, podem ser encontradas em outras regiões e Estados do Brasil, pois nunca é demais lembrar que as palavras não são estáticas e muito menos sedentárias: elas viajam, às vezes elas desaparecem, e quando se pensa que não: cá estão elas de volta.

## REFERÊNCIAS

AGUILERA, V. de A. **Atlas linguístico do Paraná (ALPR)**. Curitiba: Imprensa Oficial do Estado, 1994.

AGUILERA, V. de A.; YIDA, V. Projeto ALiB: uma análise das respostas e das não-respostas de informantes das capitais. **Signum**, Londrina, v.11, n.2, p.15-31, 2008.

BIDERMAN, M. T. C. As ciências do léxico. In: OLIVEIRA, A. M. P. P. de; ISQUERDO, A. N. (Org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. 2.ed. Campo Grande: Ed. da UFMS, 2001. p.13-22.

BIDERMAN, M. T. C. Dimensões da palavra. **Filologia e Linguística Portuguesa**, São Paulo, n.2, p.81-118, 1998.

BUENO, E. **A viagem do descobrimento**: a verdadeira história da expedição de Cabral. Rio de Janeiro: Objetiva, 1998a. (Coleção Terra Brasilis, v.1).

BUENO, E. **Náufragos, traficantes e degredados**: as primeiras expedições ao Brasil (1500-1531). Rio de Janeiro: Objetiva, 1998b. (Coleção Terra Brasilis, v.2).

CUNHA, A. G. da. **Dicionário etimológico de Língua Portuguesa**. 4.ed. Rio de Janeiro: Lexikon, 2010.

DAUZAT, A. **La géographie linguistique**. Paris: Flammarion, 1922.

FERREIRA, A. B. de H. **Dicionário Aurélio Eletrônico**. Rio de Janeiro: Nova Fronteira, 1998.

HOUAISS, A. **Dicionário eletrônico Houaiss de língua portuguesa**. São Paulo: Objetiva, 2009.

ISQUERDO, A. N. Léxico em tempo e espaço: a questão dos regionalismos. In: MARIN, J. R.; VASCONCELOS, C. A. (Org.). **História, região e identidades**. Campo Grande: Ed. da UFMS, 2003. p.165-181.

ISQUERDO, A. N.; OLIVEIRA, A. M. P. P. de (Org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2.ed. Campo Grande: Ed. da UFMS, 2001.

MOTA, J. A metodologia na pesquisa geolinguística: o Questionário Fonético-Fonológico. **Revista ProLíngua**, João Pessoa, v.2, n.2, jul./dez. 2008. Disponível em: <<http://www.okara.ufpb.br/ojs/index.php/prolingua/article/download/.../7605>>. Acesso em: 02 maio 2014.

NASCENTES, A. **Bases para a elaboração de um atlas linguístico do Brasil**. Rio de Janeiro: Casa de Rui Barbosa, 1958.

*A regionalidade na seleção lexical de informantes do  
Atlas Linguístico do Estado de Alagoas (ALEAL)*

THUN, H. La géographie linguistique romane à la fin du XXe siècle. In: CONGRES INTERNATIONAL DE LINGUISTIQUE ET DE PHILOGIE ROMANES, 22., 1998, Bruxelles. **Actes...** Tubingen: Niemeyer, 2000. p.367-388.



# CRIAÇÕES LEXICAIS LITERÁRIAS: A IRREVERÊNCIA EXPRESSA PELAS CRIAÇÕES LEXICAIS NA POÉTICA DE JOÃO CABRAL DE MELO NETO<sup>1</sup>

Rosana Maria Sant'Ana COTRIM

## As criações lexicais literárias no âmbito da neologia estilística

Tradicionalmente, a neologia tem sido conceituada como a capacidade natural de criação e incorporação de novas unidades léxicas numa língua para a nomeação de novas experiências (fatos ou objetos). Isso é o que garante, em regra, a ampliação e/ou renovação do léxico de uma língua.

Todavia, ao compreender a neologia como uma demonstração da “criatividade lexical”, Guilbert (1975) concebe, para além da neologia denominativa, uma neologia estilística, cujos propósitos estão relacionados ao aspecto expressividade do/no enunciado. A neologia estilística é definida por ele, portanto, como “[...] uma forma especial ou particular da expressividade da palavra em si ou de uma frase pela presença de uma palavra nova com o objetivo de traduzir de maneira original ou inédita certa visão pessoal do mundo.” (GUILBERT, 1975, p.44, tradução nossa).

---

<sup>1</sup> Este trabalho é parte constitutiva de nossa tese de doutorado defendida junto ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Universidade Estadual Paulista, Campus de Araraquara (UNESP/FCLAr), intitulada *As criações lexicais na poética de João Cabral de Melo Neto: contribuições aos estudos do léxico no discurso literário*, adaptada para esta publicação.

Sob esta perspectiva, admite-se que a neologia estilística, presente, sobretudo, nos discursos humorísticos, jornalísticos e nos literários, está mais para atender à discursividade do texto do que para a ampliação e/ou renovação lexical, propriamente ditas. Ora, tendo em vista que ela serve exclusivamente para executar uma tarefa expressiva que a atualiza no discurso, ela apresenta uma “existência efêmera” (MARTINS, 2000, p.111), fadada a valer pelo seu “efeito de momento” (CÂMARA JUNIOR, 2004, p.63). Isto é, a cada realização a neologia estilística é capaz de instaurar novos sentidos, de modo a levantar novas possibilidades enunciativas relacionadas aos respectivos contextos e às possíveis leituras que lhe cabem. Sem dúvida, isso garante à unidade nova de cunho estilístico uma flexibilidade interpretativa resultante do leque conceptual que o discurso literário lhe é capaz de permitir; em contrapartida, restringe seu emprego àquele discurso em que está inserida.

No que concerne especificamente ao discurso literário, vale ressaltar que, substancialmente, os processos de formação de palavras têm atendido às necessidades expressivas de escritores desde os tempos mais remotos da história da língua portuguesa. E, no tocante à produção literária do Brasil, foi, sobretudo, no Modernismo, quando os escritores buscavam reforçar seu intento criativo e, principalmente, demonstrar uma nova postura linguística frente aos seus antecessores, que as criações lexicais oriundas do domínio literário ganharam relevo e reconhecimento. Contudo, muito mais do que, em princípio, a crítica e o aviltamento da língua oficial para a implantação de uma língua que se pretendia nacional e, por essa via, a construção de uma identidade, o emprego dos neologismos na literatura modernista brasileira como um todo obteve magníficos resultados expressivos.

### **Aspectos metodológicos da análise: autor, obra e atestação das unidades lexicais da poética de João Cabral**

João Cabral de Melo Neto é um escritor pernambucano, teoricamente enquadrado na terceira fase do Modernismo no Brasil e

tem sido considerado pela crítica um dos maiores poetas brasileiros do século XX. Ele tornou-se (re)conhecido pelo rigor formal de seus versos, o que o situa junto a significativas tendências do pensamento da arte moderna herdeira, entre outros, de Mallarmé, Valéry e Mondrian.

A proposta cabralina de revalorização do cuidado com a linguagem e de uma expressão poética mais disciplinada em relação às primeiras produções da estética modernista, somada a uma perfeibilidade da forma, conferem-lhe uma composição originalmente mais racional, como o bem revela sua produção de crítica literária:

A poesia me parece alguma coisa de muito mais ampla: é a exploração da materialidade das palavras e de possibilidades de organização de estruturas verbais, coisas que não têm nada a ver com o que é romanticamente chamado de inspiração ou mesmo intuição [...] o exercício da poesia [deve ser visto [...]] como exploração emotiva do mundo das coisas, e como rigorosa construção de estruturas formais lúcidas, lúcidos objetos de linguagem. (MELO NETO, 1998, p.135).

Em tese, o poeta propõe a superação da dicotomia entre expressão e construção, que é considerada pela crítica literária como uma das maiores de suas qualidades. De fato, a combinação entre estes dois elementos de forma única e rica consubstancia sua emancipação e a da literatura frente aos seus antecessores.

Na verdade, essa “construção” poética em João Cabral opera-se sob a consciência e o olhar crítico do poeta ora para a própria linguagem, ora para a realidade social ou, em geral, para a combinação de ambas. Por essa via, sua produção encontra-se sustentada em duas tendências que se articulam: i) a poesia como resultado de um rigoroso trabalho formal; ii) a consciência de que a crítica histórico-social só se torna expressivamente relevante em poesia quando consegue mobilizar uma forma escritural também crítica.

Esse quadro torna possível, em síntese, a consideração de que a presença recorrente das criações lexicais na composição poética

cabralina e a expressividade por meio delas alcançada são fundamentais também à construção do sentido; o que, segundo Cotrim (2012), permite a inclusão do poeta no rol dos escritores da literatura modernista brasileira que recorrem sistemática e originalmente às criações lexicais.

Quanto à temática cultivada por João Cabral, esta tende a girar em torno de suas grandes paixões: o Nordeste, onde nasceu e viveu boa parte de sua vida em meio à família, aos engenhos de cana e à seca com suas mazelas; a Espanha, mais especificamente a Andaluzia, onde exerceu funções diplomáticas durante muitos anos de sua vida profissional e donde, por admiração e respeito, adquiriu e cultivou influências culturais; o elemento feminino, ao qual se entrega e através do qual canta amores, paixões, a sedução do corpo e outros prazeres; e, por fim, a própria linguagem na qual, por procedimento metalinguístico calcado na “construção” linguístico-discursiva, estão inscritas as criações lexicais.

O levantamento e análise das unidades léxicas neológicas da poética de João Cabral foram executados, neste trabalho, a partir de um compêndio de suas obras, publicado em dois volumes intitulados, respectivamente, *Serial e antes* (MELO NETO, 1997a) e *Educação pela pedra e depois* (MELO NETO, 1997b), compostos de vinte livros de poesias. Deste *corpus* de extração foram consideradas 451 unidades léxicas novas, criadas a partir de variados processos de formação de palavras, que constituíram o *corpus* de análise denominado **Neologismos na Poética de João Cabral (NPJC)** (COTRIM, 2012).

A atestação de tais unidades, não obstante a raridade de desneologização das criações lexicais literárias e, por conseguinte, a excepcionalidade de sua inserção num dicionário de língua, foi feita a partir do critério lexicográfico. Consideradas as limitações que o cercam, o critério lexicográfico permite o estabelecimento de um parâmetro que possa servir de referencial temporal, espacial, social, cultural, ou outros que corroborem para o assentamento da objetividade e precisão desejadas nos resultados da pesquisa. Em outras palavras, este critério é capaz de conferir a garantia, ainda que lábil, de que as palavras fazem parte do léxico da língua em questão ou

têm caráter neológico, por suas respectivas incursões ou não no dicionário.

Foram tomados, portanto, como *corpora* de exclusão quatro dicionários de língua, cujo parâmetro de escolha foi o temporal, visando a uma maior cobertura lexicográfica em relação à época das produções do autor. Há que se considerar, ainda, que tais dicionários configuram quatro obras de grande representatividade no português brasileiro desde aproximadamente a terceira fase da literatura modernista brasileira até a atualidade. São eles: o *Novo Dicionário Aurélio da Língua Portuguesa* (1975), o *Dicionário Aurélio da Língua Portuguesa* (2010a), o *Dicionário Houaiss de Língua Portuguesa*, versão eletrônica 3.0 (2009) e o *Dicionário Aurélio da Língua Portuguesa*, versão eletrônica 7.0 (2010b).

### **Irreverência e expressividade das/nas criações lexicais cabralinas**

A irreverência e todo tipo de expressividade que lhe é decorrente, presentes nas criações lexicais cabralinas, são oriundas, em regra, da criatividade intrínseca a alguns processos – como no caso dos amálgamas –, e da criatividade resultante da “anormalidade” de execução do processo que as origina – como no caso das derivações e composições –, seja pela aparente incongruência entre o sentido da base e a ação criada, seja pela transgressão às regras pré-estabelecidas ou pela proposição de formas esdrúxulas concorrentes às já existentes em uso na língua comum.

Os amálgamas (também conhecidos como mesclas, cruzamento vocabular, palavra *portmanteau*, palavra *valise* ou *blends*) manifestam, segundo Alves (2002, p.69), um tipo de redução em que “[...] duas bases – ou apenas uma delas – são privadas de parte de seus elementos para constituírem um novo item léxico: uma perde a sua parte final e a outra, sua parte inicial.” De modo que, eles comportam uma estrutura que favorece a criatividade, em especial, a expressa pelo sarcasmo, pela ironia, etc. Isso decorre da sua própria estrutura que se utiliza da intersecção de palavras por concatenação não-linear de bases em que a fusão propicia alguns efeitos prosódicos na combinatória sêmica. A mescla das bases é regida, sobretudo-

do, pela semelhança fônica entre elas em que se admite a exclusão da parte final de uma base e a inicial de outra, ou de apenas uma delas, desde que satisfaça as condições prosódicas da construção léxica e provoque o sentido desejado pelo inusitado da combinação, como, por exemplo, em **portunhol, espanglês, brasiguaio, apertamento**, entre outros.

Estas peculiaridades têm servido como recursos discursivos da linguagem literária e, principalmente, da linguagem publicitária e/ou jornalística como um filão gerador dos efeitos de sentido de humor e ironia nestes gêneros do discurso. Dada sua estrutura, os amálgamas são considerados criativos para a combinação sêmica por introduzir carga emocional ao discurso, o que os situam na esfera discursiva da língua. Tanto que, dentre os estudiosos da neologia, há quem os considere, um processo “deformacional” de construção de palavras, “[...] já que [junto com as truncações, as siglas e os acrônimos] constituem uma deturpação da integridade morfológica dos radicais envolvidos na construção dos seus produtos.” (CORREIA; ALMEIDA, 2012, p.56, colchetes nossos), ou um processo “marginal”, porque externam “[...] o ponto-de-vista do falante a respeito de algo ou alguém.” (GONÇALVES, 2006, p.222).

As derivações e composições, quando empregadas no discurso literário, tendem a buscar formas alternativas de expressividade, sobretudo em função da referida “anormalidade”.

Aplicados, porém, à língua comum, em geral, estes processos de formação de palavras seguem os padrões estabelecidos pelo sistema linguístico. Ora, a língua portuguesa, como qualquer língua natural, comporta a capacidade e meios para a construção de palavras. Ela dispõe, portanto, de um conjunto de regras interiorizadas e partilhadas pelos seus falantes que regulamenta a combinação de elementos pré-existentes – unidades lexicais de base e afixos – para a formação das unidades novas de estrutura totalmente transparente, ou seja, itens ou palavras cuja estrutura morfológica é perceptível e cujo significado é coerente com essa estrutura (CORREIA; LEMOS, 2009). Em outras palavras, em geral, na língua comum, a manipulação dos processos de formação de palavras segue o padrão

pré-estabelecido e/ou disponível pelo sistema da língua.

Ocorre que, no discurso literário, os processos de criação de palavras tendem a ser empregados na contramão desta lógica. Ou seja, o empenho pela busca de estratégias de manipulação e forjamento da expressividade nas unidades criadas, em geral contraria o sistema, com vistas a resultados e efeitos de sentidos específicos em relação ao enunciado em que estas aparecem.

Isso é, pois, o que se observou em NPJC pela utilização de amálgamas, derivações e composições, entre outros processos que comportam esta natureza e/ou permitem tais procedimentos de intervenção no sistema. São eles que, nesta análise, por via de regra, se considera os grandes responsáveis pelos efeitos de humor, ironia, sarcasmo, excentricidade e pejoratividade existentes e analisados na poética cabralina.

Veja-se, a seguir, exemplos de alguns destes casos com a interpretação da expressividade e do efeito de sentido em seus contextos discursivo-enunciativos, seguidos de respectivas abonações.

#### (01) **brasilento**

A unidade léxica **brasilento** foi criada a partir da mescla do gentílico **brasileiro** + o adjetivo **lento**. Trata-se, portanto, de um caso de amálgama, cujas bases sofrem as alterações necessárias e pertinentes à sua composição e cuja interpretação, no nível discursivo da unidade criada, sugere uma chamada de atenção para a imagem estereotipada do Brasil no contexto mundial, cultural e intelectualmente.

No contexto em que está inserido, **brasilento** tem o propósito de revelar a peculiaridade da língua portuguesa falada no Brasil que, por apresentar um ritmo mais lento que a língua portuguesa falada em Conacri (Nova Guiné), lembra, conforme o enunciador, a leveza e a suavidade da brisa das praias pernambucanas de Piedade, Pina e Olinda. Em regra, o português falado em África apresenta estrutura e pronúncia semelhantes às do português europeu. Contextualizada, portanto, esta comparação tende a estender-se, pejorativa e sarcasticamente, para além da aceleração de seu ritmo que decorre em função, principalmente, da neutralização e/ou exclusão da vogal pré-tônica do português europeu, aos aspectos

relacionados à evolução e desenvolvimento cultural e intelectual, também mais lento na colônia em relação à metrópole.

Além disso, a unidade criada pode remeter a outra compreensão linguística. Isto é, pode sugerir uma derivação sufixal a partir do substantivo próprio Brasil e o acréscimo do sufixo latino de valor nominal **-ento** (indicador de tendência, estado), para a formação do adjetivo **brasileto**. É preciso se considerar que no português coloquial do Brasil há grande tendência de utilização deste sufixo nos contextos pejorativos como em **sarnento**, **fedorento**, **amareloento**, etc. Para Basílio (2007), a pejoratividade é uma expressão comum da atitude subjetiva na caracterização de um ser, e a marca morfológica da pejoratividade é muito significativa nos sufixos (e prefixos) pejorativizadores, como ocorre em **-ento**. O que, ainda assim, corrobora a ideia de que a unidade criada, seja concebida por este ou pelo processo anterior, representa, naquele contexto, conforme o enunciador, a subcondição do português falado na colônia e de tudo que cerca a relação entre esta e a metrópole.

**Abonação:** “Se o que ela diz me escapa, / seu ritmo, seu acento / são esses com que falo / o português *brasileto*” (Agrestes, VII, p.256).

## (02) **intestinar**

Em **intestinar** ocorre um caso de derivação pela formação de um verbo a partir da união do sufixo de primeira conjugação (**-ar**) a um nome, em que a ação criada subsume o sentido da respectiva base. Acontece que, de modo especial, este caso de sufixação causa certo estranhamento pela “anormalidade” da execução do processo, dada a distância ou a falta de relação entre o sema de origem e a ação criada. Nesse particular, é preciso se considerar que a adaptação semântica da base à ação em NPJC, em regra, advém de processos de metaforização (COTRIM, 2012).

Na unidade criada, o sentido da ação visceral, adquirido da peristalse<sup>2</sup>, é transportado para um ambiente contextual contras-

---

<sup>2</sup> Segundo Ferreira (2010b, p.1615), a peristalse é o “[...] movimento vermiforme, progressivo da musculatura de órgãos ocos, e que impulsionam para diante o conteúdo desses órgãos, em certos casos (fezes, urina), eliminando-o para o exterior.”

tante, se considerado do ponto de vista referencial. No contexto, a proposta de relação entre o movimento peristáltico e a ação interna do sal marinho ao corroer tudo que está próximo ao mar e do cupim ao desmanchar o canal e tudo que o cerca, consegue, pela irreverência, completar a imagem que passa o enunciador acerca de sua crítica ao sistema latifundiário canavieiro que “corrói” as relações sociais no Nordeste brasileiro.

Isto é, o contexto, o conhecimento que se tem acerca do poeta e a interpretação de sua obra levam à percepção do seu julgamento crítico pela relação acintosa estabelecida entre as querelas resultantes do sistema latifundiário com os excrementos ou matérias fecais expelidos pelos animais por vias naturais, após a ação peristáltica, sugerindo a inferência de que ambos seriam “fezes”.

**Abonação:** “No canal tudo se gasta / pelo miolo, não pela casca [...] // Tudo se gasta mas de dentro: / o cupim entra os poros, lento, / e por mil túneis, mil canais, as coisas desfaz e desfaz. // Por fora o manchado reboco / vai-se afrouxando, mais poroso, / enquanto desfaz-se, **intestinal**, o que era parede, em farinha.” (Quaderna, VI, p.220-221).

### (03) **fluvante** e (04) **flutual**

Em **fluvante** e **flutual** têm-se, igualmente, casos de sufixação em que há falta de relação entre o elemento determinante e o determinado. Ou seja, nestas unidades criadas, o sufixo **-ante**, cujo papel de agente da ação prevê sua união a um verbo, aparece adicionado a um radical substantivado, enquanto que o sufixo **-al**, cujo papel de indicador de relação prevê sua adição a nomes, aparece ligado a um verbo.

Observado o contexto, as unidades criadas **fluvante** e **flutual** estão dispostas lado a lado e equivalem a uma inversão propositada do par correspondente **flutuante** e **fluvial**, unidades já dicionarizadas e consagradas pelo uso.

Compreende-se que esta inversão do final dos pares reforça o aspecto de irreverência da sufixação em NPJC por lembrar os jogos de palavras tão comuns nas brincadeiras infantis que criam uma infinidade de pares invertidos com o intuito de provocar riso, como em “garota **bonita** e **amorosa**” que pode virar “garota **bonitosa** e

**amorita**” ou em “**chefatura de polícia**” que pode virar “**politura de chefalícia**”, e outros.

**Abonação:** “[...] quando ao catar palavras: / a pedra dá à frase seu grão mais vivo: / obstrui a leitura **fluviante, flutual**, / açula a atenção, isca-a com o risco.” (A Educação pela Pedra, VII, p.17). De modo que tais contravenções sugerem, antes, estratégias linguísticas e discursivas.

(05) **itamaratício**

Em **itamaratício** há uma derivação a partir da união do sufixo **-ício**, indicador de relação ou de referência que, ligado ao ser/coisa a que se refere, serve para a formação de um adjetivo. Na língua geral, a base para esse tipo de formação, em regra, deve ser um substantivo comum, com tendência a designar algo que se intensifica. Contudo, no contexto, ele tem por base o substantivo próprio **Itamaraty**, referindo-se ao Palácio do Itamaraty do Rio de Janeiro, que fora sede do Ministério das Relações Exteriores de 1889 a 1970.

Conforme o enunciado, a unidade criada refere-se a um qualificativo dirigido a Cícero Dias. Trata-se de um pintor renomado com obras reconhecidas internacionalmente, que, ao estabelecer-se no Rio de Janeiro, no início de sua carreira, recebe críticas e rejeições, pelo seu então estilo revolucionário. Sabe-se que o período no qual estaria enquadrada a estada de Cícero Dias naquela cidade corresponde ao período de funcionamento do Palácio Itamaraty. Entende-se, portanto, que a unidade lexical **itamaratício**, criada para qualificar (ou “desqualificar”) Cícero Dias, está, então, referindo-se à relação de sua pessoa, rude, incivil e descortês no momento de sua chegada naquela cidade, ao papel diplomático, embora a princípio ignorado e desprezioso, que só mais tarde ele viria desempenhar entre o Brasil e o mundo, em virtude do ulterior reconhecimento de sua arte, pela visibilidade de suas obras.

**Abonação:** “Cícero Dias, quando foi / de Pernambuco para o Rio, / anti(e)ducado, sem prever / que o seria, **itamaratício**, / traçou na parede do hotel, / de onde a porta dava, uma seta, / que pela mão levava a vista / da visita à palavra MERDA.” (A Escola das Facas, VII, p.114).

(06) **perguntadeiro**

Esta unidade lexical criada é mais um caso de derivação sufixal, em que o sufixo nominal **-eiro** (indicador de qualidade ou estado) ligado a uma base verbal forma um adjetivo. Compreende-se, pelo enunciado, que a unidade criada reafirma a intenção do poeta de salientar o sentido depreciativo e irônico por ela expresso no discurso em que se insere. Prova disso é que se encontra na língua comum forma concorrente e já dicionarizada, a partir de outro sufixo, mais indicado pela tradição gramatical – **perguntador**.

No contexto, **perguntadeiro** refere-se ao sobrinho que se mostrava interessado e ávido das histórias contadas pelo tio. O que a diferencia da forma concorrente é sua capacidade de transmitir certo incômodo por parte do tio em relação ao sobrinho, dada exacerbada curiosidade deste nas rodas de conversa. Nesse particular, lembra-se que o sufixo **-eiro(a)** traz em sua significação, pela recorrência do ato por ele expresso, a ideia de menosprezo, desdém, depreciação nas unidades criadas como em **faladeira, zombeteira, mexeriqueiro**, etc.

**Abonação:** “O sobrinho ouvia-o atento, / e um tanto **perguntadeiro**, / do Sertão que havia atrás / da Mata doce [...]” (Museu de Tudo, VII, p.116).

(07) **todoaberto**, (08) **almiaberto** e (09) **calmoaberto**

Estas unidades, cuja disposição adjacente forma um bloco, têm parte no mesmo contexto, dimanam os mesmos efeitos de sentido e todas foram criadas pelo processo de formação de palavras. Tratam-se de casos de composição morfológica pela ligação de radicais latinos e gregos, com o elemento pospositivo **-aberto**, do particípio do verbo **abrir**. Segundo Houaiss (2009, versão eletrônica), este elemento pospositivo é usado em cultismos do séc. XVI em diante para a construção de unidades léxicas como **boquiaberto, entreaberto, perniaberto**. Acrescenta o dicionarista que as unidades deste gênero construídas com o auxílio da vogal de ligação **-i-** simulam latinismos.

De acordo com Correia e Lemos (2009), a composição morfológica consiste na construção de palavras compostas a partir de unidades não-autônomas, essencialmente extraídas de línguas anti-

gas, principalmente do latim e do grego e também de elementos emprestados de línguas influentes como a de origem anglo-americana, geralmente, adaptadas ao sistema fonológico do português. Por essa razão, ela acaba por receber várias denominações: composição culta, composição neoclássica, composição erudita, ou outras. Este tipo de composição é bastante empregado no meio científico e/ou técnico em terminologias específicas, porque tanto demanda habilidade e conhecimento linguístico para o manuseio destes radicais que não são muito comuns a todos os falantes, quanto pode resultar unidades internacionalmente traduzíveis e adaptáveis, facilitando a intercambialidade linguística nestas áreas específicas.

Em **todoaberto** o elemento pospositivo une-se ao adjetivo **todo**, originário do latim **totu**, que significa “completo, inteiro, total”. Em **almiaberto**, ao substantivo alma, também originário do latim *anima*, que significa “conjunto das funções psíquicas e dos estados de consciência do ser humano que lhe determina o comportamento; espírito”. E, em **calmoaberto**, ao elemento de composição antepositivo **calm-**, originário do grego *kaíuma*, cujo significado era “calor ardente”, passando ao português pelo latim tardio como *cauma* para adotar o sentido de “serenidade de ânimo, sossego, tranquilidade”.

Por comportar essa estrutura clássica, erudita, pouco usada e, por sua vez, estranha, beirando à excentricidade, o bloco cumpre o papel contextual de apresentar características peculiares estereotipadas de Comendadores Eclesiais, quando sentados à mesa de refeição. É consensual a fama de gula herdada de tempos áureos da Igreja Católica na Idade Média pelos seus representantes. No contexto, pois, o enunciador mostra, pela presença das composições morfológicas, a bizarrice da cena: um ambiente clássico e tranquilo, que, num primeiro momento, soa como comunal e partilha, no sentido religioso dos termos, e, num segundo momento, é desbancado por comportamentos animais e grotescos de figuras religiosas famintas e sedentas indispostas a dividirem seu alimento. A cena reporta o enunciatário à imagem formada que se tem da Igreja sobre a questão da partilha que muito se tem defendido e pouco se tem praticado, a começar pelos seus

representantes. Os compostos auxiliam, portanto, o jogo linguístico da cena enunciada.

Cumprido esclarecer que o poema está dividido em duas estrofes, correspondendo às duas fases do jantar de tais Comendadores Eclesiais: a primeira refere-se ao momento calmo, de sucumbência e espera do serviço de mesa e a segunda, à avidez da “devoração” do alimento. Em participando da primeira estrofe, portanto, as unidades criadas **todoaberto**, **almiaberto** e **calmoaberto** expressam a ironia do enunciado.

**Abonação:** “Duas Fases do Jantar dos Comendadores [título do poema] / Assentados, mais fundo que sentados, / eles sentam sobre as super-cadeiras: / cadeiras com patas, mais que pernas, / e de pau-d’áço, um que não manqueja. / Fundassentados se abrem: **todoabertos** / ante a mesa, ainda uma mesa-de-espera, / e pré-abertos, para as ótimas opções / se mini-ultimando na cozinha e adega; / **almiabertos** para que nada lhes escape / na escolha entre tudo o que a bandeja; / **calmoabertos**, podendo escolher o tudo / do que oferecem bandejas e igrejas. // Assentados fundo, ou fundassentados, / à prova de qualquer abalo e falência, / se centram no problema circunscrito / que o prato de cada um lhe apresenta. / Fundassentados se fecham: revestindo, / contra tudo em torno, bem carangueja, / a carapaça que usam, dentro do prato / e de outros círculos e áreas defesas; / se fecham: erguem fronteiras no prato, / se entrincheiram atrás das fronteiras; / se fecham até de poros, o que só fecham / quando ouvem sermão de outras igrejas.” (A Educação pela Pedra, VII, p.38-39).

### **Considerações finais**

Esta análise permitiu a observação de que as criações lexicais presentes na poética de João Cabral de Melo Neto participam ativamente do processo de “construção” da sua composição linguístico-discursiva. Isto é, que a presença significativa de criações lexicais por diferentes processos de formação de palavras em sua poética configura um elemento caracterizador da sua produção, qual seja a busca da expressividade por meio tanto das unidades

léxicas criadas quanto dos processos que as criam e, por essa via, a superação da dicotomia entre expressão e construção.

Quanto aos efeitos de sentido por elas obtidos, destacam-se nesta análise a irreverência geradora do humor, da ironia, do sarcasmo, da excentricidade e da pejoratividade, reconhecendo-se que estes elementos contribuem sobremaneira para o tal processo de construção da linguagem e do discurso cabralino. Ora, a irreverência presente nas criações lexicais analisadas e os efeitos de sentido que elas engendram no discurso em que se inserem atuam diretamente na crítica social proposta pelo poeta, de modo a contribuir na articulação das duas tendências que permeiam toda sua produção, ou seja, compondo a poesia concebida como resultado de um rigoroso trabalho formal com a linguagem e gerando a consciência de que a crítica histórico-social só se torna expressivamente relevante, em poesia, quando mobiliza uma forma escritural também crítica.

## REFERÊNCIAS

- ALVES, I. M. **Neologismo**: criação lexical. 2.ed. São Paulo: Ática, 2002.
- BASÍLIO, M. **Teoria lexical**. 8.ed. São Paulo: Ática, 2007.
- CÂMARA JUNIOR, J. M. **Contribuição à estilística portuguesa**. Rio de Janeiro: Ao Livro Técnico Editora, 2004.
- CORREIA, M.; ALMEIDA, G. M. de B. **Neologia em Português**. São Paulo: Parábola, 2012.
- CORREIA, M.; LEMOS, L. S. P. de. **Inovação lexical em português**. Lisboa: Edições Colibri, 2009. v.4.
- COTRIM, R. M. S. **As criações lexicais na poética de João Cabral de Melo Neto**: contribuições aos estudos do léxico no discurso literário. 2012. 231f. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, Araraquara, 2012.

FERREIRA, A. B. de H. **Dicionário Aurélio da Língua Portuguesa**. 5.ed. Curitiba: Positivo, 2010a. Edição comemorativa aos cem anos de Aurélio Buarque de Holanda Ferreira.

FERREIRA, A. B. de H. **Dicionário Aurélio da Língua Portuguesa**. Versão eletrônica 7.0. Curitiba: Positivo, 2010b. Edição comemorativa aos cem anos de Aurélio Buarque de Holanda Ferreira. 1 CD-ROM.

FERREIRA, A. B. de H. **Novo Dicionário Aurélio da Língua Portuguesa**. Rio de Janeiro: Nova Fronteira, 1975.

GONÇALVES, C. A. Usos morfológicos: os processos marginais de formação de palavras em português. **Revista Gragoatá**, Niterói, n.21, p.219-241, 2. Sem. 2006.

GUILBERT, L. **La créativité lexicale**. Paris: Librairie Larousse, 1975.

HOUAISS, A. **Dicionário Houaiss da Língua Portuguesa**. Versão eletrônica 3.0. Rio de Janeiro: Objetiva, 2009. 1 CD-ROM.

MARTINS, N. S. **Introdução à estilística**: a expressividade na língua portuguesa. 3.ed. São Paulo: T. A. Queiroz Editor, 2000.

MELO NETO, J. C. **Prosa**. Rio de Janeiro: Nova Fronteira, 1998.

MELO NETO, J. C. de. **A educação pela pedra e depois**. Rio de Janeiro: Nova Fronteira, 1997a. v.1.

MELO NETO, J. C. **Serial e antes**. Rio de Janeiro: Nova Fronteira, 1997b. v.2.



# INFORMATIZAÇÃO E TRABALHO HUMANO EM TERMINOLOGIA: EXTRAIR TERMINOLOGIAS EM *CORPUS* É O MESMO QUE IDENTIFICÁ-LAS?

Maria José Bocorny FINATTO

## Introdução

O objetivo desta apresentação é refletir sobre o papel das diferentes metodologias para identificação de terminologias em *corpora* com o apoio de sistemas computacionais. Nessa direção, proponho o contraste de algumas práticas de trabalho influenciadas pela participação de recursos da Linguística de Corpus (LC) com metodologias provenientes de sistemas e de ferramentas gerados no âmbito do Processamento de Linguagem Natural (PLN), uma especialidade da Ciência da Computação ainda pouco conhecida pelos linguistas brasileiros, também denominada como Processamento de Língua Naturais ou Linguística Computacional. Essa identificação de denominações e de conceitos associados a um léxico especializado situa-se no âmbito da Terminologia e da Terminografia. A Terminologia é, *grosso modo*, a disciplina voltada para a descrição dos fenômenos das práticas textuais técnico-científicas, enquanto a Terminografia é a sua contraparte mais aplicada, vinculada à dicionarização de unidades léxicas dotadas de valor de termo.

Uma preocupação aqui será distinguir as práticas de LC das de PLN, ainda que suas intersecções sejam abundantes. A distinção visa demonstrar o grande potencial de ambas para o tratamento de fenômenos da comunicação em ciências e tecnologias.

Por sua vez, o PLN, conforme Lima, Nunes e Vieira (2007) relaciona-se ao tratamento computacional da informação em língua natural, à modelagem matemática da língua e à aquisição “computacional” do conhecimento linguístico, prestando inestimável auxílio em termos do acesso universal à informação, especialmente quando pensamos na miríade de dados disponíveis hoje na internet. Transformar uma massa gigantesca de dados relativamente dispersos, repetidos e até contraditórios em informações úteis é um dos grandes desafios atuais não apenas da área de PLN, mas de toda a Ciência da Computação.

Nesse cenário, no que diz respeito ao Brasil, reconhece-se que dotar o português do Brasil de recursos para o seu processamento, em suas particularidades, é um avanço importante a conquistar. Afinal, poder interrogar uma base de dados produzidos em português, utilizando palavras de busca em português ou encontrar uma boa ferramenta de tradução automática envolvendo o português do Brasil ou localizar representações de conhecimento em português são demandas cujo atendimento deve ser tão possível quanto hoje é para o inglês, a *lingua franca* da internet.

A Linguística de Corpus (LC), por seu turno, conforme Garcia (2012), apresenta-se como uma nova metodologia de investigação de línguas em uso no âmbito dos Estudos da Linguagem. É, ao mesmo tempo, uma metodologia, uma disciplina e uma abordagem teórica. Introduzida no Brasil por Sardinha (2004), firmou-se no nosso cenário já há quase 15 anos. Caracteriza-se pela utilização de grandes acervos de textos em formato digital – com dados oriundos de escrita e também de fala – cuja exploração é feita em larga escala e com apoio computacional. Esses acervos, criteriosamente reunidos para exemplificar um determinado tipo de uso de língua – conformando um *corpus* ou vários *corpora*, são os traços de identidade dessa relativamente “nova” Linguística.

Etapas recorrentes de uma investigação com LC, seja como metodologia ou abordagem, envolvem observação de palavras em contextos, ordenação das palavras presentes em um dado *corpus* por frequências de uso, observação de variedade do vocabulário (pela relação entre o número de palavras de um texto e a quantidade de palavras repetidas que ele tem), comparação de frequências de uso de determinadas palavras em diferentes textos, associações recorrentes entre palavras, entre outros. Entretanto, é importante notar que se vai muito além de “contar palavras”, pois além de aspectos quantitativos, consideram-se os aspectos qualitativos dos seus diferentes usos.

Para essa reflexão, penso que vale a pena situar, desde já, o pano de fundo da pergunta do título deste trabalho. Afinal, essa é uma pergunta que deve ser entendida como um questionamento integrado a um cenário determinado. Tal é o cenário atual de uma Informática e de um acesso à internet que permeiam a maioria das tarefas humanas, em especial, as tarefas de pesquisa científico-acadêmica. Nesse estado de coisas, conforme entendo, a relação entre a informatização e o trabalho humano (tornando-se preciso distinguir entre **trabalho humano** e **trabalho de máquina**) influencia o modo de agir do terminólogo ou terminógrafo e não deve ser vista como uma relação antagônica, mas, sim, como complementar. A complementaridade deve prevalecer, ainda que algum descompasso ou conflito de perspectivas seja natural e inerente.

Além disso, é importante ter em mente que, se as metodologias de LC e de PLN hoje geralmente subsidiam o trabalho terminológico, também há uma via inversa – a do trabalho do terminólogo subsidiar os enfoques da LC e do PLN. Isso ocorre quando informatas, por diferentes motivos, tratam especialmente de *corpora* técnico-científicos. Entretanto, mesmo com tal reciprocidade, algumas pessoas chegam a entender que se poderia substituir ou eliminar o terminólogo, especialmente quando se pensam em aplicações como recuperação de informação ou descrição de linguagens técnico-científicas. Já, sobre dispensar o informatas, pouco se ouve de comentários em grupos de pesquisa.

A despeito dessa percepção de obsolescência, tida em alguns momentos, de pouco valor para o trabalho humano do terminólogo (geralmente um estudioso com formação em Letras/Linguística), o método tradicional do trabalho terminológico segue sendo executado. Nele temos, via de regra, a leitura atenta de muitos textos de um dado domínio ou área de conhecimento. Trata-se de uma leitura profissional, em que se processam textos um a um, para que se chegue a poder enquadrar determinadas unidades lexicais em um dado sistema conceitual, previamente desenhado pelo terminólogo, a partir de cuidadosa pesquisa bibliográfica e de consulta qualificada a especialistas. Esse método tradicional permanece reconhecido como algo dotado de grande valor de qualidade, apesar do seu, em geral, alto custo econômico. Afinal, há aqui um trabalho intelectual refinado.

Nesse ponto, considero importante estarmos atentos para a necessidade de distinção e de valorização das diferentes abordagens para a extração e para a identificação de terminologias. É importante reconhecer as especificidades, os custos e os ganhos da geração de diferentes produtos de viés terminográfico, seja essa geração totalmente automática, feita por “humanos” assistidos por computador-ferramenta ou completamente “artesanal”.

Mesmo que pese bastante o fator custo e que a rapidez de resultados seja um “valor agregado” em qualquer iniciativa, que quanto mais automatizada melhor será, não se pode desprezar a qualidade do trabalho subjetivo, tampouco a percepção e a reflexão sobre as epistemologias associadas às práticas e métodos que se escolha seguir.

Para encerrar essa já longa introdução, penso que vale dizer, mais uma vez (cf. FINATTO, 2014), que a Terminologia não tem como objetivo único sustentar uma dicionarização peculiar, que é a Terminografia. Ultrapassamos o tempo dos terminólogos “fazedores” de dicionários técnicos. No ponto atual, a sua tarefa é sistematizar descrições de uso de língua em cenários comunicativos técnico-científicos de modo a embasar produtos altamente sofisticados, entre os quais estão dicionários, ontologias e representações de conteúdo que partem de bases textuais de diferentes perfis para

que um bom dicionário “aconteça”. Ainda assim, cabe dizer que do muito que se tem tratado sobre elementos gramaticais, lexicais, semânticos, discursivos, cognitivos e pragmáticos, o que abarca definições, fraseologias e tipologias textuais, além de terminologias, é que temos os melhores subsídios para a produção de glossários, dicionários e outros repertórios, mais ou menos cibernéticos, que realmente atendam às necessidades de seus usuários.

## **Extrair e identificar terminologias**

Neste *I Congresso Internacional Estudos do Léxico e suas Interfaces*, que destaca justamente os diferentes olhares sobre o léxico, é importante ter em conta que nos acercamos a um componente da língua cuja definição será tão diferente quanto tão diferente for o foco que se escolha para a tarefa. Conforme já ensinava Rey-Debove (1984), pode-se compreender o léxico desde os morfemas até as palavras, chegando-se às lexias, às construções multipalavra e ao seu estatuto semiótico. Enfim, como ela já nos ensinou “[...] o lexicólogo tem o dever, mais delicado ainda, de reconhecer a zona das trocas entre a língua e o mundo, que se faz essencialmente pelo léxico.” (REY-DEBOVE, 1984, p.52).

Em geral, para o cientista da Computação, o léxico de uma língua é um componente aproximado a um conjunto de peças (as palavras), combinadas ou justapostas ou encaixadas de acordo com uma gramática, normalmente entendida como um sistema de regras. Cada peça-palavra se combina com uma outra seguindo determinados padrões associativos e de frequência, sempre de acordo com suas funções num sistema de regras. Esse quadro, considerando um linguista que o contemple, evocará a analogia saussuriana do jogo de xadrez com o sistema da língua.

Para um linguista, o léxico, as palavras que integram um texto, seja um texto de jornal, um poema ou um texto técnico-científico, é tudo isso e muito mais. Assim, o jogo de xadrez torna-se apenas uma ilustração para uma relação hipercomplexa entre léxico e gramática, significados, significantes e valores dos signos linguísticos.

Pois bem, se a diversidade de olhares e de entendimentos sobre o léxico é abundante entre linguistas e não linguistas, parece uma contingência dessa diversidade que extrair terminologias de um dado texto seja algo diferente do que se entende do que seja a sua identificação. Automaticamente obtidas, via de regra, o que temos, na verdade, são listas de candidatos a termo ou perfis lógico-matemáticos preenchidos com determinadas palavras capturadas automaticamente de um texto de acordo com determinados critérios de frequência e de associação. Em geral, com a extração automática, obtém-se uma lista de expressões que precisará ser mais ou menos filtrada, por métodos estatísticos e por métodos linguísticos.

Esse processo, da extração à identificação, com a participação de linguistas e de cientistas da computação – com o aval final de especialistas do domínio de conhecimento que se esteja explorando, já foi descrito por diferentes autores, entre os quais, destacamos, no Brasil, Almeida, Aluísio e Teline (2003). Essas autoras, entre elas uma terminóloga experiente, já sublinhavam a validade e as vantagens da extração automática, apontando que as listagens de candidatos a termos, antes de serem enviadas aos especialistas da área cuja terminologia se pretende repertoriar, devem ser verificadas, por um terminólogo. Esse profissional os inserirá nos campos e/ou subcampos nocionais da estrutura conceitual de forma a facilitar a validação pelo especialista consultor (ALMEIDA; ALUÍSIO; TELINE, 2003).

Desse modo, combinar-se-iam critérios semânticos, estabelecidos pelo terminólogo, com critérios de frequência e de associação de palavras – estabelecidos pelas fórmulas computacionais que apontam prováveis termos em um dado *corpus*. Esse processo, semelhantemente descrito por Estopá Bagot (1999), Vivaldi e Rodriguez (2001) e L'Homme (2005), entre outros vários autores, mostra como da extração informática chega-se à identificação ou validação conceitual e terminológico-linguística, em um trabalho nada trivial de várias sinergias reunidas.

Assim, apenas nesta seção, já se evidencia uma resposta negativa para a pergunta colocada no título deste trabalho. Extrair faz a máquina, identificar faz o ser humano. Dito isso, é importante

demonstrar o que há mais envolvido nesse processo. Nessa direção, tratamos, a seguir, dos enfoques da LC e do PLN, situando as suas metodologias para subsidiar a tarefa de identificar terminologias em acervos textuais de formato digital especialmente reunidos para esse fim.

## **Linguística de Corpus (LC) e terminologias**

Trabalhar com a linguagem técnico-científica espelhada nos corpora tal como preconiza a LC é algo reconhecidamente importante em todas as perspectivas da Terminologia – dos herdeiros da Teoria Geral da Terminologia (TGT) às perspectivas descritivas, tais como a Teoria Comunicativa da Terminologia (TCT) e outras perspectivas. Trazer a LC para a pesquisa de terminologias significa aderir a uma metodologia terminográfica associada a uma concepção de língua particular. Na LC, conforme se trata do tema no Brasil (SARDINHA, 2004), a língua é entendida como um sistema probabilístico de combinações de unidades, realizado tal sistema em meio a uma grande diversidade de usos que se poderá descrever. Tal diversidade de usos demandará uma observação extensiva da variação de usos, também das terminologias, ao longo da língua da qual se trate.

Nesse tipo de enfoque, “a verdade” sobre o que é uma língua e sobre o que são as suas palavras estará nos seus *corpora*. De tal sorte que a especificidade de um dado termo ou expressão terminológica deverá estar posta e atestada nos seus diferentes usos ao longo dos registros especializados e não especializados. Também, para identificar termos, valerá o axioma da LC “dize-me com quem andas e te direi quem és”, visto que uma palavra se definirá por suas companhias ao longo de infinitas frases que se observam com apoio de técnicas estatísticas e de programas de computador especialmente desenvolvidos.

Num trabalho com apoio da LC, tal como o de Teixeira (2010), serão comparados dois *corpora*: o *corpus* da área de conhecimento que se explora – a Oncomastologia – e o *corpus* externo à área em foco. O primeiro é chamado de *corpus* de estudo; o segundo é o

*corpus* de referência, cuja dimensão deve ser pelo menos cinco vezes maior do que a do *corpus* de estudo. Como ambos serão contrastados, para chegar-se às especificidades da Oncomastologia, o *corpus* de referência deve, preferentemente, não conter textos do gênero técnico-científico nesse tema. Dessa comparação, feita em termos da incidência de palavras em um e da sua ausência ou escassez em outro, dito de modo muito simplificado, extraem-se as *keywords* ou palavras-chave – que podem ser expressões multipalavra ou polilexicais.

Essas *keywords* são, assim, os primeiros **candidatos a termo**, que deverão ser cuidadosamente examinados nos seus usos ao longo do *corpus* de estudo. Isso porque seriam, em tese, as expressões típicas de um dado domínio frente a um uso não especializado da língua.

Depois de uma série de testes, incluem-se, nessa lista, outros candidatos a termo, e a lista de expressões passará por várias etapas de validação. Entre elas, não se dispensará a leitura de textos, ainda que uma leitura por amostragem, feita por analistas-terminólogos, e a consulta a especialistas da área de conhecimento, chegando-se, finalmente, aos termos mais relevantes do domínio em foco. Nesse processo, uma vez feito o levantamento das terminologias, parte-se, então, para a busca de suas definições, também trazidas de *corpora* especialmente reunidos para esse fim. Nem sempre termos e suas definições ocorrem em um mesmo tipo de texto.

### **Processamento de linguagem natural (PLN) e terminologias**

O PLN integra a Ciência da Computação, situando-se entre os estudos de Inteligência Artificial. Essa área de conhecimento explora as relações entre Linguística e Informática, tornando possível a construção de sistemas com capacidade de reconhecer e de produzir informação apresentada em linguagem natural (VIEIRA; STRUBE DE LIMA, 2001). Seu objetivo é, assim, essencialmente aplicado, relacionado à produção de uma ferramenta concreta, para atender uma necessidade específica. Essa ferramenta ou reconhe-

ce linguagem, geralmente escrita, como um extrator de terminologias, ou produz linguagem, como uma tradução automática, por exemplo.

Um exemplo interessante desse tipo de enfoque é o trabalho de Nazar e Cabré (2011). Nele, relata-se uma metodologia típica de PLN para a extração automática de candidatos a termos baseada em técnicas de análise estatística de textos. Diferentemente da maioria dos extratores de terminologia que aparecem na literatura sobre o assunto, essa proposta não inclui qualquer conhecimento explícito sobre a língua ou o domínio que está sendo analisado. Trata-se de um algoritmo que extrai as informações diretamente dos dados reunidos em um acervo textual, por meio de uma fase de treinamento na qual um usuário “ensina” ao computador os exemplos de unidades terminológicas (uma lista de termos validados) e unidades não terminológicas (uma coleção de textos não especializados). A partir destes exemplos, listas de **termos** e de **não-termos**, o algoritmo realiza uma percepção de padrões recorrentes que permite distinguir novas unidades terminológicas em novos textos, levando em conta esses padrões morfológicos, lexicais e sintáticos previamente “aprendidos”. A avaliação de desempenho desse algoritmo em termos de precisão e cobertura, conforme seus autores, demonstrou qualidade suficiente para ser útil no reconhecimento automático de terminologias, independentemente de domínio ou da língua com que se trabalhe. Aqui cabe ressaltar que há uma pessoa, um terminólogo (o usuário), que faz uma separação prévia entre termos e não termos. Afinal, alguém “ensina a máquina”, que funcionará partindo de um *goldstandard* de padrões para as duas categorias. O trabalho é automático, mas parte de um trabalho manual de base.

Um outro exemplo, desta vez brasileiro, dessas técnicas é o sistema ExATOlp – Extrator Automático de Termos para Ontologias em Língua Portuguesa. Conforme explicam Lopes e Vieira (2010), trata-se de uma ferramenta que recebe um *corpus* anotado e extrai automaticamente todos os sintagmas nominais (SN) dos textos, desde que esses SNs estejam na função sujeito ou objeto em suas sentenças de origem. A função primária dessa ferramenta é extrair

termos candidatos a conceitos, auxiliando a construção de ontologias, glossários e outros recursos semânticos. A ferramenta pode ser acessada em <[https://www.inf.pucrs.br/peg/lucelenelopes/ll\\_tls.html](https://www.inf.pucrs.br/peg/lucelenelopes/ll_tls.html)> conforme dados de maio de 2018.

Com o ExATOlp, os sintagmas extraídos são salvos em listas que podem conter tanto os SN na sua forma original no texto, como em sua forma canônica. A ferramenta ainda oferece algumas opções de manipulação usuais para listas de termos como a aplicação de pontos de corte, comparação de listas de palavras empregadas em textos e cálculo de medidas de associação entre palavras. As funcionalidades dessa ferramenta vão desde tarefas fortemente baseadas em conceitos linguísticos, como a extração de sintagmas nominais, até tarefas puramente estatísticas, passando por tarefas como a localização de termos extraídos identificando os textos e frases em que ocorram.

Uma amostra do funcionamento dessa ferramenta, que gera visualização desses SN dispostos em árvores hiperbólicas móveis, partindo de um *corpus* de artigos científicos que tratam do tema das Pneumopatias Ocupacionais, pode ser vista em <http://www.ufrgs.br/textecc/pneumopatias/>. Deve-se clicar nas guias **Hierarquia e Nuvem de Conceitos**.

A seguir, na Figura 1, temos uma imagem gerada por essa ferramenta no formato **Nuvem de Conceitos** a partir de um *corpus* formado pelo texto do *Curso de Linguística Geral*, de Saussure. Palavras maiores mostram-se mais relevantes, em termos de frequência e especificidade nesse domínio, dos que as menores.

Na Figura 1, os itens extraídos são palavras únicas – unigramas. O sistema ExATOlp gera tanto hierarquias quanto nuvens com itens compostos de duplas de palavras (bigramas, como **signo linguístico** ou trigramas como **princípio da arbitrariedade**) quanto com mais elementos, chamados n-gramas. No exemplo, temos que o item **língua** é o mais relevante e um outro item como **sociedade** pouco relevante. Nessa lógica, o primeiro está para termo e o segundo para não termo.

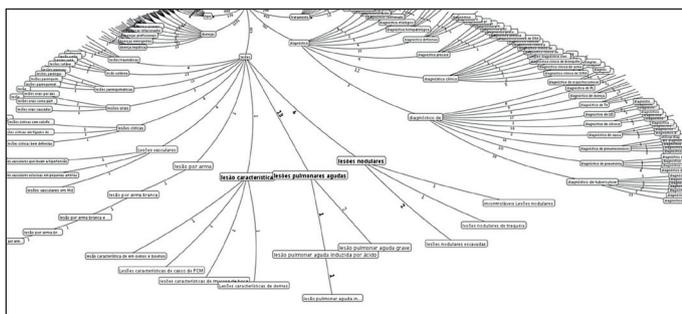
*Informatização e trabalho humano em Terminologia:  
extrair terminologias em corpus é o mesmo que identificá-las?*

**Figura 1** – Imagem Nuvem de Conceitos gerada pelo extrator ExATOLp – unigramas



**Fonte:** Extrator ExATOLp.

**Figura 2** – Imagem Hierarquia de Conceitos pelo ExATOLp no corpus CLG de Saussure



**Fonte:** Extrator ExATOLp.

Um outro recurso digno de nota, nessa linha de produtos de PLN para a representação de conteúdo e para o tratamento de material textual terminológico, é um sistema que traz relações não taxonômicas – isto é, as relações que extrapolam as tradicionais relações É PARTE DE/ É TIPO DE, presentes em um *corpus* ou em vários *corpora*, com foco para verbos, vistos como **relações**, e para os SN que se ligam a eles – vistos como elementos com valor de sujeitos e objetos gramaticais de uma dada oração ou período. Esse recurso, apresentado por pesquisadores do grupo de PLN da

PUCRS (FERREIRA; LOPES; VIEIRA, 2012), disponível gratuitamente em <<http://vhflabs.com.br/nontax/index.php>>, pode mostrar automaticamente, por exemplo, que, em um dado *corpus* de artigos científicos de Pediatria, há uma forte relação entre o verbo CAUSAR e os elementos tais como ELEVAÇÃO DA PRESSÃO ARTERIAL e LESÃO. Nesse caso, temos um passo além da identificação automática de prováveis termos num *corpus*, pois se pode aproveitar o recurso para construir um quadro de termos/conceitos e suas inter-relações.

### Diferenças entre ferramentas de LC e de PLN

A principal diferença entre um recurso de PLN e um de LC é o seu perfil de saída, o seu *output*. No primeiro caso, a saída é uma ferramenta, um produto, tal como uma representação de conteúdo ou mesmo um glossário semipronto. No caso da LC, a ferramenta, em geral fornece dados para a análise do terminólogo, que a partir deles poderá gerar o produto final.

Muitos recursos de PLN têm buscado dados em diferentes fontes, inclusive na *Wikipédia*, para a sustentação de sistemas de perguntas-e-respostas que varrem toda uma série de informações enciclopédicas na internet sobre um dado item ou palavra de busca. Após a varredura, reúnem as informações numa base de dados. Essa via de pesquisas, chamada de **mineração de textos** em Ciência da Computação, poderia inclusive selecionar, por exemplo, apenas um conjunto de dados históricos sobre uma dada terminologia/conceito ou noção. Aqui antevemos mais uma perspectiva para, num futuro próximo, ajudar a diminuir nossos esforços para preencher a lacuna da informação diacrônica sobre terminologias.

As ferramentas de PLN, conforme percebo, tendem, em mais alguns anos, a automatizar completamente a obtenção de “candidatos a termo” e a gerar recursos, com interface amigável para o linguista, que permitam que o terminólogo selecione os verdadeiros termos com maior facilidade, a partir de uma base de dados. Essa base de dados substituiria uma leitura extensiva de textos de apoio e de textos que integram o próprio *corpus* que se explore. A con-

sulta ao especialista do domínio cujos termos se busca identificar também tenderá, conforme prevejo, a ser mais facilitada e menos frequente, de modo que o *output* do processo informatizado venha a ser um glossário pronto ou semipronto, com verbetes, notas e definições.

A Terminologia e a Terminografia do futuro, conforme imagino, serão fortes aliadas das pesquisas em PLN e vice-versa. Certamente seguiremos, como terminólogos, acompanhados pela LC, ainda que suas ferramentas tenham um perfil menos automatizante, caracterizando-se por “gerar dados” para filtrar. Nesse cenário porvir, a Terminologia, a contraparte mais “reflexiva” da prática terminográfica, seguirá alimentada pelas evidências das práticas informatizadas. Afinal, o trabalho informatizado já hoje fornece muitos insumos para o trabalho de descrição dos fenômenos linguísticos associados à comunicação técnico-científica. Nesse cenário, seja no futuro ou no presente, será importante o recíproco reconhecimento de virtudes e de limites e de pontos de enriquecimento comuns, seja por parte dos linguistas (incluindo os linguistas de *corpus*), seja por parte dos informatas.

## **Agradecimentos**

Aos programas CAPES STIC-AMSud – *Projeto RITA*, CAPES-COFECUB – *Projeto CAMELEON*, CAPES-FAPERGS DOCFIX e ao CNPq pelos apoios que permitem a investigação e o intercâmbio entre Estudos da Linguagem e PLN.

## **REFERÊNCIAS**

ALMEIDA, G. M. B.; ALUÍSIO, S. M.; TELINE, M. F. Extração manual e automática de terminologia: comparando abordagens e critérios. In: WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA, 1., 2003, São Carlos. *Anais...* São Carlos: USP, 2003. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/til/2003/0012.pdf>>. Acesso em: 09 fev. 2018.

ESTOPÀ BAGOT, R. Eficiencia en la extracción automática de terminología. **Perspective: Studies in Traductology**, Oxon, v.7, n.2, p.277-286, 1999.

FINATTO, M. J. B. Orientações para a terminografia: das teorias às práticas em busca de amplitude da informação terminológica. In: ISQUERDO, A. N.; CORNO, G. O. M. Dal. (Org.). **As ciências do Léxico: Lexicologia, Lexicografia e Terminologia**. Campo Grande: Ed. da UFMS, 2014. v.7, p.433-452.

FERREIRA, V. H.; LOPES, L.; VIEIRA, R. Descoberta automática de relações não-taxonômicas a partir de corpus em língua portuguesa. In: SEMINAR ON ONTOLOGY RESEARCH IN BRAZIL, 5.; INTERNATIONAL WORKSHOP ON METAMODELS, ONTOLOGIES AND SEMANTICS TECHNOLOGIES, 7., 2012, Recife. **Proceedings of Joint**. Recife: ONTOBRAS/MOST, 2012. p.277-286.

GARCIA, A. F. Prefácio. In: SHEPHERD, T.; SARDINHA, T. B.; PINTO, M. V. (Org.). **Caminhos da Lingüística de Corpus**. São Paulo: Mercado de Letras, 2012. p.11-14.

L'HOMME, M.-C. Conception d'un dictionnaire fondamental de l'informatique et de l'Internet: Sélection des entrées. **Le Langage et l'Homme**, Paris, v.40, n.1, p.137-154, 2005.

LIMA, V. L. S. de; NUNES, M. das G. V.; VIEIRA, R. Desafios do Processamento de Línguas Naturais. In: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE; CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 26., 2007, Rio de Janeiro. **Anais...** Rio de Janeiro: SBC, 2007. V.1, p.2202-2216.

LOPES, L.; VIEIRA, R. Processamento de Linguagem Natural e o tratamento computacional de linguagens científicas. In: PERNA, C.; DELGADO, H. O.; FINATTO, M. J. B (Org.). **Linguagens Especializadas em Corpora: modos de dizer e interfaces de pesquisa**. Porto Alegre: EDIPUCRS, 2010. p.183-201.

NAZAR, R.; CABRÉ, T. Un experimento de extracción de terminología utilizando algoritmos estadísticos supervisados. **Debate Terminológico**, Porto Alegre, n.7, p.36-55, abr. 2011.

REY-DEBOVE, J. Léxico e Dicionário. **Alfa**, São Paulo, v.28, supl., p.45-69, 1984.

SARDINHA, T. B. **Lingüística de Corpus**. Barueri: Manole, 2004.

TEIXEIRA, R. de B. S. e. **Termos de (Onco)Mastologia**: uma abordagem mediada por corpus. 2010. 169f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica, São Paulo, 2010.

VIEIRA, R.; STRUBE DE LIMA, V. L. Linguística computacional: princípios e aplicações. In: NEDEL, L. P. (Org.). **IX Escola de Informática da SBC-Sul**. Porto Alegre: UFRGS, 2001. v.1, p. 27-61.

VIVALDI, J.; RODRÍGUEZ, H. Improving term extraction by combining different techniques. **Terminology**, Amsterdam, v.7, n.1, p.31-48, 2001.



# ANOTAÇÃO DE *CORPUS* DO GREGO ANTIGO: PROJETO DE TRADUÇÃO ALINHADA, *TREEBANK* E LÉXICO GREGO- PORTUGUÊS EM SERVIÇOS *WEB*

Anise D'Orange FERREIRA

O projeto na área de grego, apresentado neste texto, envolvendo tecnologias digitais tem, por finalidade, fomentar o ensino, a pesquisa, e a formação na língua grega, e publicações abertas em língua portuguesa em colaboração com projetos internacionais. A área de letras clássicas envolve uma comunidade restrita no Brasil, apesar do crescimento apontado por Correa (2014) na formação de corpo docente. Essa limitação se deve, principalmente, à praticamente inexistência de estudos clássicos nos níveis do ensino fundamental e médio e a sua presença restrita em poucas universidades. A Sociedade Brasileira de Estudos Clássicos, que existe desde os anos 90, não chega a mil associados, que são estudiosos dedicados a diferentes disciplinas ligadas à filologia, em particular: Grego, Latim, Sânscrito e outras línguas e literaturas históricas; Filosofia e História Antigas e Arqueologia. Compartilham-se *corpora* históricos, mas investigações são realizadas cada uma dentro da própria disciplina. Nas Letras, formam-se poucos que trabalham com textos antigos, traduzindo-os para a língua portuguesa, face a poucas atuações no mercado de trabalho<sup>1</sup>, embora o *corpus*

---

<sup>1</sup> Uma das perguntas mais frequentes ouvidas por quem estuda grego é “para quê?”.

das línguas clássicas a ser traduzido seja imenso. Não temos uma extensa coleção ou série editorial contínua dedicada aos textos clássicos como missão de uma editora específica, como há na Europa e EUA. Dispomos de algumas editoras universitárias e não universitárias que eventualmente publicam traduções de textos gregos. Na França, há a coleção *Belles Lettres* com edições bilíngues, criada em 1920<sup>2</sup>; nos EUA, a coleção *Loeb*<sup>3</sup>, da Harvard U. Press; na Inglaterra, Oxford e Cambridge têm suas próprias coleções de edições com aparato crítico<sup>4</sup>. Na Alemanha, a *Bibliotheca Teubneriana*, estabelecida em 1849<sup>5</sup>. Portugal, com uma tradição europeia dos clássicos, conta com algumas editoras que publicam traduções dos clássicos. No último quinquênio, em parceria com uma editora, foi criado na *web* um repositório português, mantido pela Universidade de Coimbra, de *e-books* gratuitos de obras e estudos clássicos, chamado *Classica Digitalia*<sup>6</sup>.

Apesar de tratarmos com textos pouco acessíveis aos brasileiros, estamos atuando em um campo com um tipo de saber que contempla o legado da civilização ocidental, i.e., parte do que somos hoje, no discurso, no pensamento, nas ciências, nas artes, nas ações e gestos. Em suma, uma área minimizada na educação<sup>7</sup>, mas que tem impacto implícito e dissipado em uma gama de campos do conhecimento. Assim, nós, docentes universitários de grego antigo, um grupo reduzido em número, comparativamente a outras áreas, temos a responsabilidade de manter a disciplina e a pesquisa, e de fazer alcançar à sociedade os conhecimentos gerados. Temos de dar conta de necessidades acadêmicas, didático-pedagógicas e de difusão, para que a área não desapareça no pequeno espaço que ocupa dentro do contexto universitário nacional (FERREIRA, 2014).

---

<sup>2</sup> Cf. <<http://www.lesbelleslettres.com/catalogue/textesgrecs/>> e <<http://www.lesbelleslettres.com/info/?fa=historiquefull>>.

<sup>3</sup> Cf. <<http://www.hup.harvard.edu/features/loeb/history.html>>.

<sup>4</sup> *The Oxford Classical Texts*, ou *Scriptorum Classicorum Bibliotheca Oxoniensis*; *Cambridge Library Collection: Classics*.

<sup>5</sup> Incorporada pela de Gruyter em 2007. Cf. <<http://www.degruyter.com/view/serial/36366>>.

<sup>6</sup> Cf. [https://www.uc.pt/imprensa\\_uc/catalogo/classicadigitalia](https://www.uc.pt/imprensa_uc/catalogo/classicadigitalia)

<sup>7</sup> Excluída dos níveis fundamental e médio; mantida em algumas instituições de ensino superior e dentro de poucos programas de pós-graduação exclusivos da área.

O meio para tentar cumprir tal tarefa vamos encontrar, então, nos recursos digitais do terceiro milênio, junto aos classicistas digitais, ou da era digital, que compartilham as mesmas preocupações<sup>8</sup> e trabalham na interface da linguística computacional e da linguística de *corpus* com *corpora* históricos, voltando-se tanto para o ensino, como para pesquisa e tendo como instrumento básico a produção de uma biblioteca digital de acesso aberto.

Nosso projeto voltado para a anotação de *corpus*, embora em estágio inicial<sup>9</sup>, é, de fato, um conjunto de projetos individuais e coletivos que seguem a política e os pressupostos do *Open Philology Project* (CRANE, 2013-2017) e *Perseus Digital Library* (CRANE, [20-]) sob comando de Gregory Crane (BAUMGART et al., 2014). A Biblioteca Digital Perseu (doravante PDL) é um portal que existe desde os anos 90, abrigado na Universidade de Tufts em Boston, nos EUA, com muitos recursos para leitura e pesquisa, e tem passado por grandes modificações, como começar a ser distribuído sob licença *Creative Commons* e adquirir padrões sistematizados e universais de indexação. Seu acesso já era irrestrito, mas agora seu destino é oficial. Tanto que o *corpus* de sua base de dados, bem como suas ferramentas têm sido reutilizadas para uso em várias outras ferramentas, como o *Diogenes*<sup>10</sup>, *Kalós*<sup>11</sup>, *Perseus under Philologic*<sup>12</sup>, *Alpheios*<sup>13</sup>, *Attikos*<sup>14</sup>, *Logeion*<sup>15</sup>, sendo os dois últimos disponíveis como aplicativos para *iPad*, e assim por diante.

Atualmente, essa biblioteca digital (PDL) passa por reformulação de padrões e ampliação do acervo, ações guiadas por sua associação ao projeto parceiro e irmão caçula, *Open Philology*, sediado em Leipzig, dentro do professorado Humboldt, com duração

---

<sup>8</sup> Cf. <<http://www.digitalclassicist.org/>>.

<sup>9</sup> O projeto inicial desta proposta sob minha coordenação foi financiada pelo Edital 18/2012-14 CNPq.

<sup>10</sup> Cf. <<https://community.dur.ac.uk/p.j.heslin/Software/Diogenes/>>.

<sup>11</sup> Cf. <<http://www.kalos-software.com/>>.

<sup>12</sup> Cf. <<http://perseus.uchicago.edu/>>.

<sup>13</sup> Cf. <<http://alpheios.net>>.

<sup>14</sup> Cf. <<https://itunes.apple.com/br/app/attikos/id522497233?mt=8>>.

<sup>15</sup> Cf. <<http://logeion.uchicago.edu/>>.

de cinco anos, desenvolvido por seu criador, Crane, dentro das Humanidades Digitais da Universidade de Leipzig. Ambos os projetos se desdobram em: Grego e Latim Aberto, voltado para a organização dos conteúdos, incluindo as traduções para as línguas modernas; em *E-Learning* de línguas históricas, e na biblioteca digital *Scaife*, que se encarrega de explorar as fontes disponíveis sob licenças abertas. Esses desdobramentos requerem uma ciberinfraestrutura que torne possível sua realização. O apoio das linguagens de computação, de instrumentos da linguística de *corpus* e linguística computacional é essencial. O *corpus*-alvo do projeto, em Leipzig, contém mais de 400 milhões de palavras, de obras do período de 600 a.C. até os anos 80, em virtude da lei de direito autoral alemã<sup>16</sup>, que deve ser digitalizado com OCR<sup>17</sup> adequado à natureza dos documentos, conter metadados em FRBR<sup>18</sup>, com lematização, anotação seguindo padrão TEI-XML<sup>19</sup>, marcando citações, reusos, morfologia, entidades de nomes, e *treebank* ou árvores sintáticas.

É importante destacar que tais projetos são comprometidos com o tratamento de dados e a publicação desses em licenças totalmente abertas. Isso significa ser o acesso irrestrito. Nesse sentido é que a biblioteca de Leipzig oferece uma larga possibilidade para se trabalhar com os textos.

Esses projetos apresentam três estratégias básicas para engajamento coletivo. Um deles é pressupor o ciclo virtuoso de aprendizagem e de contribuição (CRANE et al., 2012). O outro é desenvolver ferramentas de suporte, ou uma infraestrutura com-

---

<sup>16</sup> Tecnicamente, a lei alemã de direitos autorais pode considerar que uma obra acadêmica esteja em domínio público depois de 25 anos da primeira publicação. Cf. Copyright Law. Disponível em: <<http://www.iuscomp.org/gla/statutes/UrhG.htm>>.

<sup>17</sup> Sigla de *Optical Character Recognition* – programa que transforma imagens de texto em caracteres de texto, ou texto editável.

<sup>18</sup> Sigla de *Functional Requirements for Bibliographic Records*, um modelo de padronização de entidades que é independente de códigos de catalogação.

<sup>19</sup> TEI, sigla de *Text Encoding Initiative*, consórcio internacional e coletivo que desenvolve o padrão para o formato digital de textos, fornecendo orientações para métodos de codificação de textos que são lidos por máquina; XML, sigla de *Extended Markup Language* linguagem para marcação de identificação de dados.

putacional de base, ou ciberinfraestrutura, para chegar-se ao ciclo virtuoso. Por fim, estabelecer acordos de cooperação internacionais, para aumentar a quantidade de participantes no projeto. O círculo virtuoso de aprendizagem e contribuição envolve o desenvolvimento de ferramentas para várias fases de trabalho sobre o *corpus*, desde a transcrição do documento fonte, como um manuscrito, passando pelos vários elementos de etiquetagem para realização da leitura, e tradução até a produção da edição final digital. O pressuposto é que o estudante pode aprender e estudar o sistema da língua passando por essas etapas, usando suportes digitais, ajudado pela revisão dos pares e *feedback* dos professores. Ao mesmo tempo em que aprende, contribui com edições anotadas e traduções, quando o produto estiver no nível de acurácia desejado. Essa proposta, segundo Crane, tem origem na visão de (Wilhelm) Humboldt sobre universidade, em que o trabalho universitário não é simplesmente receber informação provida por outros, mas envolver-se em produção de novos conhecimentos, desde o primeiro ano de graduação. Podemos, também, do ponto de vista da aprendizagem, e de abordagens instrucionais, aproximar tal posição de propostas como *data-driven learning*<sup>20</sup>, ou aprendizagem baseada em projeto colaborativo, ou no fazer, ou ainda baseada em solução de problemas<sup>21</sup>, com raízes em teorias como o construtivismo social, teoria histórico-cultural, e mesmo centrada no aluno, à medida em que esse possa escolher o *corpus* para estudar.

É dentro desse espírito que o presente projeto foi criado e se encontra em andamento. Nosso papel nesse contexto e projeto global é de usar os serviços criados em âmbito internacional e testar as ferramentas de apoio em projetos próprios que envolvem *corpora* de textos sem tradução alinhada em português, e textos gregos ainda não anotados. No plano de ensino-aprendizagem usamos, no

---

<sup>20</sup> Aprendizagem movida a dados. O uso de exploração de *corpora* em ensino-aprendizagem de língua. Cf. McEnery e Hardie (2012, p.243) e Johns (1991).

<sup>21</sup> Cf. CSCL – *computing supported collaborative learning*, Stahl, Koschmann e Suthers (2006); construtivismo e outras teorias em M. Reigeluth (1999); Jonassen, Peck e Wilson (1999).

momento, a ferramenta *Alpheios*<sup>22</sup> para leitura e tradução, edição de alinhamento<sup>23</sup>, edição de anotação manual da árvore sintática de dependência (*treebank*). Essa ferramenta, implementada na plataforma de edição *Perseids*<sup>24</sup> foi aprimorada com os recursos de outra ferramenta, chamada *Arethusa*.

No plano de produção de material de apoio, pretendemos criar uma edição em padrão TEI-XML do dicionário e glossário grego-português para incluir nos vários serviços e ferramentas de apoio já desenvolvidas e em desenvolvimento: *Alpheios*, *Perseus Digital Library*, *Logeion* em Chicago etc.

Fazem parte desse projeto alguns colaboradores, entre eles, os professores da área de grego: Edvanda B. Rosa, Fernando B. dos Santos e M. Celeste Consolin Dezotti; a professora da área de latim, Giovanna Longo. Participam alunos de graduação e pós-graduação orientados por esses professores. Alguns participam com trabalhos coletivos, outros, com pesquisa individual; alguns realizam o projeto em sala de aula, outros, como projeto extraclasse.

Desde o início de 2013, vimos trabalhando com as ferramentas *Alpheios* em sala de aula. Essas ferramentas funcionam no navegador *Firefox*, com a instalação e ativação de um *plug-in*. O *site* inclui uma interface de edição de alinhamento e de *treebank* para facilitar o uso dos falantes da língua portuguesa<sup>25</sup>. Um dos usos do *Alpheios* é a consulta ao *corpus* grego<sup>26</sup> com tradução alinhada já existente no *site*, como na Figura 1. Realizamos a edição de alinha-

---

<sup>22</sup> Ferramenta elaborada em colaboração com a equipe de Crane, da Tufts, sob gerência de Harry Diakof, e execução da programadora Bridget Almas, também da equipe da Perseus Digital Library. Cf. <<http://alpheios.net>>. O projeto *Alpheios* começa a passar por atualização em 2017, mas a versão original permanece acessível.

<sup>23</sup> O projeto de edição de alinhamento recebeu apoio do Edital 18 do CNPq de dezembro de 2012.

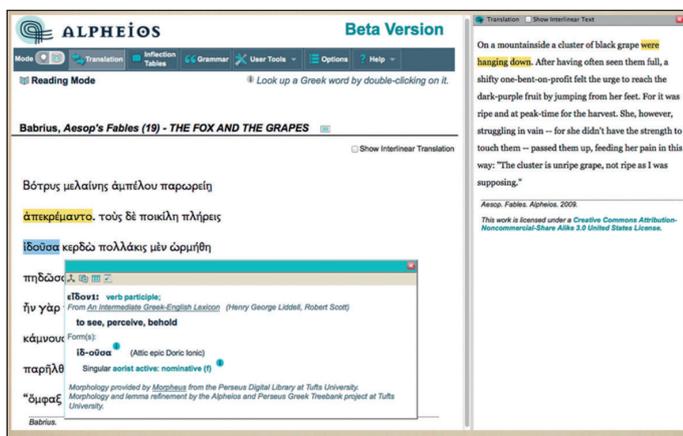
<sup>24</sup> Cf. Beaulieu e Almas, 2013; Berti e Almas, 2013 e *Perseids website* - <<http://sites.tufts.edu/perseids/>>. A partir de 2015 esta plataforma passou a ser usada em nova fase do projeto, cf. Ferreira, 2017 <<http://dx.doi.org/10.23925/2318-7115.2017v38i1a9>>.

<sup>25</sup> Portuguese Entry Points: Alignment Editor <[http://repos1.alpheios.net/exist/rest/db/app/align-entersentence\\_por.xhtml](http://repos1.alpheios.net/exist/rest/db/app/align-entersentence_por.xhtml)>. Treebank Editor: <[http://repos1.alpheios.net/exist/rest/db/app/treebank-entertext\\_por.xhtml](http://repos1.alpheios.net/exist/rest/db/app/treebank-entertext_por.xhtml)>.

<sup>26</sup> O *corpus* existente no *Alpheios* provém da PDL. São as edições gregas da Loeb, U. Harvard Press, com acesso aberto.

mento do grego com a tradução em português, de alguns desses textos: excertos da *Biblioteca* de Apolodoro e de algumas fábulas de Esopo (Bábrio); de excertos da *História* de Heródoto; da edição de alinhamento da tradução de excertos de tragédia: *Persas* e *Prometeu Acorrentado*, de Ésquilo; edição de alinhamento de tradução de outros textos de prosa: excertos de Heródoto. Esse material está sendo armazenado nos formatos XML e XHTML. No primeiro, para compor nosso próprio banco de dados e o repositório do *Alpheios*. No segundo formato, para ser visualizado no navegador.

**Figura 1** – Consulta a texto grego com tradução alinhada no *Alpheios*



**Fonte:** *Alpheios*

Além da consulta ao texto alinhado, a ferramenta dispõe de consulta a tabelas de flexão, dicionários, árvore sintática, palavras aprendidas, etc. (Figura 2). Na Figura 3, vemos uma consulta à árvore sintática de dependência (*treebank*). A anotação já realizada por outras pessoas e exposta no *site* do *Alpheios* segue um guia de anotação sintática para anotação do *treebank* do grego antigo (BAMMAN; CRANE, 2008<sup>27</sup>).

<sup>27</sup> Cf. <<http://nlp.perseus.tufts.edu/syntax/treebank/greekguidelines.pdf>>. As versões em português estão disponíveis na FCLAr para os interessados pelo contato com este autor <[anise@pq.cnpq.br](mailto:anise@pq.cnpq.br)>.

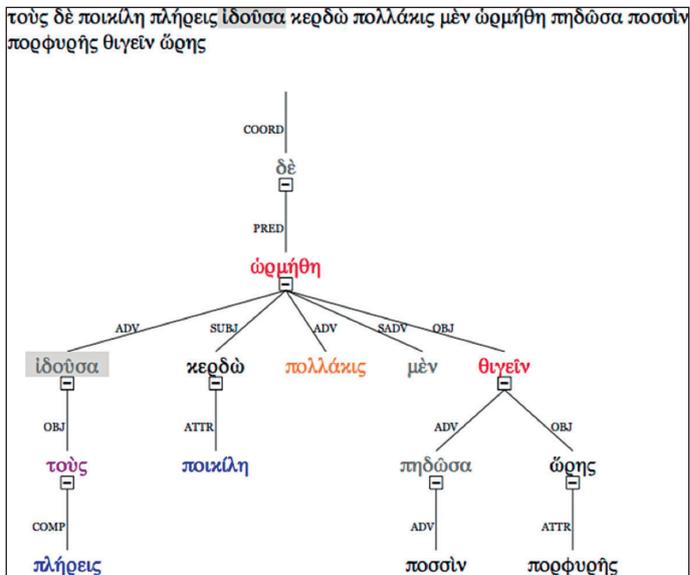
Figura 2 – Consulta à tabela de flexão verbal e verbete de dicionário

The screenshot shows a web browser window titled 'Inflections'. The main content is a table of Greek verb forms for the verb 'εἶδον' (to see). The table is organized by gender (masculine, feminine, neuter) and number (singular, dual, plural). The forms listed include nominative (n.v.), genitive (g.), dative (d.), and accusative (acc.) for each gender and number. A pink arrow points to the title of the table: 'Participles in -ών, -ούσα, -όν (strong aorist active; present of εἶμι and εἶμι)'. Below the table, there is a dictionary entry for 'εἶδον1: verb participle; From An Intermediate Greek-English Lexicon (Henry George Liddell, Robert Scott) to see, perceive, behold'. The entry also includes the form 'Form(s): ἰδ-ούσα (Attic epic Doric Ionic)' and 'Singular aorist active: nominative (f)'. At the bottom, there is a note: 'Morphology provided by Morpheus from the Perseus Digital Library at Tufts University. Morphology and lemma refinement by the Alpheios and Perseus Greek Treebank project at Tufts University.'

|          |            | masculine | feminine | neuter  |
|----------|------------|-----------|----------|---------|
| singular | n.v.       | ἰδών      | ἰδούσα   | ἰδόν    |
|          | genitive   | ἰδόντος   | ἰδούσας  | ἰδόν    |
|          | dative     | ἰδόντι    | ἰδούσαι  | ἰδίῃ    |
| dual     | n.v.       | ἰδόντε    | ἰδούσα   | ἰδόν    |
|          | g.d.       | ἰδόντων   | ἰδούσων  | ἰδόν    |
| plural   | n.v.       | ἰδόντες   | ἰδούσαι  | ἰδόντες |
|          | genitive   | ἰδόντων   | ἰδούσων  | ἰδόντων |
|          | dative     | ἰδούσιν   | ἰδούσαις | ἰδίῃσιν |
|          | accusative | ἰδόντας   | ἰδούσας  | ἰδόντας |

Fonte: Alpheios

Figura 3 – Consulta à árvore sintática do texto grego



Fonte: Alpheios.

A consulta às ferramentas de morfologia e dicionário não se faz exclusivamente com textos presentes no *Alpheios*. Qualquer texto em grego politônico Unicode na *web*, visto pelo navegador com o *plug-in* do *Alpheios* ativado, permite o uso associado desses recursos (Figura 4), exceto da árvore sintática, cuja visualização depende de o *corpus* ter sido anotado previamente.

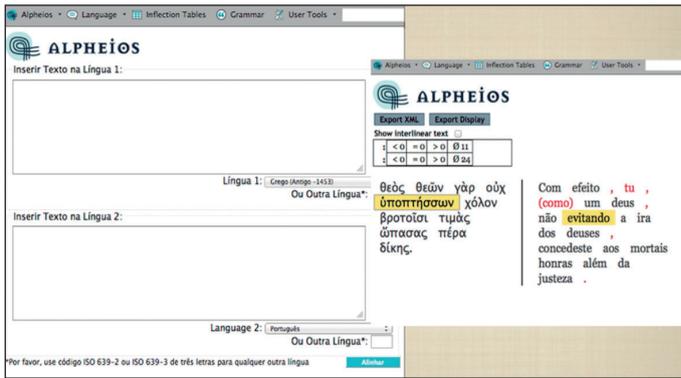
**Figura 4** – Consulta às ferramentas com outros textos gregos disponíveis na *web*



Fonte: *Alpheios*

Na Figura 5, a primeira janela exibe o editor, a página de entrada dos textos da língua 1 e 2. Essa ferramenta efetua o alinhamento de quaisquer idiomas. O arquivo alinhado é exportado em dois formatos: XML para ser incluído nas bases de dados, e XHTML para poder ser visualizado pelo navegador após a realização do alinhamento.

**Figura 5** – Interface do editor de alinhamento de tradução do Alpheios com uma janela exibindo o alinhamento do texto



Fonte: *Alpheios*

A Figura 6 mostra a interface de entrada do texto do editor de *treebank*. Ele pode ser feito “colando” o texto grego no quadro ou carregando um arquivo XML com uma árvore já feita ou iniciada. A etiquetagem do editor de *treebank* é dupla (Figura 7). Uma é feita escolhendo-se as etiquetas da categoria gramatical e flexões: pessoa, número, gênero, etc. A outra é feita escolhendo-se as etiquetas da árvore sintática: SBJ, sujeito, OBJ, objeto, e assim por diante, lembrando que o entendimento dessa etiquetagem sintática depende dos critérios e definições propostos no guia de anotação. Além da etiquetagem, o usuário movimenta os “galhos” da árvore, colocando cada item lexical no seu ponto de dependência. Não cabe aqui explicar o procedimento, uma vez que esse está descrito no guia.

**Figura 6** – Interface de entrada do texto do editor de treebank



Fonte: Alpheios

**Figura 7** – Modos de etiquetagem da árvore sintática



Fonte: Alpheios

Dos trabalhos coletivos dentro do projeto, destacam-se nossa tradução do guia de anotação sintática do Grego, com a colaboração dos professores Filomena Hirata (USP) e Fernando B. Santos (UNESP); a tradução do guia de anotação do latim, por alunos da FCLAr-UNESP, de graduação Kelli Fauth Claro, e da pós-graduação, Adriana Sandrin e Ana Maria Riccioli, com a revisão da professora Giovanna Longo (UNESP). Ainda, os vídeos instrucionais do uso das ferramentas por Caio Vieira e Kelli Claro. Em andamento, em sala de aula, temos a edição de alinhamento de excertos da peça *Medeia*, de Eurípides sob supervisão da professora Edvanda B. Rosa; a edição de árvores sintáticas de sentenças avulsas e de passagens de Heródoto sob minha supervisão. Em grupo extraclasse: tivemos o alinhamento de excertos da tragédia *Héracles*, sob supervisão de Edvanda B. Rosa, executado por aluno de graduação (Michel Reis). Em trabalhos individuais, tivemos a tradução da *Biblioteca*, de Apolodoro, pelo doutorando Caio Vieira, os TCCs com anotações e tradução de excertos de *Heródoto* (Katia Stamberk) e *Sentenças* de Menandro (Paula Boschiero), uma pesquisa de iniciação científica (Mariana Masotti), com alinhamento da peça *O Ciclope*, sob orientação de M. Celeste C. Dezotti. Em preparação, como trabalho individual, o alinhamento de um dos livros de Artemidoro de Daldis, *Oneirokritika* (Anise D'O. Ferreira). Vislumbramos a oportunidade de colaborar, futuramente, com outros projetos internacionais de inserção de traduções de grego em língua portuguesa, como os de epigrafia digital<sup>28</sup> que vêm utilizando recursos desenvolvidos em parceria com o projeto *Open Philology* de Crane<sup>29</sup>.

Como vimos nas ferramentas, os dicionários exibidos são de língua inglesa. Assim, é nossa meta, e está em preparação, como trabalho coletivo, a elaboração de um léxico digital Grego-Português

---

<sup>28</sup> Por exemplo, o projeto de epigrafia digital de Michèle Brunet, que inclui a coleção de epígrafes gregas do Louvre. Cf. <[http://www.hisoma.mom.fr/mb/IG\\_LOUVRE/EPIGRAMME-EN.html](http://www.hisoma.mom.fr/mb/IG_LOUVRE/EPIGRAMME-EN.html)> e <[https://www.academia.edu/3340253/La\\_collection\\_des\\_inscriptions\\_grecques\\_du\\_Musee\\_du\\_Louvre](https://www.academia.edu/3340253/La_collection_des_inscriptions_grecques_du_Musee_du_Louvre)>.

<sup>29</sup> A partir de 2017, a plataforma Ugarit para alinhamento de tradução de até três línguas começa a ser utilizada <<http://ugarit.ialigner.com>>

que possa ser incluído nesses vários serviços. Em sua fase inicial, foi realizado um glossário dos termos mais frequentes com desenvolvedores de ferramentas (*DC Core Vocabulary*, do Dickinson College)<sup>30</sup>. Em fase seguinte, pretendemos rever e ampliar o Dicionário Grego-Português<sup>31</sup> de modo a ser inserido futuramente na PDL, ferramentas e plataformas associadas. Além disso, os alinhamentos de palavras feitos em um grande número futuramente poderão gerar um léxico dinâmico com a pesquisa feita tanto em grego como em português, tal como executado, para o grego e inglês, na proposta de Bamman e Crane (2008).

## **Algumas lições**

Do ponto de vista didático-pedagógico, o trabalho elaborado em sala de aula com edição de alinhamento de tradução oferece ao professor uma oportunidade ímpar para perceber e entender com clareza os tipos de dificuldades de seus alunos, muito antes desses serem avaliados do ponto de vista formal, ainda quando realizam as tarefas<sup>32</sup>. O alinhamento palavra a palavra é uma representação gráfica do entendimento ou desentendimento do aluno no nível morfológico, sintático e semântico. A edição de alinhamento de tradução também exige cuidados com pontuação e hifenização na língua portuguesa; precisamos deixar espaço entre os sinais de pontuação para que esses não fiquem alinhados de forma “colada” ao item lexical. Se o sinal de pontuação ficar “grudado” no alinhamento, com a palavra, a leitura pela máquina vai computar aquele item como um novo item lexical<sup>33</sup>. As traduções durante o alinhamento são constantemente reavaliadas, em virtude de um esforço que se faz para aproximar, do grego, os papéis sintáticos e a ordem

---

<sup>30</sup> Coordenado por Christopher Francese, com português implementado por Caio Camargo <<http://dcc.dickinson.edu/pt/greek-core-list-port>>

<sup>31</sup> Projeto DDGP- Dicionário Digital Grego-Português, inicialmente apoiado com um projeto CNPq/PQ/ 2017 e com auxílio regular FAPESP Out2017-2019 <<http://perseidas.fclar.unesp.br>>.

<sup>32</sup> Cf. Exposição mais detalhada de alinhamento em Ferreira, 2015.

<sup>33</sup> Isso não ocorreu mais no Alpheios da Plataforma Perseids.

em português. Elas fazem ampliar a discussão sobre as leituras, traduções e estilos; parecem fazer acentuar a conscientização e memorização lexical, e portanto, das traduções, pela quantidade de repetições de alinhamento que as tomadas de decisão obrigam a fazer. Além disso, provoca uma mudança na dinâmica de sala de aula, com maior comprometimento dos alunos, e todos se transformam, de certa forma, em pesquisadores.

Em relação à anotação sintática, o acompanhamento pelo guia é indispensável. As etiquetas resumem em um único grupo certas categorias que normalmente são detalhadas nas gramáticas tradicionais. Sua elaboração é descrita em Bamman e Crane (2008) O guia é baseado na gramática de dependência de Praga (BAMMAN; CRANE, 2009; BÖHMOVÁ et al., 2003) e na gramática latina de Pinkster (1990). Para um professor, essa tarefa requer uma espécie de acomodação entre o esquema proposto de anotação em árvore, e o conhecimento prévio das gramáticas gregas. A segunda versão<sup>34</sup> da anotação foi detalhada e se aproxima das categorias da gramática de Smyth (1956).

O trabalho de anotação exige, ainda, uma conscientização sobre eventuais diferenças de conceitos e nomenclatura das gramáticas da língua portuguesa e grega. Aqui também a dinâmica se altera. O professor pesquisa junto com aluno, discutindo opções de etiquetagem e debate com colegas essas opções. A anotação da árvore sintática obriga a tomar decisões ou aceitar a possibilidade de mais de uma leitura sintática do texto, em virtude das ambiguidades que a língua grega nos deixa. O efeito é maior quando os alunos têm de elaborar a árvore. Mas a navegação pelas árvores já construídas ajuda a leitura, no sentido de que orienta por meio de uma representação gráfica o jogo da sintaxe.

Nosso trabalho até o momento é ainda pequeno. Mas é importante ressaltar que os projetos dos classicistas da era digital terão vida longa e são promissores para novas atuações profissionais; enquanto houver um texto grego não digitalizado, não anotado e

---

<sup>34</sup> Elaborado por Giuseppe Celano, agosto de 2014: comunicação pessoal. Com tradução para português. Atualmente disponível em <[http://perseusdl.github.io/treebank\\_data/](http://perseusdl.github.io/treebank_data/)>

não traduzido, há trabalho a fazer. Além disso, quanto mais houver *corpora* anotados, mais perguntas de pesquisa poderão ser respondidas e novas ferramentas são desenvolvidas para buscar as respostas. Incentivamos as traduções inéditas em português, no nível mais avançado, mas também, como formação, precisamos lidar com textos que são tradicionais do currículo dentro da área. De todo modo, temos muitas oportunidades de atuação em projetos que podem ter um grande impacto, não só localmente, mas em colaboração com classicistas de várias partes do mundo. A era digital nos permite isso. Assim, procuramos contemplar as necessidades acadêmicas, didáticas e sociais apontadas no início.

## REFERÊNCIAS

BEAULIEU, M.-C.; ALMAS, B. Developing a new integrated editing platform for source

documents in Classics. **Literary and Linguistic Computing**, Oxford, v.28, n.4, p.493-503, 2013.

BAMMAN, D.; CRANE, G. Building a dynamic lexicon from a digital library. **JCDL'08**, Pittsburgh, June 2008. Disponível em: <<http://www.perseus.tufts.edu/~ababeu/fp135-bamman.pdf> >. Acesso em: 07 fev. 2018.

BAMMAN, D.; CRANE, G. **Guidelines for the syntactic annotation of Ancient Greek treebanks**. Version 1.1. Medford: Tufts Digital Library, 2009.

BAUMGARDT, B. et al. Open Philology at the University of Leipzig. **Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)**, Reykjavik, Iceland: European Language Resources Association (ELRA), 2014. Disponível em: <[http://www.lrec-conf.org/proceedings/lrec2014/pdf/940\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/940_Paper.pdf)>. Acesso em: 07 fev. 2018.

BERTI, M.; ALMAS, B. Perseids Collaborative Platform for annotating Text Re-Uses of fragmentary authors. **DH-CASE 2013**:

Collaborative Annotations in Shared Environments: metadata, vocabularies and techniques in the Digital Humanities, Florence: ACM Digital Library, 2013. Disponível em: <<http://www.fragmentarytexts.org/wp-content/uploads/2013/09/DH-Case-2013-Paper.pdf>>. Acesso em: 07 fev. 2018.

BÖHMOVÁ, A. et al. The prague dependency treebank: a three-level annotation scenario. In: ABEILLÉ, A. (Ed.). **Treebanks: building and using Parsed corpora**. New York: Springer, 2003. p. 123-126

CORREA, P. C. The Amorim Lima Project. **Classics in Brazil**, Centro de Estudos Helênicos de Harvard, Washington, maio 2014. Apresentação oral.

CRANE, G. **Open Philology Project**: Digital Humanities. Leipzig: Universät Leipzig, 2013-2017. Disponível em: <<http://www.dh.uni-leipzig.de/wo/open-philology-project/>>. Acesso em: 08 fev. 2018.

CRANE, G. (Ed.). **Perseus Digital Library**. Boston: Tufts University, [20-]. Disponível em: <<http://www.perseus.tufts.edu>>. Acesso em: 08 fev. 2018.

CRANE, G. et al. Student researchers, citizen scholars and the trillion word library. **Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL 2012)**, Washington: ACM Digital Library, 2012. p. 213-222. Disponível em: <<https://dl.acm.org/citation.cfm?id=2232857>>. Acesso em: 08 fev. 2018.

FERREIRA, A. D'O. Aprendendo grego antigo no mundo digital do terceiro milênio. **The ESPECIALIST**: descrição, ensino e aprendizagem, São Paulo, v.38, n.1, 2017. Disponível em: <<https://revistas.pucsp.br/index.php/esp/article/view/32220>> Acesso em: 08 fev. 2018.

FERREIRA, A. D'O. Traduzindo e produzindo dados abertos nas letras clássicas digitais. In Santos, F. B. e Oliveira, J.K. (Eds) **Estudos Clássicos e Seus Desdobramentos**: artigos em

homenagem à professora Maria Celeste Consolin Dezotti. São Paulo: Cultura Acadêmica, 2015. p.229-244.

FERREIRA, A. D'O. Applied linguistics, digital instructional design and teaching Greek with tools from PDL/Open Philology Projects. **Classics in Brazil**, Centro de Estudos Helênicos de Harvard, Washington, maio 2014. Apresentação oral.

JONASSEN, D. H.; PECK, K. L.; WILSON, B. G. **Learning with technology: a constructive perspective**. Upper Saddle River: Merrill, 1999.

JOHNS, T. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. **ELR Journal**, Birmingham, v.4, p.27-46, 1991.

MCENERY, T.; HARDIE, A. **Corpus linguistics: method, theory and practice**. Cambridge: CUP, 2012.

PINKSTER, H. **Latin syntax and semantics**. Londres; New York: Routledge, 1990.

REIGELUTH, C. M. (Ed.). **Instructional design theories and models: a new paradigm of instructional theory**. Mahwah: Lawrence Erlbaum Associates, 1999. v.2

SMYTH, H. W. **A Greek Grammar for colleges**. Cambridge: Harvard U. Press, 1956.

STAHL, G.; KOSCHMANN, T.; SUTHERS, D. Computer-supported collaborative learning: an historical perspective. In: SAWYER, R. K. (Ed.). **Cambridge handbook of the learning sciences**. Cambridge: Cambridge University Press, 2006. p.409-426. Disponível em: <<http://GerryStahl.net/elibrary/global>>. Acesso em: 08 fev. 2018.



## **SOBRE OS ORGANIZADORES E AUTORES**

### **ALDO LUIZ BIZZOCCHI**

Doutor em Semiótica e Linguística Geral pela USP (1994). Pesquisador do NEHiLP, USP, Universidade de São Paulo. São Paulo. SP. Brasil. CEP: 05508-900. E-mail: [contatos@aldobizzocchi.com.br](mailto:contatos@aldobizzocchi.com.br)

### **ÁLVARO IRIARTE SANROMÁN**

Doutor em Ciências da Linguagem – Linguística Aplicada, pela UMINHO (2000). Membro do GALABRA. Professor Associado do Departamento de Estudos Portugueses e Lusófonos, do Instituto de Letras e Ciências Humanas da UMINHO - Universidade do Minho. Campus de Gualtar. Braga. Portugal. CEP: 4710-057. E-mail: [alvaro@ilch.uminho.pt](mailto:alvaro@ilch.uminho.pt)

### **ANA SUELLY ARRUDA CÂMARA CABRAL**

Doutora em Linguística pela University of Pittsburg (1995). Professora Associada, Coordenadora do (LALLI), do Instituto de Letras da UNB - Universidade de Brasília. Brasília. DF. Brasil. CEP: 70910-900. E-mail: [asacc@unb.br](mailto:asacc@unb.br)

### **ANISE DE ABREU GONÇALVES D'ORANGE FERREIRA**

Doutora em Letras Clássicas pela USP (2002) e em Psicologia Experimental pela USP, Universidade de São Paulo (1993). Membro dos grupos Alter e Linceu, e do consórcio internacional Sunoikisis DC. Docente da UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”. Faculdade de Ciências e Letras – Departamento de Linguística. Araraquara. SP. Brasil. CEP: 14800-901. E-mail: [anise.a@gmail.com](mailto:anise.a@gmail.com)

### **APARECIDA NEGRI ISQUERDO**

Doutora em Letras (Linguística e Língua Portuguesa) pela Universidade Estadual Paulista “Júlio de Mesquita Filho” UNESP-Araraquara(1996). Docente aposentada da UFMS – Universidade Federal do Mato Grosso do Sul. Campo Grande. MS. Brasil. CEP: 79070-90. E-mail: [anegri.isquerdo@terra.com.br](mailto:anegri.isquerdo@terra.com.br)

### **CLAUDIA ZAVAGLIA**

Livre-Docente em Lexicografia e Lexicologia pela UNESP (2009). Doutora em Linguística e Língua Portuguesa pela UNESP (2002). Docente da UNESP, Universidade Estadual Paulista “Júlio de Mesquita Filho”, IBILCE, Departamento de Letras Modernas. São José do Rio Preto. SP. Brasil. CEP: 15054-000. E-mail: zavaglia@ibilce.unesp.br

### **CLOTILDE DE ALMEIDA AZEVEDO MURAKAWA**

Doutora em Letras, Linguística e Língua Portuguesa pela UNESP (1991). Docente aposentada, atuando no Programa de Pós-Graduação em Linguística e Língua Portuguesa, UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho” – Faculdade de Ciências e Letras. Araraquara. SP. Brasil. CEP: 14800-901. E-mail: jtm.jau@uol.com.br

### **CONCEIÇÃO DE MARIA DE ARAUJO RAMOS**

Doutora em Linguística pela UFAL - Universidade Federal de Alagoas (1999). Coordenadora do ALiMA. Professora associada IV da UFMA - Universidade Federal do Maranhão. São Luís. MA. Brasil. CEP: 65080-805. E-mail: conciefma@gmail.com

### **CRISTINA DE CÁSSIA BORELLA**

Mestre em Linguística pela UNICAMP - Universidade Estadual de Campinas (2000). Docente da UFAM - Universidade Federal do Amazonas. Manaus. AM. Brasil. CEP: 69060-000. E-mail: cristina.borella@gmail.com

### **CRISTINA MARTINS FARGETTI**

Livre Docente em Línguas Indígenas pela UNESP (2015). Doutora em Linguística pela UNICAMP - Universidade Estadual de Campinas (2002). Líder do grupo LINBRA. Docente da UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho” – Faculdade de Ciências e Letras – Departamento de Linguística. Araraquara. SP. Brasil. CEP 14800-901. E-mail: cmfarget@gmail.com

### **ENEIDA ALICE DOS SANTOS**

Mestre em Linguística pela UNB – Universidade de Brasília (2008). Docente da UFAM - Universidade Federal do Amazonas. Manaus. AM. Brasil. CEP: 69060-000. E-mail: eneidaalice@gmail.com

### **FABIANE CRISTINA ALTINO**

Doutora em Estudos da Linguagem pela UEL (2007). Docente na UEL - Universidade Estadual de Londrina, Centro de Letras e Ciências Humanas - Departamento de Letras Vernáculas e Clássicas. Londrina. PR. Brasil. CEP 86057-970. E-mail: fabiane\_altino@uol.com.br

### **GISELE APARECIDA RIBEIRO**

Doutora em Estudos Linguísticos pela UFMG - Universidade Federal de Minas Gerais (2016). Docente da UEMG - Universidade do Estado de Minas Gerais. Passos. MG. Brasil CEP: 37900106. E-mail: gisele.ribeiro@uemg.br

### **GUSTAVO LUIZ VIEIRA**

E-mail: gustavoluizvieira@gmail.com

### **JÉSSICA PAULA VESCOVI**

Mestre em Letras -Linguagem e Sociedade pela UNIOESTE (2014). Docente colaboradora da UNIOESTE - Universidade Estadual do Oeste do Paraná. Marechal Cândido Rondon. PR. Brasil. CEP: 85960-000. E-mail: gilvescovi@hotmail.com

### **JORGE DOMINGUES LOPES**

Doutor em Linguística pela UNB - Universidade de Brasília (2014). Docente da UFPA – Universidade Federal do Pará, Campus Universitário do Tocantins/Cametá. Cametá. PA. Brasil. CEP: 68400-000. E-mail: jdlopes@ufpa.br

### **JOSÉ DE RIBAMAR MENDES BEZERRA**

Doutor em Linguística pela UFAL - Universidade Federal de Alagoas (2001). Pesquisador do ALiMA. Professor associado II da UFMA - Universidade Federal do Maranhão. São Luís. MA. Brasil. CEP: 65080-805. E-mail: comendesufma@gmail.com

### **LAÍSA BAUERMEISTER STELO**

Graduada em Letras pela UFMS. E-mail: laisa.stelo@gmail.com

### **LUÍS HENRIQUE SERRA**

Doutorando em Filologia e Língua Portuguesa na USP – Universidade de São Paulo. Docente da UFMA - Universidade Federal do Maranhão, Departamento de Pedagogia, campus VII. Codó. MA. Brasil. CEP: 65085580. E-mail: luis.ufma@gmail.com

### **MARANÚBIA PEREIRA BARBOSA DOIRON**

Doutora em Estudos da Linguagem pela UEL - Universidade Estadual de Londrina (2017). E-mail: maranubiadoiron@gmail.com

### **MARIANA GIACOMINI BOTTA**

Doutora em Linguística e Língua Portuguesa pela UNESP - Universidade Estadual Paulista “Júlio de Mesquita Filho” (2011). Docente no UniRitter - Centro Universitário Ritter dos Reis. Porto Alegre. RS. Brasil. CEP: 05508-900. E-mail: marianabotta@yahoo.com.br

### **MÁRCIA SIPAVICIUS SEIDE**

Doutora em Letras de USP – Universidade de São Paulo (2006). Docente na UNIOESTE - Universidade Estadual do Oeste do Paraná, Centro de Ciências Humanas, Educação e Letras. Marechal Cândido Rondon. PR. Brasil. CEP: 85960-000. E-mail: marciaseda4@hotmail.com

### **MARCO DIMAS GUBITOSO**

Doutor em Ciências da Computação pela USP (1996). Docente da USP – Universidade de São Paulo, Instituto de Matemática e Estatística. São Paulo. SP. Brasil. CEP: 05508-900. E-mail: marco.gubitoso@gmail.com

### **MARIA DE FÁTIMA SOPAS ROCHA**

Doutora em Linguística pela UFC - Universidade Federal do Ceará (2013). Pesquisadora do ALIB. Docente da UFMA - Universidade Federal do Maranhão. São Luís. MA. Brasil. CEP: 65080-805. E-mail: fsopas@yahoo.com.br

### **MARIA JOSÉ BOCORNY FINATTO**

Doutora em Letras pela UFRGS (2001). Vice-coordenadora do TERMISUL. Docente da UFRGS - Universidade Federal do Rio Grande do Sul, Departamento de Linguística, Filologia e Teoria Literária. Porto Alegre. RS. Brasil. CEP: 91540000. E-mail: mariafinatto@gmail.com

### **MÁRIO EDUARDO VIARO**

Doutor em Filologia Românica pela USP (2001). Vice-Coordenador do NEHiLP. Docente da USP - Universidade de São Paulo. São Paulo. SP. Brasil. CEP: 05508-900. E-mail: maeviaro@usp.br

**NELLY CARVALHO**

Doutora em Letras pela UFPE (1993). Membro da diretoria da Aliança Francesa, conselheira do Conselho Estadual de Educação de Pernambuco. Professora Emérita da (UFPE) - Universidade Federal de Pernambuco. Recife. PE. Brasil. CEP: 50740-530. E-mail: nellycar@terra.com.br

**ODAIR LUIZ NADIN DA SILVA**

Doutor em Linguística e Língua Portuguesa pela UNESP (2008). Docente da UNESP - Universidade Estadual Paulista “Júlio de Mesquita Filho” - Faculdade de Ciências e Letras – Departamento de Letras Modernas. Araraquara. SP. Brasil. CEP 14800-90. E-mail: odairnadin@fclar.unesp.br

**ROSANA MARIA SANT’ANA COTRIM**

Doutora em Linguística e Língua Portuguesa pela UNESP - Universidade Estadual Paulista “Júlio de Mesquita Filho” (2012). Docente da UFMT - Universidade Federal de Mato Grosso. Rondonópolis. MT. Brasil. CEP: 78735-000. E-mail: rocotrim@outlook.com.br

**VANDERCI DE ANDRADE AGUILERA**

Doutora em Letras pela UNESP - Universidade Estadual Paulista “Júlio de Mesquita Filho” (1990). Coordenadora do ALIB. Docente na UEL - Universidade Estadual de Londrina - Centro de Letras e Ciências Humanas - Departamento de Letras Vernáculas e Clássicas. Londrina. PR. Brasil. CEP 86057-970. E-mail: vanderci@uel.br



## SOBRE O VOLUME

Série Trilhas Linguísticas, n.32

Formato: 14 x 21 cm

Mancha: 10 x 18,5 cm

Tipologia: Garamond 11/13,5

Papel: Pólen Bold 90 g/m<sup>2</sup> (miolo)

Cartão Supremo 250 g/m<sup>2</sup> (capa)

1ª edição: 2019

Para adquirir esta obra:

STAEPE – Seção Técnica de Apoio ao Ensino, Pesquisa e Extensão

Laboratório Editorial

Rodovia Araraquara-Jaú, km 01

14800-901 – Araraquara

Fone: (16) 3334-6275

E-mail: [laboratorioeditorial.fclar@unesp.br](mailto:laboratorioeditorial.fclar@unesp.br)

Site: <http://www.fclar.unesp.br/laboratorioeditorial>

