

O DICIONÁRIO ETIMOLÓGICO DA LÍNGUA PORTUGUESA (DELPo): CONCEITOS DE METALEMA, HEMILEMA, HIPERLEMA E ULTRALEMA

Mário Eduardo VIARO¹

RESUMO

O presente texto visa à divulgação das ferramentas computacionais desenvolvidas pelo Núcleo de apoio à pesquisa em Etimologia e História da Língua Portuguesa (NEHiLP/USP) para a confecção do Dicionário Etimológico da Língua Portuguesa (DELPo), as quais são necessárias para a organização hierárquica de dados referentes tanto à etimologia quanto à origem das palavras investigadas. Essas ferramentas são apresentadas de maneira conceitual, uma vez que formam os pressupostos da construção daquilo que a Lexicografia tradicionalmente chama de *lema*.

PALAVRAS-CHAVE: Lexicografia; Banco de Dados; Lema; Etimologia; Linguística Computacional.

1. A tarefa de coleta para um dicionário etimológico

Uma das tarefas do etimólogo é *grosso modo* coletar palavras extraídas de contextos em determinadas obras e associá-las à sua data de publicação. Consabidamente, essas obras deve ser edições confiáveis. Por meio da comparação dessas datações, é possível estabelecer a data mais antiga da palavra investigada e estabelecer etimologias. Tradicionalmente, o etimólogo é entendido como uma pessoa extremamente erudita, que sabe onde localizar as melhores fontes e não raro conhece detalhes de conteúdo das obras que consulta/cita. Essa postura, contudo, é ultrapassada pelo avanço tecnológico atual.

Obviamente, ter erudição e ser organizado são qualidades sempre desejáveis para o pesquisador desse trabalho, mas não é o ponto principal, se quisermos garantir a

¹ USP - Faculdade de Filosofia, Letras e Ciências Humanas - Departamento de Letras Clássicas e Vernáculas. Endereço: R. Luciano Gualberto, 403 - FFLCH - DLCV. CEP 05508-900 Cidade Universitária - São Paulo/SP - Brasil. Endereço eletrônico: nehilp@usp.br

fidedignidade e a *veracidade* dos dados, e mais ainda o *valor* desses dados. O ideal é que o menor número de erros seja cometido. Mesmo que saibamos que erros dependentes exclusivamente das propensões humanas sempre existirão, há alguns tipos de erros que podem ser evitados, como os motivados por cansaço e pela falta de atenção, dado o manuseio de um grande número de informações. Nesse ponto, acreditou-se, desde o início, que seria de extrema utilidade a existência de um programa que fizesse a comparação automática da data atribuída à mais antiga abonação da palavra com a data da obra analisada.

As vantagens de um programa computacional sobre a pesquisa manual seriam várias,² dentre as quais:

(a) numa pesquisa manual, o pesquisador não tem disponíveis em sua erudição os *termini a quo* de todas as palavras de uma língua, de modo que suas descobertas se pautam, na maioria das vezes, apenas pela “sensação” de que a palavra “não deveria estar naquele texto”, por ele supor (baseado exclusivamente na sua experiência de falante e/ou pesquisador) que a palavra seja mais recente do que a data do texto investigado;

(b) mesmo se for extremamente organizado, o pesquisador não consegue fazer uma investigação exaustiva de um texto e pauta-se apenas naqueles dados que chamam sua atenção. Assim sendo, outros dados importantes acabam por passar involuntariamente despercebidos ao escrutínio do investigador, por mais minucioso que seja;

(c) para sanar a impossibilidade de uma investigação exaustiva, o pesquisador acaba às vezes especializando-se, e isso não contribui para a criação de um amplo dicionário etimológico. Assim sendo, seu recorte acabará necessariamente incidindo apenas na forma (por exemplo, procura palavras com um determinado sufixo) ou no significado (por exemplo, busca palavras de um determinado campo semântico), quando sua busca não é enviesada por uma questão *a posteriori* (por exemplo, procura palavras de suposta origem africana).

2 Outra parte do projeto do DELPo se encontra em um artigo (ainda no prelo) de Viaro et al. Sobre a pesquisa de Linguística baseada em ferramentas computacionais há obras de referência como a de Sardinha (2004).

A conclusão é que o mesmo texto precisaria ser revisto inúmeras vezes por uma grande quantidade de pesquisadores para estarmos seguros que dele foram extraídas todas as informações interessantes para um estudo etimológico.

2. Estrutura do Dicionário Etimológico da Língua Portuguesa (DELPo)

Para a confecção do DELPo (Dicionário Etimológico da Língua Portuguesa), projeto principal do Núcleo de pesquisa em Etimologia e História da Língua Portuguesa (NEHiLP, endereço: www.nehilp.org), concentrou-se não sobre os lemas (tais como entendidos tradicionalmente pela lexicografia), mas sobre as acepções, uma vez que as datações das acepções podem levar à datação do lema principal.

Um exemplo bastaria para fundamentar essa decisão: num dicionário temos lemas que são sempre distintos entre si, exceto se forem homônimos. Por *homonímia* a lexicografia entende o fenômeno da semelhança (por vezes meramente gráfica) entre dois lemas que têm étimos diferentes. No caso de o étimo ser igual, essas formas são chamadas de *acepções*. Contudo, flagrou-se aqui uma contradição: se o objetivo último do NEHiLP é a determinação das etimologias, não pareceu ser uma boa decisão definir *a priori* se duas palavras com mesma sequência de caracteres são homônimas ou fruto de polissemia. Além disso, havia complicações técnicas para a determinação desses conceitos. Suspendeu-se, nesse momento, a distinção entre *homonímia* e *polissemia*, com vistas à solução desse problema técnico, e nasceu daí um conceito novo, o de *metalema*³.

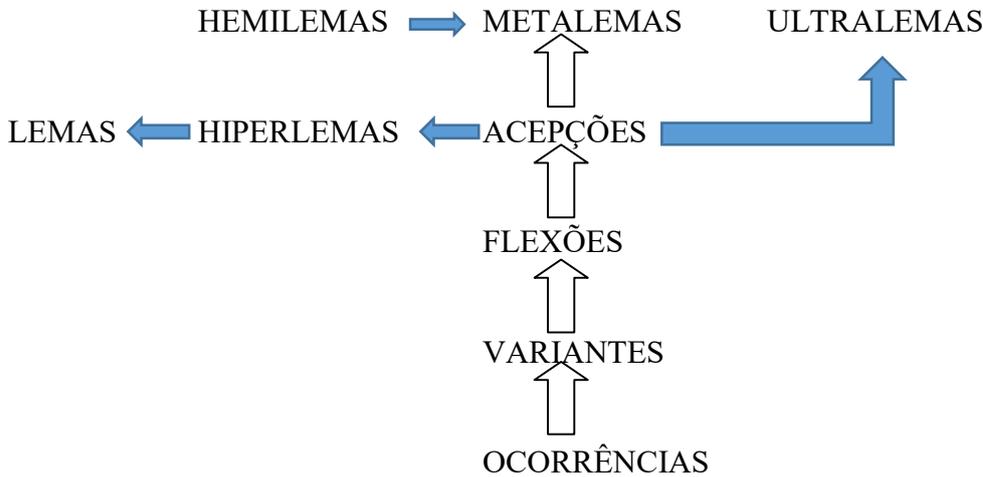
Obviamente, foi construída toda uma estrutura para que, ao analisarmos os dados, pudéssemos recuperar *a posteriori* as condições de homonímia e polissemia com os chamados *hiperlemas*, que uma vez aprovados se tornariam nos *lemas* utilizados pelos consulentes. Por fim, conjuntamente com outros pesquisadores, concluiu-se que palavras abreviadas, truncadas e mutiladas não podem ter o mesmo *status* de palavras legíveis e completas, e, nesse momento, surgiu o conceito de *hemilemas*⁴. A última estrutura apresentada foi a separação dos dados do dicionário com vistas à etimologia

3 O termo *metalema* surgiu, espontaneamente, em discussões sobre a estrutura do banco de dados, no IME-USP, durante uma reunião ocorrida em 08/10/2013, às 17h21.

4 Inicialmente chamado de *paralema*.

recente daqueles que explicitam a etimologia remota e a linguística histórico-comparativa e, com isso, surgiram os *ultralemas*⁵.

Conforme o estabelecimento teórico decidido em reuniões do NEHiLP e sua implementação técnica posterior, os dados linguísticos armazenados para o DELPo obedecem a seguinte estrutura:



Essas estruturas hierárquicas são importantes de ser visualizadas para o entendimento das questões apresentadas a seguir com maior detalhamento.

3. O Metalema

Um metalema difere de um lema tradicional em vários sentidos. Computacionalmente, o metalema é apenas uma sequência de caracteres, e não existem metalemas homônimos no Dicionário. Sob o mesmo metalema estarão palavras homônimas, como *manga* (fruta) e *manga* (parte do vestuário).

Do ponto de vista lexicográfico, o *lema* é a forma básica da entrada do verbete de dicionário, visto que o lema é, na verdade, um feixe de flexões⁶ (ao menos nas línguas flexivas como o português). Além disso, o conceito de *lema*, do ponto de vista lexicográfico, depende da etimologia, uma vez que duas palavras homógrafas podem

5 Os termos *hemilema*, *hiperlema* e *ultralema* surgiram em julho de 2014 para denominar essas estruturas computacionais.

6 Estritamente falando, mesmo palavras invariáveis (preposições, advérbios, conjunções e interjeições) devem ser vistas como um conjunto de variantes flexionais, apenas com a ressalva de que, nesse caso, se trataria de um conjunto unitário. Subordinadas às *flexões* estão, na lógica do banco de dados, as *variantes ortográficas*. Variantes não ortográficas de uma mesma palavra são recuperáveis por meio dos *hiperlemas*.

ser, dependendo do caso, dois lemas ou um só, dependendo se são, respectivamente, fruto de homonímia (etimologias distintas) ou de polissemia (têm uma origem comum⁷). Essa distinção, preciosa para o NEHiLP, é, no entanto, embaraçosa para a computação, uma vez que tais distinções não conseguem ser feitas *a priori*. Além disso, o NEHiLP entende que a certeza de estarmos, em alguns casos complexos, diante de uma homonímia ou de uma polissemia é fruto da própria investigação etimológica. Assume-se, portanto, mesmo para os casos aparentemente óbvios, que a decisão de distintas acepções homógrafas serem casos de homonímia ou polissemia é algo *a posteriori* (que será feito pelo *hiperlema*).

Uma vez que praticamente todas as palavras são polissêmicas e algumas delas possuem homônimos, define-se:

metalema é a forma significativa que corresponde a uma sequência única de letras na língua, não importa a quantos lemas ou acepções de lemas distintos corresponda.

As fichas do Dicionário adotam, para o *metalema*, a mesma metodologia da lexicografia tradicional utilizada para o lema, ou seja:

- O *metalema* de um verbo é sua flexão no infinitivo;
- O *metalema* do substantivo é sua flexão no singular;
- O *metalema* de um adjetivo é sua flexão no masculino singular.

Além disso, é adotada a ortografia atual para o *metalema*. Portanto, os *metalemas* (e também as flexões) seguirão a grafia atual, pois apenas as variantes mantêm a grafia do documento do qual foram extraídas. Isso significa, na prática, que, para chegar à forma de *metalema*, a forma lematizada da variante passa ainda por uma *atualização ortográfica*, que respeita as seguintes convenções:

- O uso dos acentos na versão atual é respeitado (*carâ* → *cará*, *magica* → *mágica*, *idéia* → *ideia*, *bêsta* → *besta*);
- O uso atual de maiúsculas é respeitado (*lisboa* → *Lisboa*, *Rei* → *rei*)⁸;
- Grafias etimológicas que envolvam as letras *y* e *h* são atualizadas (*arraya* → *arraia*, *physica* → *física*, *asthma* → *asma*, *pharmacia* → *farmácia*, *omem* → *homem*, *hontem* → *ontem*);
- Consoantes dobradas (exceto *rr*, *ss* e casos excepcionais) são convertidas em simples (*accusar* → *acusar*, *elle* → *ele*, *anno* → *ano*);

⁷ Para a distinção entre *étimo* e *origem*, vide Viaro 2001: 106-109.

⁸ Somente topônimos e antropônimos terão o *metalema* e a flexão escritos em maiúsculo.

- Vogais dobradas (exceto em casos excepcionais de conservação na ortografia atual) são convertidas em simples (*coor* → *cor*, *irmãa* → *irmã*)⁹;
- O til deve ser substituído por *m/n* antes de consoantes e no final de palavras (*cãssado* → *cansado*, *coïbra* → *Coimbra*), exceto se o til permanece na grafia atual, como em *amanhã*;
- As terminações *-am/-om* de substantivos e suas variantes se converterão em *-ão* (*rrezom* → *rezão*), exceto em palavras como *acordeom*, *marrom* e *edredom*¹⁰;
- Palavras que estejam juntas/separadas de uma forma distinta da ortografia atual precisam ser preparadas previamente.

Atualizam-se as grafias:

- *i/j/y* (*jgreya* → *igreja*, *ieyum* → *jejum*);
- *u/v* (*vntar* → *untar*, *cauallo* → *cavalo*);
- *s/ss* (*ssabio* → *sábio*, *asar* → *assar*), mas não entre *c*, *sc*, *xc*, *ç*, *sç*, *xç*;
- *s/z/x* antes de consoante ou em final de palavra (*ezquerdo* → *esquerdo*, *ferox* → *feroz*);
- *c/ç* (*reçeber* → *receber*, *acucar* → *açúcar*);
- *r/rr* (*rrei* → *rei*, *amarar* → *amarrar*);
- *g/j/gu* (*guato* → *gato*, *loguo* → *logo*, *gamays* → *jamais*);
- *gu/gü/go* (*lingoa* → *língua*, *lingüiça* → *linguiça*);
- *c/k/q/qu* (*quasa* → *casa*, *qeira* → *queira*, *kalenda* → *calenda*);
- *qu/qü/co* (*cincoenta* → *cinquenta*, *freqüente* → *frequente*);
- *e/i* e *o/u* postônicos e semivogais (*gatu* → *gato*, *titolo* → *título*, *cabi* → *cabe*, *escarneo* → *escárnio*, *cacao* → *cacau*);
- *m/n* antes de consoante e no final de palavras (*emtemder* → *entender*, *conprido* → *comprido*, *cãosado* → *cansado*);
- *c=ç/s=ss* (*çapato* → *sapato*, *alicerse* → *alicerce*) ou *s/x/z* antes de vogal (*caza* → *casa*)¹¹.

9 Esta decisão não foi fácil, uma vez que uma palavra como *coor* pode ter, dependendo da época e do local, uma sílaba ou duas. No entanto, como a avaliação disso depende muito de pressupostos assumidos por autores especialistas e há discordâncias, preferiu-se tomar essa decisão.

10 Como no caso das vogais dobradas, essa decisão não foi simples de ser tomada, mas em muitas situações (sobretudo depois do século XV) é muito difícil decidir se a grafia corresponde ou não à pronúncia e, uma vez que há discussões entre os autores sobre o problema, convencionou-se atualizar também essas formas.

Por outro lado, não se padronizam:

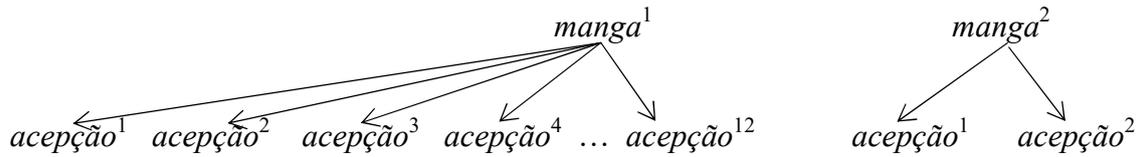
- *e/i* e *o/u* tônicos e pretônicos (*filis* → *filiz* e não *feliz*, *aquelo* → *aquelo* e não *aquilo*);
- vogais epentéticas (*idea* → *idea* e não *ideia*);
- vogais nasais em hiato (*corõa* → *corõa* e não *coroa*);
- oscilações entre *e/em* finais (*ome* → *home* e não *homem*);
- oscilações entre *b/v* (*estorbar* → *estorbar* e não *estorvar*);
- oscilações entre *l/r* (*pubrico* → *público* e não *público*);
- outras variações de consoantes (*celebro* → *célebro* e não *cérebro*);
- metáteses (*estrobar* → *estrobar* e não *estorvar*);
- ausências de sons (*sobiar* → *sobiar* e não *assobiar*)¹².

Um exemplo do tratamento distinto do NEHiLP em relação à Lexicografia tradicional: segundo o dicionário de Houaiss & Villar (2001), o lema *manga*¹, que significa “parte de vestimenta”, remonta ao latim *manīca*; já seu homônimo *manga*², “fruto da mangueira”, vem do malaiala *manga* (por sua vez originário do tâmil *mānkāy*). Pois bem, *manga*¹ tem 12 acepções distintas (apenas 8 na versão eletrônica de 2009.3) e *manga*² pelo menos duas. Além disso, o dicionário Houaiss estabelece o *terminus a quo* de *manga*¹ no século XIII e o de *manga*² em 1554. Para a primeira forma, não há variantes ortográficas, mas para *manga*² menciona-se *manguas* (1554), *mangue* (1583) e *mangas* (1616), que, ao mesmo tempo, são variantes e flexões dos lemas acima.

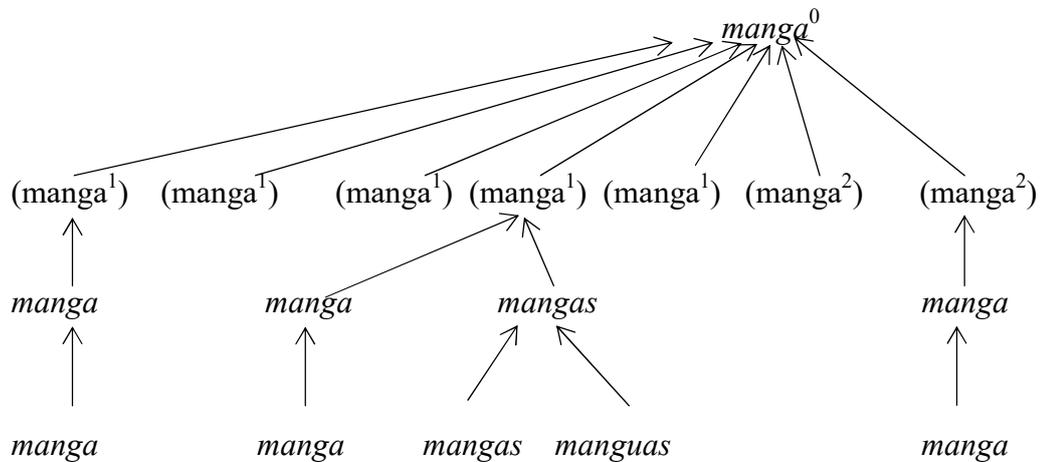
A estrutura de um dicionário criado sob os parâmetros da Lexicografia tradicional privilegiaria, portanto a seguinte hierarquia, estritamente do ponto de vista etimológico e semântico (a saber *manga*¹ e *manga*² estão separadas por *homonímia*, definida como palavras de étimo distinto, ao passo que as acepções subordinadas a uma palavra, cujo étimo pode não ser idêntico, mas sua origem sim, formam a sua *polissemia*):

11 Como no caso das vogais dobradas e das terminações nasais, essa decisão varia muito conforme os séculos, e é muito difícil decidir se a grafia corresponde ou não à pronúncia. Visto que há discussões entre os autores sobre o problema, convencionou-se atualizar também essas formas.

12 Casos especiais: *ũa* → *uma* (flexão), mas *um* (metalema) e *algũa* → *alguma* (flexão), mas *algun* (metalema).



No entanto, organizar essa estrutura apenas seria possível se os significados de *manga*¹ e de *manga*² pudessem ser identificados automaticamente, algo que o programa *Moedor* não consegue fazer, pois apenas consegue apresentar *sugestões de lema*¹³. Dessa forma, o programa analisa apenas a sequência de caracteres *manga* à qual se associam variantes ortográficas já flexionadas, extraídas de um texto. A atribuição de sentido deve ser feita pelo pesquisador. O Dicionário, porém, inverte a situação. Imagina-se uma *manga*⁰ que seja apenas uma cadeia de caracteres que equivale parcialmente à sugestão de lema das moagens. Essa *manga*⁰ é o significante não só de *manga*¹ e de *manga*² mas de todas as acepções de *manga*¹ e de *manga*². O trabalho de vinculação hierárquica por meio da etimologia é estabelecido *a priori* na lexicografia tradicional, mas é uma tarefa *a posteriori* do NEHiLP. Assim sendo, muitas das informações presentes marginalmente num dicionário são aproveitadas e organizadas diferentemente no Dicionário:



Desse modo, dizemos que *manga*⁰ é o metalema de todas as acepções de *manga*, que, por sua vez se manifestam na forma de flexões, as quais se concretizam em variantes ortográficas. Só a *variante* é real. Os demais níveis acima dela (flexões, acepções e metalemas) são abstrações e seguem atualizações ortográficas. Se a antiga grafia *manguas* é considerada como variante ortográfica da flexão *mangas*, isso se justifica por meio da pesquisa filológica. Já *mangues*, que aparece no dicionário Houaiss, não pertence ao metalema *manga*⁰ e sim ao metalema *mangue*⁰, pois não se

13 A sugestão de lema, contudo, segue um processo inteligente de autoalimentação, ou seja, aprimora-se à medida que são preenchidas as fichas do DELPo.

trata de uma variante ortográfica, mas de outro som, associado a outro fonema. A recuperação de *mangues* como variante não ortográfica de *manga* se efetua por outro caminho, ou seja, pelo *hiperlema*.

Da mesma forma que, baseado numa *variante lematizada e atualizada ortograficamente* obtemos o *metalema*, também por meio do mesmo raciocínio chegamos à *flexão*, que, diferentemente do *metalema*, precisa ser digitada na ficha. A única diferença da forma que aparece no campo do *metalema* e no da *flexão* é o fato de a *flexão* não ser lematizada. Ou seja, apesar de sua ortografia ser atualizada, respeitam-se as flexões da variante escolhida como abonação da ficha.

Exemplos:

- se a variante analisada é *primeyras* (com ortografia antiga, no feminino plural), o campo do *metalema* deverá aparecer como *primeiro* (na ortografia atual e masculino singular, como nos dicionários). No campo da *flexão* deve-se digitar *primeiras* (na ortografia atual, mas no feminino plural).
- Igualmente, se a variante analisada é *disserom*, o campo do *metalema* deve aparecer como *dizer* e no da *flexão* digita-se *disseram*.
- Por fim, se a variante for *assouiassemos*, o *metalema* deve ser *assoviar* e a *flexão*, *assoviássemos* (e não “assobiar” e “assobiássemos” porque a mudança na formação do *metalema* e da *flexão* digitadas deve ser ortográfica e não fonética).

4. Hemilema

Um *hemilema* é uma espécie de palavra que necessita de uma interpretação de uma acepção (por sua vez ligada a um *metalema*) ou de várias interpretações. Esse recurso é usado para abreviaturas em geral e para palavras truncadas (devido ao mau estado de conservação do manuscrito). Muitas vezes é possível interpretar também a *flexão*, por exemplo, “*mtas*” é um *hemilema* que tem a interpretação de uma certa acepção de “muito”, mas reflete a forma flexionada “*muitas*”. Igualmente, pelo contexto, é possível interpretar que o *hemilema* “*qr*” é a *flexão* “*quer*” de uma determinada acepção de “querer”.

O conceito de *hemilema* nasceu da tendência do NEHiLP a privilegiar as edições diplomáticas às críticas, uma vez que para o estudo etimológico é preciso atribuir graus de certeza aos dados parcialmente documentados, como abreviaturas, palavras truncadas por dano nos manuscritos e outras ocorrências lexicais que requerem interpretação extra e erudição. Em textos muito antigos, as abreviaturas e palavras truncadas podem dar margem a muitas interpretações, e escolher uma única dessas interpretações para vinculá-las a uma acepção é bastante arriscado. Mesmo nos casos em que as abreviaturas são óbvias, convencionou-se tratá-las sempre como um hemilema, e sua consulta não será no dicionário, mas na busca de abreviaturas. A discussão etimológica no campo *Etimologia* das fichas de acepção fará menção às abreviaturas, no caso de retrodatações baseadas em interpretações das mesmas, mas isso não pode ser feito automaticamente como nas demais variantes.

A decifração de um hemilema pode conduzir a uma palavra flexionada; no entanto, os hemilemas remetem às acepções e não às suas flexões: por exemplo, *mtos* ou *mta* é interpretado como uma das acepções do metalema *muito* e não às flexões *muitos* ou *muita*, respectivamente. Também são considerados hemilemas formas como *senhor(a)s*, *amigo(a)(s)*, *amigx*, *amigxs*, *amigx(s)*, *amig@s*, *amig@(s)*, etc.

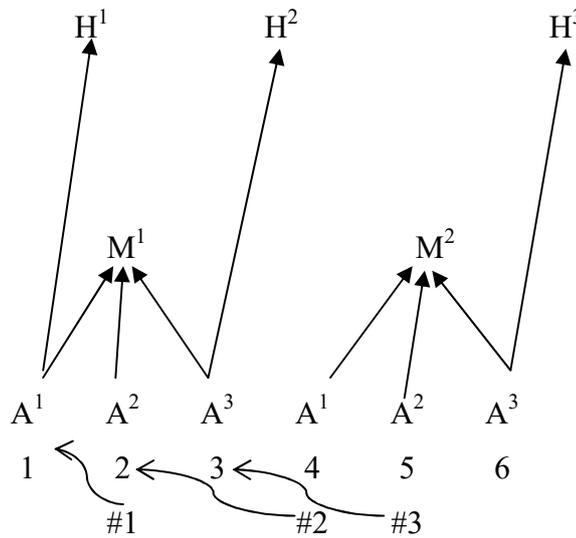
É preciso observar que uma *sigla* é distinta de uma abreviatura, pois a decomposição de uma sigla pode equivaler a mais de uma palavra (o que, na verdade, é sua etimologia). Normalmente, as siglas são escritas somente com letras maiúsculas (por exemplo, USP), mas há algumas siglas que se tornam nomes comuns e somente o pesquisador poderá diferenciá-las. Convencionou-se que uma sigla não é um hemilema e sim um metalema.

5. O Hiperlema

O campo *Etimologia* das fichas do banco de dados do DELPo pode ser preenchido com uma única informação, a saber, uma *remissão*. Nesse caso, diz-se que a acepção é *subordinada* a uma outra. A remissão tem o seguinte formato:

ID da acepção subordinante

Uma acepção subordinada pode remeter a uma acepção subordinante sob o mesmo metalema ou em outro metalema. A acepção subordinante que não for ao mesmo tempo subordinada chama-se *acepção principal*. Toda acepção principal gera um hiperlema, como pode ser visto na *Busca Homonímia*. Um metalema pode ter duas ou mais acepções principais. Nesse caso gerará *hiperlemas homônimos*¹⁴. Dessa forma, distingue-se homonímia e polissemia sob um mesmo metalema. Exemplo:



Na figura acima, o primeiro metalema tem duas acepções principais (subordinantes e não subordinadas) e o segundo, apenas uma, o que resulta em três hiperlemas. A acepção 2 (isto é, de ID 2) é subordinante em relação à acepção 4, mas é subordinada em relação à acepção 1.

O hiperlema gerado H¹ terá as acepções (1), (2) e (4), ou seja, ou seja, as acepções A¹ e A² do metalema M¹ e A¹ de M²; o hiperlema H² terá as acepções (3) e (5), isto é, A³ de M¹ e A² de M²; finalmente, o hiperlema H³ terá a acepção (6) ou A³ de M². Cada hiperlema gerará lemas, da seguinte forma para a consulta externa:

O lema L¹ terá dois homônimos L¹₁ e L¹₂, correspondentes a H¹ e a H² e o lema L² terá três lemas homônimos L²₁, L²₂ e L²₃. O lema L²₁ remeterá a L¹₁; o lema L²₂ remeterá a L¹₂ e o lema L²₃ remeterá a H³.

6. Ultralema

Algo digitado no campo *Origem* nas fichas do banco de dados do DELPo gera

¹⁴ Como já dito, não há metalemas homônimos, mas há hiperlemas homônimos.

um *ultralema*. Também é possível gerá-las manualmente. Os ultralemas foram criados para discussão de radicais, raízes e elementos de formação e distinguem-se em *ultralemas comuns* e o *ultralemas remotos*, dependendo do grau de antiguidade desses elementos. O formato dessa inserção deve ser:

língua ULTRALEMA

Se o ultralema for um sufixo, deve ter o formato *-x-* (se for um sufixo não-final, que requer, por exemplo, uma vogal temática) ou *-x* (se for um sufixo final); um prefixo deverá ser *x-*. Demais elementos, como raízes, são escritos sem hífen à esquerda ou à direita. No campo *Origem remota*, devem vir informações semelhantes às que aparecem no campo *Origem*; no entanto, como se trata de uma língua reconstruída (*indo-europeu*, *camito-semítico*, *bantu*, *proto-jê*, etc.), o elemento de formação terá sempre um asterisco antes:

língua *ELEMENTO DE FORMAÇÃO

Além do comentário etimológico, a utilidade dos ultralemas se vê no campo *Cognatos*, no qual estarão, automaticamente, todas as acepções cujos ultralemas foram declarados no campo *Origem* das suas fichas¹⁵.

Assim sendo, o metalema *cabo* deverá ter dois hiperlemas, um que agregue os significados de “início”, “cabo geográfico” e “cabo militar” e outro que agregue os de “cabo de panela” e “cabo de aço”.

Além de reunir as acepções do metalema de modo que se possam distinguir os lemas homônimos das acepções em polissemia, tal como entende a Lexicografia, outra utilidade do hiperlema está na escolha de um metalema principal num conjunto de metalemas alótropos. Por exemplo, “emboaba”, “imbuava” e “emboava” são três metalemas distintos, apesar de terem o mesmo significado, uma vez que a sua distinção não é meramente ortográfica. Para reuni-los é preciso que “imbuava” e “emboava” remetam, por exemplo, a “emboaba”, o qual, por sua vez, gerará automaticamente um hiperlema. No hiperlema “emboaba” estarão entre as suas acepções também as formas “imbuava” e “emboava”.

¹⁵ É preciso, portanto, organizar-se para que apenas a acepção principal tenha os ultralemas a fim de evitar duplicações desnecessárias.

Os dois hiperlemas que se formam da palavra **cabo** referem-se, cada um, a um ultralema distinto. Um deles, além da palavra “cabo”, também terá as palavras **cabeça**, **capital**, **decapitar**, etc. O outro, além da palavra “cabo”, também terá, por exemplo, palavras como **caber**, **receber**, **concepção**, etc. *Uma palavra pode remeter a mais de um ultralema. É o caso, por exemplo, de palavras formadas por composição. Por exemplo, a palavra **sanguessuga** tem duas raízes, portanto, dois ultralemas. Um deles, além do hiperlema **sanguessuga**, também está vinculado aos hiperlemas **sangue**, **sanguíneo**, etc. e o outro, além de **sanguessuga**, também está vinculado a **sugar**, **sucção**, etc.*

7. Conclusões

Diferentemente de pertencerem a alguma teoria lexicográfica, os conceitos de metalema, hemilema, hiperlema e ultralema são meramente operacionais, mas permitem, mediante correta análise construir aquilo que a Lexicografia chama de *lema*. Assim sendo, sob o ponto de vista da Etimologia, um *lema* é algo *a posteriori*. Não se sabe de antemão se duas formas lematizadas pertencem ou não ao mesmo lema sem uma investigação etimológica e sem operações mecânicas, resultado de decisões de organização. Nesse sentido, o *metalema* reúne formas lematizadas idênticas, mas mediante as remissões, o *hiperlema*, criado automaticamente, as separa, mas ainda é um passo anterior ao *lema*, que remete da forma tradicional (com “vide lema x”).

O vínculo entre as acepções e seus elementos de formação formam ainda um outro mecanismo de busca, tal como provam os *ultralemas*. Da mesma forma o vínculo entre *interpretações* de acepções e um lema abreviado, por exemplo, formam um terceiro mecanismo de busca distinto.

Poder-se-ia pensar num quarto mecanismo de busca, relacionado a determinadas “classificações semânticas” das acepções e que caracterizam buscas de antropônimos e topônimos. Mimetizar uma consulta de acordo com a tradição lexicográfica requer ainda este tipo de cuidado.

REFERÊNCIAS BIBLIOGRÁFICAS

Houaiss, Antônio; Villar, Mauro de S. 2001. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva.

Sardinha, Tony B. 2004. *Linguística de corpus*. Barueri: Manole.

Viaro, Mário E. 2011. *Etimologia*. São Paulo: Cortez.

Viaro, Mário E. *et alii*. No prelo. *O desafio da retrodatação para os estudos etimológicos de língua portuguesa*. Anais do I CINEFIL. Araraquara.

www.nehilp.org