

# RECONSTRUÇÃO LEXICAL DO PORTUGUÊS ANTIGO PAULISTA

Mário Eduardo VIARO<sup>1</sup>

■ **RESUMO:** Neste artigo discute-se o problema da reconstrução lingüística, em especial a reconstrução do antigo português paulista. Nessa tarefa, será levada em consideração a proposta de Jespersen para uma escala de incerteza em afirmações científicas a respeito de fenômenos diacrônicos. Mencionar-se-ão alguns aspectos da sistemática filogenética de Hennig e se considerará a possibilidade de sua aplicação em dados lingüísticos.

■ **PALAVRAS-CHAVE:** Reconstrução. Português. Diacronia.

## Reconstrução e cognição<sup>2</sup>

As chamadas regras fonéticas da gramática comparativa não surgiram no século 19, como se costuma afirmar nos manuais: elas já estão na gramática de Nebrija (1492) e, em língua portuguesa, aparecem pela primeira vez em Duarte Nunes de Leão (1576), embora somente tivessem sido estudadas mais exaustivamente pelo esquecido lingüista Francisco Evaristo Leoni (1858).

É natural que a comparação tenha sido o primeiro passo da lingüística. A *comparação* acompanha o discurso científico, não sendo nenhum exagero dizer que é um dos pilares cognitivos do ser humano, sem o qual não se pode conceber se dois objetos são ou não são iguais e mereçam assim o mesmo nome.

<sup>1</sup> Departamento de Letras Clássicas e Vernáculas – Faculdade de Filosofia, Letras e Ciências Humanas – USP 05508-900 São Paulo/SP Brasil – E-mail: maeviaro@usp.br

<sup>2</sup> O título do presente artigo é homônimo de um dos 11 subprojetos associados ao projeto *História do Português Paulista* (Projeto Caipira), coordenado pelo Prof. Dr. Ataliba Teixeira de Castilho (USP). Agradeço ao Prof. Dr. Nelson Papavero (Museu de Zoologia – USP) pela leitura cuidadosa do original e sugestões.

Um outro pilar é a *valoração*. Diante da multiplicidade dos seres, a língua atua dando pesos desiguais às coisas e separando o que é importante, bom e distintivo daquilo que não merece atenção. Os seres, naturalmente infinitos, díspares e inefáveis, se reúnem, como feixes, em signos com nomes e conceitos.

A *reconstrução* é o terceiro pilar. Servimo-nos dela quando estamos no papel de ouvintes ou leitores, valendo-nos da comparação e da valoração, para entender o que o outro diz. Remexendo as reminiscências de nossa memória, lembramo-nos também de experiências que nos auxiliam nessa reconstrução.

Por fim, o quarto pilar é a *criação*. Em vez de ficarmos impotentes com a dificuldade de reconstruir, propomos, conscientemente ou não, uma série de hipóteses e, assim, também podemos recontar o naturalmente incontável, por exemplo, aquilo que já passou.

Essas quatro bases sempre foram utilizadas pelos lingüistas em seus modelos e métodos científicos, independentemente da corrente a que se afiliam. A língua real está longe de ser um todo homogêneo. Não é e nunca foi.

Quando falamos em protolínguas, apenas por ingenuidade podemos pensar em conjuntos coesos de fenômenos. O ideal de língua uniforme, reforçado a partir do século 17 e aplicado às línguas nacionais, contudo, entra subrepticiamente nas discussões e nas afirmações de muitos lingüistas. Deparamo-nos com um código, composto de milhares de unidades de léxico, cada uma herdada à sua maneira. Pior que isso: cada flexão de cada palavra é herdada à sua maneira, como comprova a vitalidade e o surgimento das irregularidades que povoam as línguas. As regras sintáticas e os componentes gramaticais também são herdados e formam a única garantia de afiliação das línguas. Por isso, em vez de falarmos de *história das línguas*, expressas em árvores genealógicas, mais correto seria falar de uma *história do léxico* independente de uma *história dos componentes gramaticais*. Nos dois casos, o esquema arbóreo é problemático, apenas fundado no mito da monogênese em Babel. O que os dados nos mostram é exatamente o contrário: muitas variantes sempre convivem ao mesmo tempo. Se só uma sobrevive na norma culta é porque as outras deixaram de existir ou foram relegadas a variantes diatópicas ou diastráticas. Em outras palavras, quanto mais para trás nos voltamos, tantas mais variantes encontraremos: uma árvore ao contrário seria uma forma mais correta de

representar o que se passou. A aceleração do processo de seleção de uma única variante, como se vê nos últimos séculos, se deve mais ao *inutilia truncat* dos pré-iluministas do que propriamente a algum mecanismo natural da língua. A política de gramáticos como Vaugelas (1647) comprova isso muito bem. Eles também trouxeram a idéia de que a língua é lógica, subordinando assim a realidade à metalinguagem.

## **Existiu um protocaipira?**

Falar, portanto, de um protocaipira no sentido de um todo coeso seria um anacronismo. Nunca existiu algo assim, como nunca houve um único latim vulgar (VÄÄNÄNEN, 1962) ou um único indo-europeu (LEHMANN, 1993). As línguas reconstruídas devem assemelhar-se a línguas naturais. Devemos partir, portanto, sempre da variação, senão estaremos no terreno da fantasia.

Para tal, o que temos diante de nós são dados escritos. A freqüência de uso das variantes portuguesas que vieram para o Brasil é desconhecida. Mesmo assim, quanto mais o dado ocorrer (nos textos ou nas variantes faladas), maior é nossa certeza. O asterisco utilizado ao lado das formas reconstruídas é muito pobre, pois não revela o grau de certeza da reconstrução. Sobre isso já havia alertado Jespersen (1922):

It is of course, impossible to say how great a proportion of the etymologies given in dictionaries should strictly be classed under each of the following heads: (1) certain, (2) probable, (3) possible, (4) improbable, (5) impossible – but I am afraid the first two classes would be the least numerous.

Haudry (1994) faz distinção entre uma reconstrução segura do indo-europeu e uma hipótese: afirma que os dados apontam, em vários temas nominais, para um \*-s no nominativo de seres animados e um \*-m no acusativo de seres animados e neutros, bem como no nominativo dos neutros. Isso seria seguro e apontado por milhares de dados. No entanto, é possível imaginar, pela hipótese da ergatividade do indo-europeu, que um \*\*-s do nominativo animado se opunha a um \*\*-Ø do nominativo neutro, sendo \*-m marca apenas do acusativo (seja animado, seja neutro).

Cuidados de diferenciação de certeza, como esse, contudo, são raros e revelam forte embasamento filosófico, pois é costume misturarem-se informações obtidas indutivamente com hipóteses (que, por definição, são

dedutivas), criando modelos, teorias e outros construtos parcialmente científicos. Quando a história é totalmente descartada e só se vale da competência do falante, no entanto, tal confusão é ainda mais freqüente.

Reconstruir as bases do chamado dialeto caipira, portanto, muitas vezes revela-se tão difícil quanto reconstruir o indo-europeu. No século 16, se nos valermos de textos como o do Dr. Diogo Gouveia, o “Velho”, ao rei D. João 3º (1538), vemo-nos diante de fenômenos muito curiosos. Novamente, estamos reféns da língua escrita (e dos filólogos que estabeleceram o texto); no entanto, abstraindo isso, é possível localizar alguns indícios curiosos da língua falada na época. Mas, que significa exatamente estarmos diante de formas como *milhor*, *siguir*, *custume*, *emprestemo* (=empréstimo), *pidia*, *pubricamente*, *offericia*, *recibiam*, *vigairo*, *propios*, *exempro*, *testimunharom*? Antes de responder a isso, vejamos como outras ciências lidam com a reconstrução.

## Reconstrução em lingüística e em zoologia

Na zoologia, Hennig (1950) desenvolveu o método rigorosíssimo da sistemática filogenética, que forma o alicerce da filogenia moderna. Resumindo grosseiramente esse método, postula da seguinte forma: uma *espécie* é um conceito obtido a partir da soma dos *indivíduos* e de suas relações de parentesco e, portanto, não é um dado observável. Contrariamente ao senso comum, o próprio *indivíduo* é inobservável, uma vez que ele, muitas vezes, é irreconhecível em dois momentos de sua existência (qual seria, por exemplo, a semelhança entre o feto e o indivíduo adulto ou entre o ovo, a larva, a pupa e o inseto holometabólico adulto?). Cada fase morfológica do indivíduo é um *semaforonte*. Analogamente, dentro de uma mesma linhagem evolutiva, cada fase morfológica dessa linhagem é um *eidoforonte* (PAPAVERO; LLORENTE-BOUSQUETS, 1992). Em relação aos dados evolutivos, a maioria absoluta das espécies extintas não se preservou como fósseis. Mesmo com relação aos seres atualmente existentes, o conhecimento sobre eles é tremendamente limitado, uma vez que ele se revela sob inúmeros aspectos (biológicos, bioquímicos, citológicos, etológicos, fisiológicos, histológicos, morfológicos etc.): “a veracidade das afirmações sobre nosso conhecimento da evolução dos organismos limita-se aos *táxons* e aos *caracteres* amostrados” (AMORIM, 2002, p.19). Os *táxons* são construtos genéricos de caráter hiperonímico para as

classificações científicas (reinos, filos, classes, ordens, famílias, gêneros), pautados no mecanismo cognitivo da *comparação*, a que já nos referimos. Os caracteres são as únicas coisas observáveis, e é pela sua comparação que se chega aos táxons. Comparando indivíduos, suas estruturas que ocupam a mesma posição relativamente a outras estruturas do corpo são *homólogas*. A mão humana é, evolutivamente, homóloga à barbatana da baleia ou à asa do morcego; sob a luz da teoria da evolução, pretende-se, pois, que esses caracteres, hoje morfologicamente distintos, provenham de um mesmo ancestral. No entanto, a asa do morcego e a asa da borboleta são caracteres *análogos*, pois possuem a mesma função, mas não ocupam a mesma posição relativa (e portanto não indicam uma ancestralidade comum). Dados dois *estados* (diferentes formas) de um certo caráter homólogo, dizemos que temos um estado *plesiomórfico*, que antecedeu no tempo e originou, por alguma mutação, um segundo estado *apomórfico*. Dessa forma, cada caráter tem sua história de origem e diferenciação. Os caracteres, porém, não existem independentemente da espécie. Quanto à distributividade dos estados, um estado apomórfico é dito *sinapomórfico* quando compartilhado por dois ou mais táxons. Da mesma forma, um caráter é *simplesiomórfico* para o conjunto de todos os táxons que compartilham a condição plesiomórfica de um caráter. Diz-se que um estado é *autapomórfico* quando for apomórfico para uma única espécie, num determinado horizonte de tempo. Os conceitos de sinapomorfia e autapomorfia são tremendamente parecidos com os conceitos de *erros conjuntivos* e *erros disjuntivos* da ecdótica, respectivamente (PAPAVERO; LLORENTE-BOUSQUETS, 1999, p.94-112), criados por Maas (1927) a partir das idéias de Lachmann (1850).

### Aplicação do modelo de Hennig em dados lingüísticos<sup>3</sup>

Se utilizarmos a mesma nomenclatura para os estudos diacrônicos, veremos que um indivíduo pertencente à espécie *exemplo* remonta a um ancestral latino *exemplum*. A passagem  $l > r$ , justificável pela homologia posicional, tem um elemento plesiomórfico  $l$  e um outro apomórfico  $r$ . Outro indivíduo da mesma espécie *exemplo* existe na carta de D. João 3º a D. Pedro Mascarenhas (1539). No entanto, outra espécie aparentada,

<sup>3</sup> Um exemplo de reconstrução clássica, com a indicação de caracteres homólogos, se encontra no anexo encontrado ao final do artigo (STANG, 1966; ENTWISTLE; MORISON, 1949; e MATTHEWS, 1960).

*exemplo* com *l* plesiomórfico, aparece nas cartas do Padre Manuel da Nóbrega (década de 40 e 50 do século 16). Podemos dizer que *exemplo* e *exempro* são *eidoforontes*, por causa de outras características para além da sua forma (morfofossintaxe, semântica etc.). Na mesma carta de 1539, existe uma analogia da passagem *l > r* em *supricai*, que remontaria a um latim *supplicatis*. Pensando assim, seria curioso perceber que uma das regras fonéticas mais conhecidas (o rotacismo do *l*) é, na verdade, um fenômeno analógico (pois *exempro* e *supricai* são espécies distintas), motivado por sua condição homológica (segunda posição da sílaba com ataque complexo, iniciado por oclusiva). Abstraindo o conceito de língua, encontrá-lo-emos ainda em outros sistemas. Assim, Teyssier (1959) mostra que em Juan de Encina (1496) se encontram casos parecidos: *habrar*, *cramar*, *cremencia*, *prazer*, *obrigar* etc. Casos análogos de *l > r* também se encontram em romeno, mas não são homólogos (só ocorrem em situação intervocálica). Ex.: *soare* remonta ao latim *solem*. Os dois rotacismos, portanto, são incomparáveis e isso não é também um fenômeno desconhecido da zoologia: trata-se do fenômeno da *homoplasia convergente*.

## Homoplasias e derivas

*Homoplasia* é a presença da mesma apomorfia em diferentes táxons, oriundos de ancestrais diferentes, em tempos diferentes. Diz-se que a homoplasia é *convergente* se se deve a diferentes bases genéticas, que originaram, por alguma mutação, a mesma expressão fenotípica (tanto quanto se possa julgar), sempre em ancestrais diferentes e em diferentes tempos. Assim, os insetos da família *Mantidae* (vulgo louva-a-deus), que são heterometabólicos, têm uma aparência incrivelmente grande, no tocante ao par anterior de pernas, transformadas em garras raptorais, com os *Mantispidae*, insetos neurópteros holometabólicos.

Em língua, semelhanças de caracteres chegam nesse nível também: *bad* significa “ruim” seja em inglês, seja em persa, e *mati* significa “olho” em malaio e em grego moderno. Podemos rastrear sua história e percebermos tratar-se de uma coincidência. Igualmente, no nível dos sons, a incomum passagem de *o > i* se dá em ucraniano e no romanche da Surselva, sem que isso possa ser atribuído a algum substrato comum, gerando inclusive formas próximas como, respectivamente, *niev* e HIB para “novo” (BERNARDI, 1994; PUGH; PRESS, 1999).

Voltando ao rotacismo de *exempto*, vemos que ocorreu em inúmeras fases da história da língua, ou seja, não é fenômeno datado, que pudesse ser explicado por substrato: aparentemente surge naturalmente. Nada garante que alguém que hoje fale *exempto* tenha herdado essa forma ou simplesmente realize a líquida nessa posição sempre como não-lateral. Esse fenômeno é conhecido pela lingüística com o nome de *deriva*. Quando Sapir (1921) o propôs, fê-lo nos seguintes termos:

The desire to hold on to a pattern, the tendency to "correct" a disturbance by an elaborate chain of supplementary changes, often spread over centuries or even millennia—these psychic undercurrents of language are exceedingly difficult to understand in terms of individual psychology, though there can be no denial of their historical reality.

Embora Sapir tivesse razão com relação à existência das derivas, esse tom misterioso deu-lhes uma aura demasiadamente transcendente e quase mística. É verdade que o *\*d* indo-europeu tornou-se *\*t* em germânico e essa transformação *d > t* ocorreu outra vez na passagem do germânico para o antigo alto alemão e novamente em dialetos modernos (por exemplo, em alemão suíço). O alemão, de modo geral, tem uma tendência ao ensurdecimento das oclusivas sonoras, fenômeno comum a outras línguas germânicas (dinamarquês, islandês etc.), tendência que inexiste nas línguas românicas. Por outro lado a passagem *s > \*h > Ø* em posição de coda ocorreu no francês e, em outras épocas, no espanhol. As derivas seriam vistas, pela cladística, como *homoplasias paralelas*, ou seja, devidas à mesma base genética que, pela mesma mutação, originou a mesma expressão fenotípica, em diferentes ancestrais e diferentes tempos.

Resta explicar como isso ocorre. Vimos que, pelo mecanismo cognitivo da valoração, separamos o que é distintivo do que não é. No entanto, a realidade dos elementos não-distintivos não deixa de existir por causa de nossa intervenção cognitiva. A distinção entre *l* e *r* reside basicamente no esforço muscular que promove a lateralização. A distinção entre *b* e *p* está no vozeamento. A distinção entre um *s* e um *h* está na fricção. Lateralização, vozeamento, fricção e outros componentes articulatórios (nasalidade, aspiração, tensão, intensidade acentual etc.) são foneticamente graduais e não absolutos como os traços binários da fonologia fazem supor. Além disso, toda articulação é composta não só de elementos sistêmicos, mas

também de inúmeros componentes da norma (COSERIU, 1958), de modo que podemos lançar hipóteses sobre situações areais: o hábito articulatorio de menor lateralização do *l* existiria em determinadas regiões da Península Ibérica, uma maior tensão muscular existiria em uma extensa área da zona germânica, uma menor fricção das sibilantes ocuparia ampla zona românica ocidental. As apomorfias lingüísticas convivem com as plesiomorfias, de modo que a eclosão de uma mudança lingüística se deve não só a uma maior expansão areal da forma apomórfica, mas também a uma maior valoração de seu uso. Como as línguas oficiais privilegiam as plesiomorfias (com muitíssimas exceções), que conservam caracteres do ancestral documentado, formas apomórficas são vistas freqüentemente como “desvios” daquelas.

A abundância de *exemplo* nos textos anteriores às gramáticas do século 17 mostram que, nesse momento, as espécies apomórficas tinham o mesmo prestígio das espécies plesiomórficas e estabeleciam até mesmo uma tradição escrita. Seria um *non sequitur*, contudo, imaginar que sejam mais coloquiais: simplesmente a oposição binária falado/escrito não se impunha. O que havia eram práticas de escrita. A influência das gramáticas não se efetua com prescrições em nome de alguma pretensa “lógica”, como fariam mais tarde. Quando Diogo de Gouveia escrevia em 1538 a D. João 3º, a gramática de Fernão de Oliveira (1536) havia sido recolhida e a de João de Barros ainda não tinha vindo a lume (1540). Mesmo nas cartas de Manuel da Nóbrega, a influência dos conselhos lingüísticos de João de Barros é mínima. A prática escrita medieval ainda contava. Nas Atas da Câmara da Vila de São Paulo, grafias como *thrynta* (1572) apontam tenuemente para o movimento da ortografia etimológica, propalada por Nunes de Leão, embora ainda seja anterior à sua obra principal (1576).

Os caracteres da escrita só passam a uma interpretação sonora por meio de formulações de hipóteses e não pela reconstrução. Entra aqui o quarto pilar cognitivo a que nos referimos no início.

Se o rotacismo de *l* em ataques complexos no português brasileiro é facilmente comprovável como herdado do português europeu, uma vez que há continuidade na documentação, o mesmo fenômeno na posição de coda é mais difícil de ser comprovado, embora Vasconcelos (1928) a descreva em vários dialetos europeus e a encontremos não só no dialeto caipira, mas também em outras variantes não-portuguesas. Flutuação de formas como as que aparecem nas Atas da Câmara: *almotasell* (1562)/



*dallmotaser* (1564) parecem estar ligadas a flutuações do tipo *aluguel/aluguer*, não generalizadas. Pode ser que não estejamos aí diante de uma mesma tradição, mas manifestações separadas da mesma deriva (ou seja, outras homoplasias paralelas).

## Graus de incerteza de Jespersen (1922)

É com esse espírito cauteloso que devemos agir diante de formas isoladas que retrocedam a datação dos fenômenos diante de formas curiosas como *obitter* (Atas da Câmara, 1572: “pedio p<sup>a</sup> hordenar q~ se fizese hũa fïta p<sup>a</sup> obitter dinheiro”), que “comprovaria” a presença do *-i-* epentético, típico do português brasileiro *atual*. A certeza dessa afirmação só se dá por meio da absoluta certeza de transcrição paleográfica e da descoberta de outros exemplos dessa regra, o que daria à afirmação “a epêntese de *-i-* remonta ao último quartel do século 16” um **grau 3 de incerteza** na escala de Jespersen, uma vez que outros exemplos da mesma época ou anteriores não foram encontrados.

A partir dos fósseis muitas coisas se reconstroem em paleontologia; outras, apenas preenchem nossa necessidade de exaustividade de informação. A espetacular reconstrução de dinossauros se entredevorando que presenciamos em programas televisivos tem enorme fundamentação indutiva, alguma fundamentação hipotética e uma grande dose de fantasia.

Imaginar que se possa reconstruir línguas em grau 1 pode causar uma grande frustração, como afirma Meillet (1913, p.59): “*pour une étymologie sûre, les dictionnaires en offrent plus de dix qui sont douteuses et dont, en appliquant une méthode rigoureuse, on ne saurait faire la preuve*”.

Um documento de 1564 das Atas da Câmara comprova a existência de uma espécie *hareseo* (abonada como *arreceo* nos dicionários, cf. HOUAISS; VILLAR, 2001). Isso “comprovaria” que a forma protética de *receio* era usada na época. Tal fenômeno do *a-* protético, porém, dispõe de vários paralelos analógicos: *alevantar* (carta de Manuel da Nóbrega ao padre Simão Rodrigues, 1549), *alevantada* (carta do padre Pero Doménech, 1550) e em outros textos portugueses mais antigos. Uma afirmação como “as formas protéticas conviviam com as não-protéticas em meados do século 16” tem **grau 2 de incerteza** na escala de Jespersen. Muito embora *alembrear* seja uma forma citada por Amaral (1920), devemos trabalhar indutivamente

na coleta no nível dos indivíduos e das espécies antes de partir para hipóteses. Se não encontramos a forma *alembrrar* na mesma época, não é possível afirmar que *arreccio*, *alevntar* e *alembrrar* tenham todos o mesmo grau de antigüidade (sobretudo porque talvez todas sejam apomorfias independentes).

Outro caso interessante são as curiosas palavras *estãodo*, *sãotos*, *avãogelhos*, *evãogelhos* que se vêem nas Atas da Câmara de 1562 e nos anos seguintes, inclusive quando se muda o escrivão. Essas variantes não têm continuidade na atualidade, como ocorre em *bom/bão*, embora se conheça a transição de palavras como *-am > -ão*: *pam/pão*, *cam/cão*. Antes de afirmar que se trata de uma pronúncia extinta, devemos nos lembrar de variantes como *estavam/estávão*, tendo sobrevivido a segunda na língua culta embora a primeira tenha sido adotada pela grafia. Indaga-se, portanto, se *estãodo* não seria apenas uma prática escrita e não um reflexo de um fenômeno fonético. Fora das Atas, essas formas autapomórficas não ocorrem, o que não quer dizer que não tenham existido, pois, assim como os fósseis, os testemunhos escritos formam uma pequena parcela do que de fato existiu. Dessa forma, até que se encontrem outros exemplos fora das Atas, a afirmação "o som \*[ã] passou a \*[ãw] nas tônicas não-finais e pretônicas do português paulista da segunda metade do século 16" tem grau 4 de incerteza na escala de Jespersen.

Haveria algum caso de grau 1? Se sim, podemos atribuí-lo ao fenômeno do alçamento das pretônicas. Em muitas palavras, *e > i* e *o > u* em situação pretônica, em muitas fontes. O alçamento de *e > i*, erroneamente tratado como um fenômeno atual (sob o mal-aplicado rótulo de "harmonia vocálica"), é, sem dúvida, uma herança do português europeu. As formas *pedial/pidia* remontam ao latim *petēbat*. A espécie apomórfica (por assimilação) com [i] foi preferida pela fala, e a forma com *e*, pela escrita. A forma *pidia* já se encontra no século 13 e o alçamento não pode ser considerado um fenômeno recente. De um pretérito imperfeito ancestral \*[pi'dia] (e não de \*[pe'dia]) derivaram-se as espécies mais correntes da fala brasileira [pi'dia] e [pi'd<sub>3</sub>ia]. Do infinitivo ancestral \*[pi'dir] (testemunhado no português medieval *pidir*) formaram-se o português brasileiro [pi'di] ou [pi'dir] e o crioulo caboverdiano ['pidi] (QUINT, 2000). Uma afirmação como "muitas palavras com alçamento das pretônicas e > i formam a base das variantes atuais do português brasileiro, sendo as não-alçadas restritas à grafia" tem tantos exemplos da dialetologia portuguesa (VASCONCELOS,

1928) e em tantos textos antigos, que poderíamos seguramente atribuir-lhe **grau 1 de incerteza**, na escala de Jespersen.

Há, contudo, dois tipos, que muitas vezes se confundem. As espécies *minino/menino* vieram para o Brasil e, malgrado a dificuldade de estabelecer um étimo para esses eidoforontes, tudo nos faz crer que, ao contrário do que se julga, a primeira é plesiomórfica e a segunda, apomórfica (por dissimilação), ao contrário do exemplo anterior. A forma *minino* também já se encontra no século 13. No tocante à expressão, na fala prevalece atualmente a primeira, de norte a sul do país (não tem variação diastrática nem diatópica); já na escrita, prevalece a segunda. É apenas o *status* da escrita sobre a fala que nos faz julgar nos modelos transformacionais que a forma falada seja um desvio ou uma imperfeição da escrita (VIARO, 2005), daí falar de alçamento de  $e > i$  até mesmo em casos em que o termo não se justifica, como este (pois, a imaginar que a forma mais antiga não seja com \*[e] pretônico, há, na verdade, abaixamento da vogal  $i > e$ ).

Excluir totalmente a certeza de uma afirmação é mais fácil do que formular uma afirmação certa. Por exemplo, diante da bizarra grafia *quaza* que aparece nas Atas da Câmara de 1562, algum pesquisador ingênuo poderia reconstruir uma pronúncia \*[kwaza]. No entanto, vastos exemplos mostram que \*[kw] > [k] e nunca o inverso. Exemplos similares não faltam: *quamara* (1562), *quaregadas* (1563), *brãoquos* e *enquãopar* (1564). A julgarmos pela total inexistência de conseqüências atuais de uma pronúncia de tais oclusivas velares bilabializadas apomórficas em toda a área lusófona, quer em posição inicial quer em qualquer outra posição, estaríamos em uma situação bem mais difícil de comprovar do que a do caso *estãodo*, citado acima. Uma afirmação do tipo “a grafia *qua* equivale sempre a \*[kwa] independente da palavra” teria **grau 5 de incerteza**, na escala de Jespersen.

## Conclusões

Antes de avaliarmos todas as hipóteses que envolvem a reconstrução ou as interpretações de grafias, no tocante ao português brasileiro antigo, em especial ao paulista, devemos nos perguntar:

(a) o caráter em questão é ocorrência isolada a uma única palavra (autapomórfica) ou ocorre abundantemente?

(b) é possível provar que o caráter em questão existe ou existiu na palavra ou, analogicamente, em outra palavra, em alguma outra variante

falada no mundo lusófono?

(c) o caráter em questão existe ou existiu em alguma outra língua românica ou está presente arealmente? Se sim, trata-se de simpliesiomorfia ou de homoplasia?

A possibilidade ou impossibilidade de responder a essas questões ajudará no estabelecimento dos graus de incerteza envolvidos na reconstrução. Para isso, precisamos de dados, sobretudo os que rastreiem as datações mais antigas e as últimas ocorrências de uma forma extinta. Oliveira e Kewitz (2002) localizaram no *Correio Paulistano* de 1865 um texto que pode ser considerado para o caipira um equivalente ao *Juramento de Estrasburgo* para o francês. Deve haver outros, mais antigos. Os fósseis estão à disposição. Busquemos.

## **Lexical Reconstruction of the Old Portuguese from São Paulo/ Brazil**

■ **ABSTRACT:** *This paper deals with the problem of language reconstruction and especially with the reconstruction of the old spoken Portuguese from São Paulo/Brazil. In this task, Jespersen's proposal for uncertainty degrees in scientific statements about diachronic phenomena is taken into consideration. Some aspects of Hennig's phylogenetic systematics are mentioned. The possibility of its application in linguistic data is also considered.*

■ **KEYWORDS:** *Reconstruction. Portuguese language. Diachrony.*

## **Referências**

AMARAL, A. **O dialeto caipira**. São Paulo: Hucitec/Instituto Nacional do Livro/ Ministério da Educação e Cultura, 1981 [1ª ed 1920].

AMORIM, D.S. **Fundamentos de sistemática filogenética**. Ribeirão Preto: Holos, 2002.

BARROS, J.de. **Gramática da língua portuguesa**. Organizada por José P. Machado. Lisboa: Sociedade Astória, 1957 [1ª ed. 1540].

BERNARDI, R. et alii. **Handwörterbuch des rätoromanischen**. Zürich: Offizin, 1994. 3v.

**COSERIU, E. Sincronía, diacronía y historia: el problema del cambio lingüístico.** Madrid: Gredos, 1958.

**ENTWISTLE, W.J.; MORISON, W.A. Russian and the slavonic languages.** London: Faber & Faber, 1949.

**HAUDRY, J. L'indo-européen.** Paris: PUF, 1994.

**HENNIG, E.H.W. Grundzüge einer Theorie der phylogenetischen Systematik.** Berlin: Deutscher Zentralverlag, 1950.

**HOUAISS, A.; VILLAR, M.S. Dicionário Houaiss.** Rio de Janeiro: Objetiva, 2001.

**JESPERSEN, O. Language: its nature, development and origin.** London: George Allen & Unwin, 1954 [1ª ed 1922].

**LEHMANN, W.P. Theoretical bases of Indo-European linguistics.** London/New York: Routledge, 1993.

**LEONI, F.E. Genio da lingua portugueza ou causas racionais e philologicas de todas as formas e derivações da mesma lingua, comprovadas com innumeraveis exemplos extrahidos dos auctores latinos e vulgares.** Lisboa: Typographia do Panorama, 1858. 2v.

**MATTHEWS, W.K. Russian historical grammar.** London: University of London/The Athlone, 1960.

**MEILLET, A. Aperçu d'une histoire de la langue grecque.** Paris: Klincksieck, 1913.

**NEBRIJA, A.de. Gramática castellana.** Madrid: D. Silverio Aguirre, 1946 [1ª ed 1492]

**LEÃO, D.N.de. Ortografia e origem da língua portuguesa.** Introd. Maria Leonor C. Buescu. Lisboa: INCM, 1983 [1ª ed. 1576]

**OLIVEIRA, F.de. Gramática da linguagem portuguesa.** Lisboa: Academia das Ciências de Lisboa, 2000 [1ª ed. 1536]

**OLIVEIRA, M.; KEWITZ, V. A representação do caipira na imprensa paulista do século XIX. In: DUARTE, M.E.L.; CALLOU, D. (orgs.) Para a história do português brasileiro.** Rio de Janeiro: UFRJ/FAPERJ, 2002. v.4, pp.125-154.

PAPAVERO, N.; LLORENTE-BOUSQUETS, J. Un nuevo concepto en biología comparada: el 'eidofronte'. **Publicaciones especiales del Museo de Zoología de la UNAM**. México 5: p.21-29. 1992.

\_\_\_\_\_. **Principia taxonomica: una introducción a los fundamentos lógicos, filosóficos y metodológicos de las escuelas de taxonomía biológica**. México:UNAM, 1996. v.8.

PUGH, S.M.; PRESS, I. **Ukrainian: a comprehensive grammar**. London/ New York: Routledge, 1999.

QUINT, N. **Dictionnaire cap-verdien-français: créoles de Santiago et Maio**. Paris/ Montréal: L'Harmattan, 1999.

\_\_\_\_\_. **Le cap-verdien: origines et devenir d'une langue métisse**. Paris/ Montréal: L'Harmattan, 2000.

SAPIR, E. **Language: an introduction to study of speech**. New York: Harcourt, 1921.

STANG, C.S. **Vergleichende Grammatik der baltischen Sprachen**. Oslo/ Bergen/Tromsø: Universitetsforlaget, 1966.

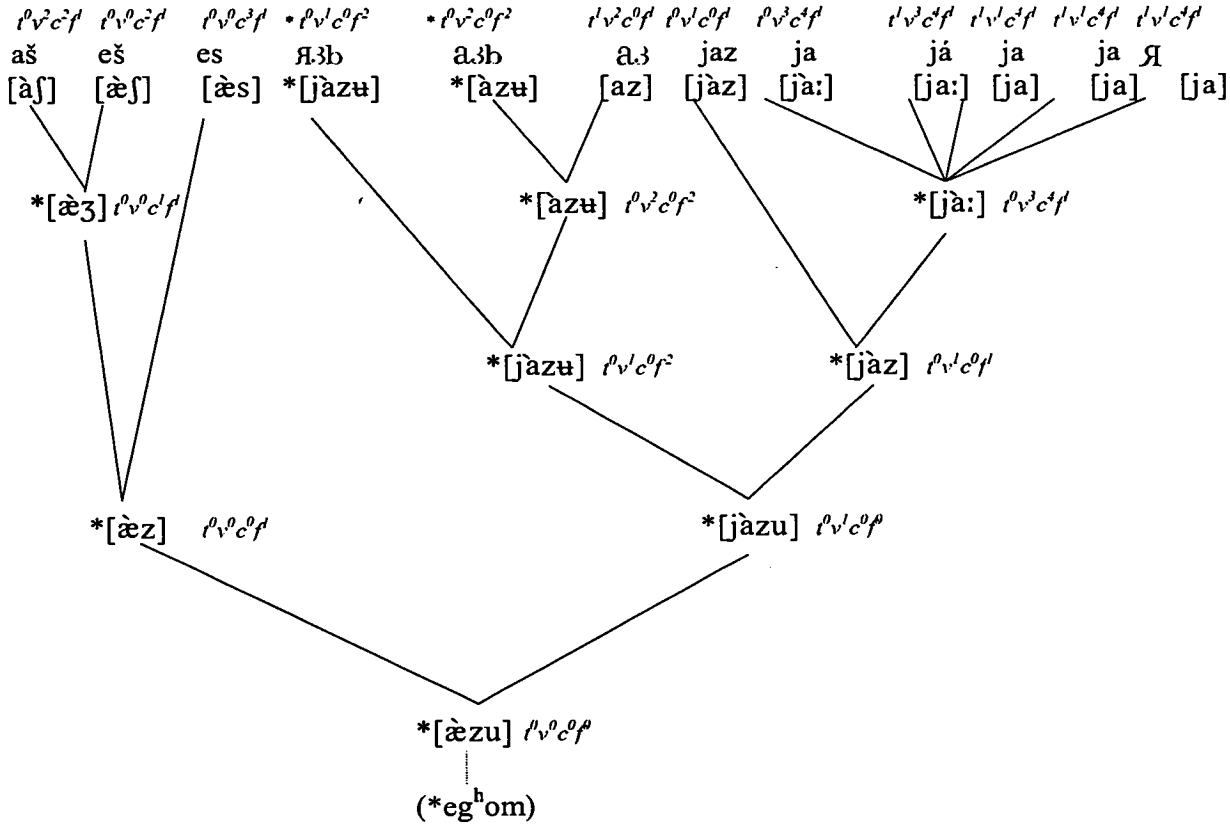
TEYSSIER, P. **La langue de Gil Vicente**. Paris: Klincksieck, 1959.

VASCONCELOS, J.L. **Opusculos**. Coimbra: Imprensa da Universidade, 1928.

VÄÄNÄNEN, V. **Introducción al latín vulgar**. Madrid: Gredos, 1988 [1ª ed. 1962].

VIARO, M.E. Semelhanças entre o português brasileiro e as variedades africanas e asiáticas. *In: SILVA, L.A.da (org). A língua que falamos*. São Paulo: Globo, 2005. pp.211-252.

**ANEXO: EXEMPLO CLÁSSICO DE RECONSTRUÇÃO COM BASE CARACTERES: ramo balto-eslavo do ide \*eg<sup>h</sup>om "eu"**



Caracteres: t = tom; v = primeira vogal ; c = consoante \*g<sup>h</sup> > \*z nas línguas *satem*; f = vogal final. O \*-m final não foi considerado.