



Sobre o problema da existência dos dados linguísticos

Mário Eduardo Viaro¹

RESUMO:

Há diversas formas de entender se a existência de um dado linguístico. Sua relevância não pode ser medida apenas pela materialidade, uma vez que a Linguística Histórica, a Filologia e a Etimologia também reconhecem a realidade de vários dados não materiais como, por exemplo, as reconstruções. Esses dados são utilizados também pela Linguística Geral e pela Lexicografia em particular. Há constantemente alguns problemas relativos à atualização filológica de seus grafemas, que é tradicionalmente feita de modo que se evite o anacronismo, o qual pode conferir falta de cientificidade ao dado linguístico e às análises pautadas nele. Com base em três obras de Jerónimo Cardoso (1508-1569), apresentam-se neste artigo várias situações envolvendo a questão da existência e a realidade de dados linguísticos.

PALAVRA-CHAVE:

Dado linguístico;
Existência;
Linguística Histórica;
Lexicografia;
Filologia

¹ Professor livre-docente da área de Filologia e Língua Portuguesa (FLP), do Departamento de Letras Clássicas e Vernáculas (DLCV) da Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) da Universidade de São Paulo (USP). Bolsista PQ-1D (CNPq). ORCID: <https://orcid.org/0000-0001-5714-1611>

1 Introdução

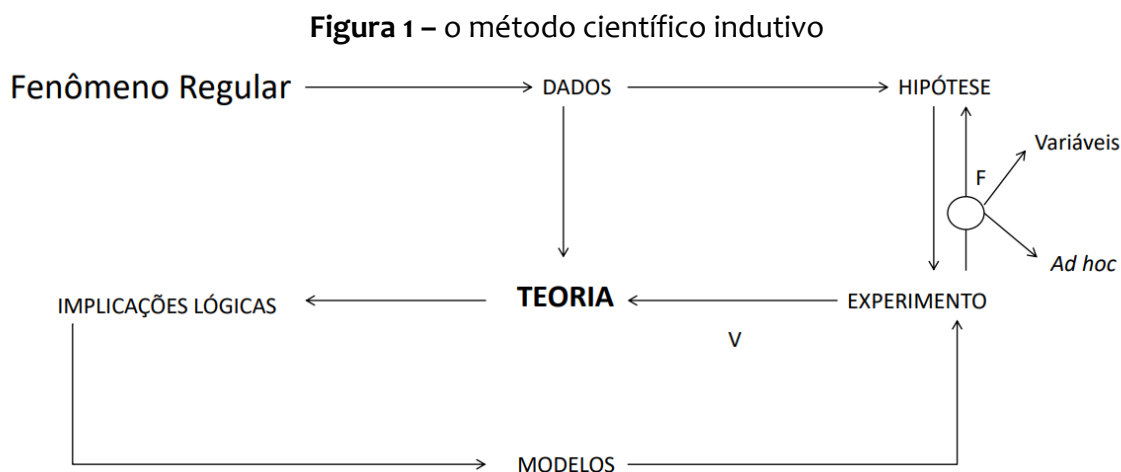
Em Teoria do Conhecimento, **dados** são fatos regulares e verificáveis, que podem ser controlados. Um espécime coletado por um biólogo é um dado, um fenômeno astronômico detectado por aparelhos há milhares de anos-luz também é, assim como é uma unidade mental da linguagem, como um morfema. Alguns dados linguísticos são reprodutíveis, como a pronúncia de um determinado fonema numa língua viva moderna: tanto o fonema quanto a pronúncia são dados, de diferente espécie. Outros dados não são reprodutíveis, como a realização de um fonema específico ou numa palavra latina, tal como soava durante o tempo de Júlio César. A situação de impossibilidade de percepção direta não impede o dado não reprodutível de ter tratamento científico: um biólogo estudará, com metodologia distinta, tanto uma espécie de réptil atual, quanto uma que viveu na era mesozoica e é assim que se estabelece o diálogo interdisciplinar entre Biologia e Paleontologia.

Imprescindível é que haja interesse do cientista nas informações que dado carrega, mas esse interesse só pode ser entendido como **científico** se o dado for verificável e descritível, independentemente da sua reprodução em laboratório. Dados são elementos dos quais nenhuma teoria prescinde, pois são a fonte de todas as conjecturas, as quais se submetem a um tipo de experimentação. Na verificação da sua verdade, inicia-se um **método**. A verdade contida no dado é complexa. Segundo Hegenberg (1969, p. 47), no conceito ocidental de “verdade” existente nas explicações científicas, entrelaçam-se significados tradicionais muito distintos: a de um relato que descreve os fatos sem omissão (latim *veritas*), uma relação de compatibilidade entre a descrição e as coisas (grego *alétheia* - ἀλήθεια), assim como uma revelação de um conhecimento latente (hebraico *emunah* - אמונה).

2 O modelo indutivo em Linguística

Um **modelo** é o resultado esperado de um método científico. Num modelo indutivo, os dados são a base sobre a qual qualquer hipótese se sustenta, cuja comprovação por meio de testes é o que alimenta o material, a partir do qual nascem, mediante implicações lógicas, as afirmações teóricas. Um modelo, dessa forma, não advém diretamente dos dados ou de uma hipótese, mas das **implicações lógicas** de afirmações que compõem uma **teoria**, alimentada por hipóteses corroboradas. Um modelo submete-se também à verificação experimental, como qualquer hipótese.

Não raro, apesar dessa circularidade, nesse momento de verificação, o modelo pode mostrar-se falso e isso só é explicável pelo fato de as teorias (de qualquer área do conhecimento), mesmo no percurso indutivo, não serem alimentadas exclusivamente de dados testados, mas também de dados brutos, como mostra a figura 1, que entram em contradição com afirmações da teoria e com suas decorrentes próprias implicações lógicas.



Fonte: própria

Quando a verificação de uma hipótese ou de um modelo se revela falsa, duas situações são possíveis para a **manutenção do modelo** (que é menos resistente do que o conjunto de afirmações conhecido como teoria):

1. A descoberta de **variáveis** que mantêm o modelo funcionando;
2. A situação menos nobre da introdução de explicações *ad hoc*, que facilmente são detectáveis pela crítica e fragilizam a defesa do modelo.

Um exemplo simples basta para entender esse percurso: à luz de uma teoria linguística conhecida como Estruturalismo, tanto fones quanto fonemas são dados de diferente **essência**:

- 1) **Fones** são o resultado de uma produção mecânica conhecida como **articulação**, que subentende um emissor humano e um aparelho fonador previamente definidos. Um fone, independentemente de sua análise acústica ou perceptiva, é, portanto, um fato controlável, verificável, descritível e, na maioria das vezes, reproduzível.

- 2) Um **fonema** é um dado completamente distinto: trata-se de uma abstração mental, de modo que não é diretamente verificável pelos órgãos do sentido (como o fone o é, pela audição), nem mesmo por indícios de escaneamentos cerebrais do tipo PET, uma vez que faz parte da mente humana e se confunde com as áreas específicas da linguagem nos cérebros humanos. O termo **fonema**, impôs-se aos poucos, ainda que com uma polissemia indesejável para um termo teórico (HEGENBERG, 1974, p. 35-68).

A verificação da existência e do comportamento de fonemas depende de um método e de modelos, uma vez que se trata de um elemento basilar do eixo sintagmático, sobre o qual se constroem outros elementos formais da língua, como sílabas, morfemas, itens lexicais e outras unidades maiores. Para responder a uma pergunta aparentemente simples, do tipo “quantos fonemas tem um determinado sistema linguístico?”, recorreu-se ao chamado **método da comutação**. Assim, *grosso modo*, valendo-se de dois itens lexicais semelhantes, que se distinguem por um só fone, ambos pertencentes ao mesmo sistema linguístico em questão, teremos: (1) **Dois fonemas** quando, da sua comparação, verificamos significados distintos para cada item lexical; (2) Quando isso não ocorre, entende-se que há **apenas um fonema**, que abstrai a diferença desses dois fones distintos

Valendo-nos de dados de língua portuguesa, diremos que /z/ e /s/ são fonemas porque “casa” [ˈkaza] e “caça” [ˈkasa] são itens lexicais distintos, mas [ˈkasa] e [ˈkase] remetem ambos à mesma palavra do português, *i.e.* “caça”, de modo que o mesmo fonema, chamado /a/ por definição, corresponde a dois fones distintos: [a] e [e]. Explica-se que isso só é possível porque, durante a aquisição de linguagem, o fonema foi adquirido mediante exercício da **abstração** compatível com a de outros falantes. Escolhendo outro sistema linguístico, esse resultado não é verdadeiro: em romeno, [ˈkasa] e [ˈkase] correspondem respectivamente a *casa* “a casa” e *casă* “casa”, portanto, não se trata, a rigor, de itens lexicais idênticos, de modo que podemos dizer que em romeno temos dois fonemas - , por definição /a/ e /e/ - e não um só, como em português.

3 O modelo dedutivo em Linguística

A diferença entre a concretude do dado fonético e a abstração do dado fonológico incita questionamentos sobre a existência e a realidade dos dados linguísticos:

- Por **existência** de um dado linguístico entende-se não só a sua **materialidade**, mas também a sua **funcionalidade**, a qual, por sua vez, é responsável pelo

entrelaçamento do seu significado cultural com uma referência mental, sua **essência** (RICKEN, 2002).

- Por **realidade** de um dado linguístico entende-se a possibilidade de **reconstrução** dessa essência.

Essa distinção entre existência e realidade se fundamenta no fato de que algo não é necessariamente real apenas por causa dos sentidos, como preconiza uma filosofia baseada no Realismo ingênuo. Obviamente, para essa definição de realidade não é preciso aprofundar-se em questões filosóficas sobre a possibilidade de uma realidade não-mental (BRANQUINHO; MURCHO; GOMES, 2006). Tampouco as raízes de nossa argumentação têm a ver com questões do Fenomenalismo. Por **real** entende-se algo que seja passível de reconstrução referencial e que reflita o que não seja exclusivamente subjetivo (BUNGE, 1999). Assim sendo, **apenas dados reais**, passíveis de reconstrução, com ou sem existência concreta, podem ser objeto da ciência linguística: o que não se pode reconstruir, por meio de modelos, ou é **irreconhecível** (se tiver existência material, por exemplo, uma palavra completamente manchada e ilegível em um manuscrito) ou é **incognoscível** (se, além de irreconstrutível, também não tiver existência material).

A realidade do fone é possível de se reconstruir não só diretamente, por meio da audição, mas também indiretamente, por meio de grafemas (tradicionais ou IPA) ou espectrógrafos; a realidade do fonema se reconstrói não só pela sua eficiência em gerar novas palavras, derivada de sua funcionalidade na linguagem, mas também pela verificação dos modelos, tais como o acima descrito, cujo método se representa na figura 1.

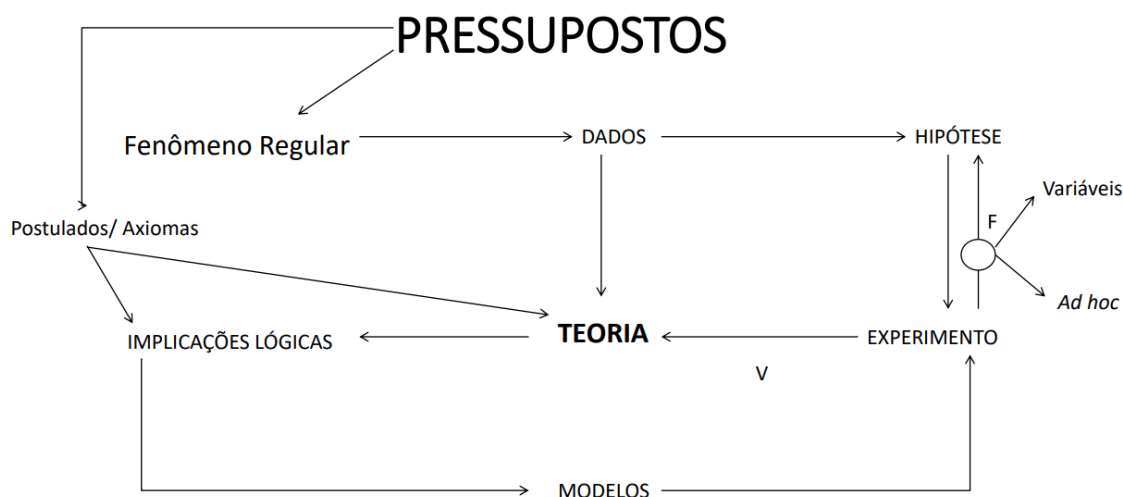
Fora do percurso do método indutivo, é possível que não só dados advindos de fenômenos regulares alimentem uma teoria, mas também **postulados** e **axiomas**, decorrentes da mesma **lógica** em que se fundam as implicações da teoria no método indutivo. Quando isso ocorre, temos o método dedutivo, empregado também em modelos científicos. Ambos os métodos, contudo, se sustentam inevitavelmente, do ponto de vista epistemológico, em **pressupostos** de um paradigma histórico e cultural, que permanece subentendido durante certo tempo (KUHN, 1996).

Chomsky & Halle (1968, p. 177) apresentam em suas descrições traços obtidos por indução e por dedução. Por exemplo, o traço “coronal”, apesar de referir-se a articulações concretas, é um postulado, de modo que pode-se afirmar que uma das proximidades entre [k^w] e [p] seja o fato de terem o traço [- coronal]. Esse traço é útil, por exemplo, para entendermos alguns dados provenientes dos estudos diacrônicos, como a passagem do latim *aquam* para o romeno *apă*, ambos itens lexicais com o mesmo significado “água”. A

coronalidade *per se* não é um fato verificável, tal como a vibração das cordas vocais, mas é um dado, postulado dedutivamente. Sua ausência também é um dado.

A funcionalidade demonstrada em outros dados revela a existência da coronalidade, ainda que sua essência seja apreendida dedutivamente. O reconhecimento dessa funcionalidade, entretanto, é - diferentemente da funcionalidade do modelo indutivo - garantida apenas *a posteriori*, ou seja, após a aplicação do modelo.

Figura 2 – o método científico dedutivo



Fonte: própria

No caso da Fonologia, acima apresentado, pressupostos de fenômenos como “língua” conduzem necessariamente a modelos indutivos, que dialogam com dados pelos quais se interessa a teoria estruturalista. O modelo resultante produz uma conceituação com pressupostos platônicos, sobre os quais se sustenta o conceito da “língua mental”. Uma mente seria algo necessariamente individual. No entanto, a Linguística pressupõe que uma língua seja um fenômeno social (a *langue* saussuriana). Isso, porém, não a identifica com outro conceito, o de “sistema linguístico”, que, para Saussure (1967 [1916], p.140) , como diz o **Cours** na sua primeira parte, cap. 3 § 9 só ocorreria em sincronia:

La linguistique synchronique s’occupera des rapports logiques et psychologiques reliant des termes coexistants et formant système, tels qu’ils sont aperçus par la même conscience collective.

La linguistique diachronique étudiera au contraire les rapports reliant des termes successifs non aperçus par une même conscience collective, et qui se substituent les uns aux autres sans former système entre eux.

A confusão entre língua e sistema linguístico, na área dos pressupostos, aparentemente tem suas raízes na gramática normativa e foi herdada pela linguística oitocentista e pela Linguística moderna, seja por qual linha teórica sustente o credo do linguista, com raríssimas exceções, tais como a postura schuchardtiana, a polêmica sapiriana das derivas, ou as isoglossas da Geolinguística e a Sociolinguística, que defendem a existência de **diassistemas**, que nada mais são que sistemas, integrados política- e historicamente num conjunto conhecido como “língua”.

4 A existência de dados reais

Resumidamente, a realidade de um dado é banal quando a verificação de sua existência é imediata. Também duvidar da realidade dos fonemas, pelo simples motivo de que não sejam inexistentes pela observação direta, seria algo ingênuo. Defender a realidade da coronalidade ou uma área física correspondente à produção de um dado [+coronal] exigirá, contudo, não só um método científico, mas um esforço argumentativo muito maior que defenda a dedução: numa perspectiva kuhniana, a permanência de um paradigma de termos científicos requer não somente a necessidade da convicção da parte do linguista proponente e de seus seguidores, mas também das instituições que sustentem esses dados teóricos e da própria teoria. É nessa zona alheia ao método científico que nascem as discussões que introduzem refutações, o surgimento de linhas teóricas alternativas e confissões de incredulidade, mas nada disso tem, de fato, aporte epistemológico consistente e legítimo, pois transita entre os pressupostos de um momento histórico.

Inversamente, um dado existente pode não ter realidade para a Ciência, como ocorre no caso das **falsificações**, que, uma vez descobertas, deixam de participar do diálogo e das relações causais que envolvem as explicações científicas. Como visto, a existência não tem a ver com a materialidade, uma vez que fonemas existem e não têm elementos que materialmente o sustentem. Além de real, um dado científico pode ser único materialmente, embora revele um fato que, um dia, **foi regular**. As ciências têm tratamentos distintos para dados únicos. Espécies novas devem ser descritas em Biologia e os exemplares em que são baseadas, mesmo que sejam únicos, devem estar à disposição de todos os pesquisadores em um museu. É por isso que sabemos que um dodô (*Raphus cucullatus*), uma moa gigante (*Dinornis robustus*) ou o pombo azul da Ilha Maurício (*Alectroenas nitidissimus*) existiram regularmente. Diferentemente, a fênix, os dragões de diversas narrativas épicas ou o unicórnio não são animais conservados em museus ou coleções particulares. Um nome científico, que siga as normas do ICZN não lhes cabe e não

são levados em consideração na Teoria da Evolução. Animais fictícios são dados para outros estudos científicos, como a História do Folclore, uma vez que dados fictícios são cognoscíveis e passíveis de reconstrução, por exemplo, em ilustrações. Um animal não fictício, desenhado, fotografado ou filmado cujo único exemplar foi destruído no incêndio de um museu onde ficou durante décadas ou séculos também pode ter *status* de existência para a Biologia, muito embora tentativas de reorganização de clados, nesses casos, sejam mais complexas, dada a situação de atual inexistência do holótipo, e isso vale mais do que uma narrativa espetacular de um animal presenciado por alguém, por mais idôneo e respeitável que seja o declarante.

Dados arqueológicos e paleontológicos têm tratamento muito semelhante aos dados biológicos, entretanto, dados linguísticos, contudo, não são sistematicamente catalogados: uma lista de palavras da língua oti basta para comprovar a sua existência no oeste paulista. A maior parte dos dados linguísticos não foi gravada, filmada, arquivada, classificada.

Muito da existência de dados linguísticos antigos se deve à escrita, que, como sabemos, mantém distância da sua realização fonética. Dito de outro modo, é preciso **reconstruir** o elo que liga escrita à fala para que possamos trabalhar com dados de sincronias pretéritas. Isso já é sabido desde meados do século XIX, quando se trata de dados de sistemas linguísticos completamente reconstruídos, como o indo-europeu, mas há extrema banalização quanto à acurácia desses dados quando se trata de dados linguísticos não obtidos por meio de um falante real.

5 O tratamento de dados linguísticos de sincronias pretéritas

Há diferenças qualitativas entre as línguas mortas: o hitita é uma língua cuja expressão se restringiu basicamente à época em que foi empregada por seus falantes. O latim não se enquadra na mesma tipologia. Exceptuando testemunhos como o de Montaigne, ao final do capítulo XXVI do primeiro volume de seus *Essais*, composto por volta de 1572-1573, que afirma ter aprendido o latim antes do francês, é possível dizer que há duas fases na língua latina:

- uma *lingua latina original* adquirida como língua materna, com transmissão de uma geração mais antiga a outra subsequente. No que já em meados do século IV, devido à fragmentação do Império Romano, não se reconhece o

mesmo sistema de séculos anteriores em reconstruções da língua falada usada (MAURER JR, 1951, 1959, 1962; SILVA NETO, 1956, 1977; VÄÄNÄNEN, 1988; BUCHI, SCHWEICKARD, 2014)

- Uma *língua latina exclusivamente escrita*, adquirida como língua estrangeira, a qual sobreviveu muito tempo depois do século IV. Baseada num tipo de variante escrita culta, mais ou menos semelhante ao latim clássico, serviu de língua veicular. Era preponderantemente escrita e, às vezes, falada por pessoas cultas, durante toda a Idade Média e séculos subsequentes. Sobrevive, de certo modo, até hoje.

Obviamente, dados do latim anterior ao século IV teriam de ter necessariamente tratamento distinto dos do “latim” posterior ao século IV. Nessa sincronia posterior, verifica-se com maior frequência um fenômeno comprovável apenas em línguas com tradição escrita: a **ressurreição** de dados linguísticos, mediante a leitura de textos antigos. Se morfológica- e sintaticamente, os dois latins são sistemas praticamente idênticos, uma diferença significativa se encontra no plano lexical, quer na presença de elementos neológicos, quer na diferença de frequência de uso de itens lexicais. Atribuir características fonológicas do latim clássico ao latim medieval ou ao latim moderno é, no mínimo, um erro grosseiro de análise e tem o mesmo valor de uma falsificação para a Filologia, mas confirma a existência de uma recorrente confusão entre o plano idealizado política- e ideologicamente do que se conhece como “língua latina” e o sistema linguístico obtido por meio de dados históricos, sobre o qual deveria interessar-se o linguista quando discorre sobre o “sistema linguístico latino” numa certa sincronia.

É complexa, portanto, a natureza dos dados linguísticos, que advêm seja da **língua** (entendida sociopoliticamente e, não raro, anacronicamente, isto é, sob uma perspectiva que não é nem diacrônica nem sincrônica), seja de um de seus **sistemas linguísticos** específicos (que, segundo a perspectiva saussuriana, seriam necessariamente sincrônicos, pois visam à comunicação) dos quais se compõe uma língua (VIARO, 2012). Não bastasse essa complexidade, falta ao dado linguístico extraído de sincronias pretéritas o elemento que ratifique a sua existência ou que, ao menos, gradue a verossimilhança de seus dados.

Parte dessa dificuldade se dá por uma certa incúria na simbologia que deve acompanhar um dado linguístico: a mais flagrante delas se revela na falta de marcação do dado por meio de um símbolo que diferenciem os dados quanto à sua existência, como o asterisco schleicheriano. Desde meados do século XIX, a Linguística sentiu a necessidade de apontar um dado reconstruído por meio de um asterisco, o qual reflete que o dado é real

mas não verificável por causa dos acidentes históricos, portanto inexistente. A realidade do dado se deve às reconstruções do sistema linguístico em sincronias pretéritas. O uso do asterisco usado para marcar reconstruções, portanto, diz-nos que o dado oferecido pelo autor existiu materialmente, embora não seja testemunhável. O asterisco schleicheriano é empregável, portanto, em duas situações: **(a)** reflete dados de uma língua ágrafa de uma sincronia pretérita; **(b)** reflete dados não encontrados nos manuscritos sobreviventes. Assim sendo, o asterisco schleicheriano nos apresenta não só um dado real, mas um dado que o autor proponente defende como **possível** (VIARO; BIZZOCCHI, 2016).

Nos estudos sincrônicos do século XX, porém, que não trabalhavam com dados diacrônicos ou com sincronias pretéritas, o significado do asterisco passou a ser interpretado de outro modo: trata-se de um dado irreal, portanto, **impossível**. Tal asterisco foi divulgado com o gerativismo, mas aparentemente já circulava nos escritos do estruturalismo norte-americano, por exemplo, no cap. 10.4 do *Language* de Bloomfield (1966 [1933¹], p. 163): “on the other hand, **Bill John hit* is not an English form (...); similarly, *play-ing* is a form, but **ing-play* is not”. Veja, do ponto de vista historiográfico, a apresentação do mesmo problema em Graff (2002).

A convivência de modelos sincrônicos e diacrônicos nos textos linguísticos do final do século XX e do século XX criou o paradoxo de se usar o asterisco com os dois significados, às vezes na mesma página ou no mesmo parágrafo. Essa prática equivale à adoção em Aritmética do símbolo “+” não só para a adição mas também para a subtração, de modo, que, num mesmo cálculo, algo como $2+3+4=1$ fosse entendido como verdadeiro, pois, mediante alguma exegese, dever-se-ia entender que o autor quis dizer que o primeiro sinal deve ser entendido como adição ($2+3=5$) e o segundo como subtração ($5-4=1$).

A incúria também surge em transcrições fonéticas de línguas mortas. Quando se diz que a pronúncia de uma palavra como *homō* “ser humano” em latim clássico era [‘homo:] sem asterisco, não há clareza de que essa pronúncia, na verdade, se trata de uma conclusão obtida por meio da **reconstrução** do sistema do latim clássico. Esse cuidado existe, por exemplo, no DÉRom (BUCHI; SCHWEICKARD, 2014) mas a falta de cuidado é ubíqua e também tem longa tradição, como vê no *Language*, no cap. 11.3 (BLOOMFIELD, 1966 [1933¹], p. 175-176).

Rigorosamente, afirmações sobre a realização fonética de qualquer dado linguístico anterior ao período de invenções que permitam que tais dados possam ser ouvidos sem a presença do sujeito enunciador (por exemplo, gravações magnetofônicas, rádio, televisão, cinema etc.) necessitam de reconstrução. A pronúncia de **homō** deve ser indicada cientificamente apenas por *[‘ho:mo:]. O mesmo

é válido para a pronúncia de dados linguísticos provenientes de qualquer língua anterior à segunda metade do século XIX. Longe de ser um cuidado exagerado ou uma atitude radical, o tratamento adequado dos dados históricos constrói as bases da Linguística Diacrônica, não a convicção que os autores possam ter sobre eles. Qualquer afirmação sobre a fonologia de uma língua cuja sincronia estudada se encontra em um estágio pretérito da história necessita do mesmo rigor. Dizer que tal cuidado não é importante, por ser “óbvio” e por requerer “bom senso” equivale a dizer que os sinais da Aritmética poderiam ser ambíguos, pois a interpretação de $2+3+4=1$ conduziria a uma probabilidade de 25% de verdade. O problema se transfere assim, ao leitor, de quem se exige sagacidade, mas essa postura entra em contradição com a clareza da linguagem científica postulada desde Francis Bacon. Além disso, para a Ciência, sempre foi preferível esclarecer os dados a torná-los mais obscuros.

6 Relação entre dados existentes e corpora

Partindo não de uma entidade tão ampla quanto uma língua específica ou um determinado sistema linguístico, mas para um tipo de **recorte**, conhecido em Linguística como *corpus*, a existência do dado também se revela algo bastante complexo. Suponhamos que queiramos analisar dados de três testemunhos de um mesmo autor numa sincronia pretérita, a saber, os vocabulários do quinhentista português Jerónimo Cardoso (1562-1563, 1570), a saber: **(a)** as duas primeiras edições do vocabulário português-latim desse autor, respectivamente de 1562-1563 e a edição póstuma de 1570 (chamada 1570¹); **(b)** a primeira edição do vocabulário latim-português, chamada 1570². Confrontando esses dados com a informação presente em outras obras que tratem da História da Língua Portuguesa (por exemplo, TEYSSIER, 1990, p. 49-52; MAIA, 1986, p. 438-468), não é possível ignorar a hipótese bastante robusta de que havia dois fonemas distintos em sincronias pretéritas da língua portuguesa antiga: **(1)** um fonema */s/, grafado como <c> ou <ç>, de realização pré-dorsoalveolar; **(2)** um segundo fonema */ʃ/, grafado <s> ou <ss>, de realização apicoalveolar.

Esses fonemas teriam permanecido em sincronias mais recentes: em alguns lugares há ainda as duas realizações fonologicamente distintas, em outros, apenas uma delas, como resultado de uma neutralização desses fonemas. Os lugares onde houve a neutralização correspondem ao “português padrão” e a neutralização foi difundida por todo Império Português, de modo que não há notícia da existência de dois fonemas hoje a não ser de forma residual no português lusitano setentrional, como é testemunhado ainda no início do

século XX por Vasconcelos (1928, 1985). Tais fonemas do português antigo seriam, ainda, comparativamente equivalentes aos fonemas do castelhano e do galego modernos, a saber, /θ/, grafado <c> e <z>, de realização apicodental, e /s/, grafado <s>, de realização apicoalveolar. O castelhano meridional aparentemente tendeu, como o português centro-meridional, a neutralizar os dois fonemas, de modo que só herdou um fonema no espanhol americano. Desse modo, a dúvida que se impõe face a essas informações, quando se elencam os dados obtidos nas obras de Cardoso na confecção de um suposto dicionário de ocorrências se refere a problemas de **modernização da grafia**. Interessa à Lexicologia, como à Filologia, em edições críticas e semidiplomáticas, que a modernização deva ser equilibrada e não ser anacrônicas (SPINA, 1994, p. 49-62; SPAGGIARI; PERUGI, 2004, pág. 60-168; CAMBRAIA, 2005, p. 109-131). A modernização lexicográfica não se preocupa com vários problemas da Filologia, como pontuação, mas diversos outros problemas são similares. Não há certeza se a situação linguística que Cardoso quer refletir é a de dois fonemas */s/ e */ʃ/ ou apenas um só */s/. Alguns dados aparentemente apresentam oscilação:

çafira, pedra. Saphirus, i. [1562-1563¹, 38r, A, 11]; **Çafira**, pedra. Saphirus, i. [1570², 28v, A, 4]; Sapphirus, i. A **çafira**, pedra preciosa. [1570¹, 218v, B, 18]; Lapis saphir9. A **saphira** pedra preciosa [1570¹, 117v, A, 39]; Sapphiratus, a, ã. Cousa chea de **safiras**. [1570¹, 218v, B, 20].

A informação presente entre colchetes tem quatro itens: (1) edição com a qual se trabalhou; (2) número da a folha do impresso com marcação de (r) frente ou (v) verso; (3) coluna, que pode ser A ou B; (4) e a linha da coluna. Uma análise propensa a imaginar, pelos dados acima, que <ç> e <s> refletem um único fonema pretérito */s/ e apenas estão em variação livre chegará, possivelmente, às seguintes conclusões:

- 1) Como a escrita não segue qualquer normatividade, isso é indício de que há neutralização da pronúncia, como na norma culta atual;
- 2) Conseqüentemente, é possível dispor essas ocorrências como exemplos do **mesmo verbete**, pois se trataria de uma única variante da unidade lexical;
- 3) Esse verbete será escrito com <s>, ou seja, **safira**. Essa decisão se pauta tanto na prática ortográfica atual quanto no étimo (latim *sapphirus*). O dado <saphira> parece apontar para isso.

Uma característica importante da solução da escrita única é a ausência de questionamentos na construção do único verbete **safira**, enviesado pelo pressuposto de que “a pronúncia da época em questão já era igual à de hoje em dia”. No entanto, isso somente não seria anacrônico se refletisse uma pronúncia como *[ʂa'fira], com o som apicoalveolar. Se admitirmos que apenas havia uma pronúncia pré-dorsoalveolar, *[sa'fira] a grafia única **çafira** também seria justificável. Trata-se, portanto, de duas hipóteses distintas, mascaradas sob uma só. Tais análises que revelam um fonema pretérito único não são isentas de questionamentos, tais como:

- 1) Qual seria a “pronúncia de hoje em dia”, uma vez que há diversas pronúncias no português atual? [sa'fira], [sa'fire], [se'fire], [ʂe'fire]?
- 2) Cardoso, como estava num momento de transição da padronização da escrita, não teria tido, em sua prática e convívio, algum contato com a distinção */s/: */ʂ/, para a qual teria alguma sensibilidade perceptiva, refletida na sua oscilação da representação gráfica entre <ç> e <s>?
- 3) A escrita com <s> da palavra portuguesa seria apenas uma influência etimológica da palavra latina *sapphirus* nas decisões de representação fonética da grafia de Cardoso ou a interferência de Stockhamer (citado na capa), do editor ou do tipógrafo na composição de sua obra póstuma de 1570?

Tais questionamentos justificariam a presença de **dois** verbetes lexicograficamente relevantes: **çafira** e **safira**, nascidos da dúvida. Duas novas situações, ao menos, se anteveem, com a presença da dúvida sobre os dados:

- 1) a análise final de todos os dados da expressão reconstruída de Cardoso poderá apresentar indícios de que **não havia** mais distinção fonológica, em contexto algum, entre */s/: */ʂ/, embora houvesse, em sua época, ambas as pronúncias *[s] e*[ʂ] de maneira um tanto caótica. Os dados <çafira> e <safiras> refletiriam também uma utilização fonética um tanto caótica dos grafemas <ç> e <s>, de modo que a presença de dois verbetes **safira** e **çafira** seria defensável. Caberia, necessariamente, ao dicionarista fazer a remissão de um verbete para outro e, nesse caso, é comum a solução de escolher um deles arbitrariamente (por exemplo a forma atual) para conter os dados, enquanto, por meio do outro, se faria simplesmente a remissão. Para diminuir

a arbitrariedade, pode pesar o fato também, na escolha do **verbeta principal**, do fato de o étimo iniciar-se com <s> e isso ser um critério explícito ou pressuposto.

- 2) a análise final de todos os dados da expressão reconstruída de Cardoso poderá apresentar indícios de que **havia** distinção fonológica entre */s:*/s/ em **alguns contextos fônicos** e consequente neutralização em outros, por exemplo, no início de palavras ou antes de um */a/. Nesse caso, a oscilação entre <çafira> e <safiras> também pode ser entendida como aleatória. A única diferença na defesa de dois verbetes aqui seria a conclusão da presença da oposição fonológica em alguns contextos, no entanto os mesmos problemas lexicográficos de remissão se apresentam.

Obviamente, soluções mistas também podem existir, como, por exemplo, seguir cegamente a ortografia atual em casos de dúvida, mas essa solução só parece ser válida quando, por exemplo, há explicitação de listas de casos excepcionais num prefácio. A situação da existência de dois dados a partir de diferenças ortográficas, <çafira> e <safiras> para o mesmo item lexical, encontra paralelos em outras palavras, como: <çafra> e <safra>, <çagacidade> e <sagacidade>, <çapinho> e <sapinhos>, <çaramago> e <saramago>. No entanto, o paralelo não é encontrado sempre. Por exemplo, **çafar** é o único verbeta possível nos seguintes dados:

çafar. Detero, is. [1562-1563¹, 38r, A, 9]; **Çafar**. Detero, is. [1570², 28v, A, 2]; Extero, is, triui, tritum. **Çafar** [1570¹, 79r, B, 18]; Penulam æstate deteris. **Çafas** ho be/dem no estio. Dirseha dos que gastã/ as cousas em tempo desnecessario,co/mo sam os que gastam ho dinheyro/ sem necessidade, & despois não no/ têm quando ho hão mister. [1570¹, 177v, B, 38 – 178r, A, 2]; **çafada** cousa. Detritus, a, um. [1562-1563¹, 38r, A, 10]; **Çafada** cousa. Detritus, a, um. [1570², 28v, A, 3]; Attritus, a, um. Cousa **çafada**. [1570¹, 34v, B, 15].

Há casos intermediários em que uma das flexões não aparece no *corpus*, por exemplo, para o seguinte dado: “**çafões**. Calige pelliceæ. [1562-1563¹, 38r, A, 23]; **Çafões**. Calige pelliceæ. [1570², 28v, A, 18].”. Perante esses dados, não se dirá que ***safão** é uma reconstrução válida, a não ser que a hipótese do único fonema */s/ seja evidente, mas já vimos as consequências dessa decisão. Por outro lado ***çafão** não reflete a situação da **reconstrução parcial**, uma vez que há apenas o plural dessa palavra abonado no *corpus*. Uma possível solução é a construção de um verbeta **çafão***, com asterisco à direita no lema

do dicionário, que deve estar no singular. O recurso do asterisco à direita é comum na bibliografia de línguas de parco testemunho escrito, como é a língua gótica (cf. MILLER, 2019) e é usado para reconstruir inexistentes no *corpus*.

Se a intenção é afirmar acerca da **inexistência absoluta** em qualquer *corpus* numa determinada sincronia (ou em todas as sincronias), a ambiguidade do asterisco bloomfieldiano deve ser desfeita e, nesse caso, propõe-se um símbolo distinto, a saber, a **estrela** em vez do asterisco (★). Em suma, (1) pelo símbolo *x se afirmaria que não há testemunhos (numa sincronia ou num *corpus*), mas o dado foi reconstruído; (2) pelo símbolo x* se afirmaria que há testemunhos, mas não em determinada flexão; (3) a notação ★x equivaleria a afirmar que a forma não existe e não é real, sendo essa a razão da falta de testemunhos. Em todos esses casos há hipóteses em que a **certeza** sobre o dado está em questão (VIARO, 2014; VIARO; BIZZOCCHI, 2016).

7 O problema da modernização em Lexicografia

O asterisco à direita permite-nos falar não só sobre o lema em Lexicografia, mas em outras questões de Linguística Histórica. Pode-se afirmar sobre a inexistência no *corpus* de determinadas flexões como **çafei***, **çafamos***, **çafaríamos***, **çafava***, **çafado***, **çafando*** etc. A situação de **çafão*** no *corpus* de Cardoso é a mesma de **çafumar**, **çamarra**, **çambarco**, **çamponha**, **çapa**, **çapata** (e derivados como: **çapatada**, **çapataria**, **çapateiro**, **çapatear**, **çapateta**, **çapatinho**, **çapato**), **Çaragoça**, **çarrafar** (e **çarrafadura**), **çarrar**. Não existem, portanto, no *corpus*, itens lexicais como ***safumar**, ***samarra**, ***sambarco**, ***samponha**, ***sapa**, ***sapata** (e derivados), ***Saragoça**, ***sarrafar** (e ***sarrafadura**), ***sarrar**. Inversamente, há **sábado**, **sabão**, **saber**, **sábio**, **sabor**, **Saboia**, **sambixuga**, mas não ***çábado**, ***çabão**, ***çaber**, ***çábio**, ***çabor**, ***Çaboia**, ***çambixuga**. Observe-se que a sequência fonológica de todos esses dados envolve uma labial */b/, */p/, */m/ ou uma vibrante */r/ ou */r/ na sílaba seguinte, bem como uma posição de ataque na primeira sílaba e o centro vocálico */a/ ou */ã/. Questões como a apresentada no subcapítulo anterior, sobre a neutralização ou não da oposição */s:*/ʃ/ deve levar em conta também outros contextos antes de se chegar a uma conclusão que possa ser espelhada lexicograficamente numa ortografia modernizada.

A modernização de verbetes traz também algumas questões relativas à existência dos dados em dicionários etimológicos. Se a intenção do dicionarista é reunir sob o mesmo **hiperlema** variantes ortográficas e fonéticas (VIARO, 2017), surge aí uma solução lexicográfica mas um grande problema para a Linguística Histórica e para a Etimologia. O

anacronismo não é necessariamente um erro metodológico quando infringe exclusivamente problemas filológicos, mas é algo grave quando revela uma realidade fonética ou fonológica distinta da imaginada nas reconstruções da sincronia pretérita em questão. Uma modernização excessiva que revele fonemas e estruturas em contradição com a época do manuscrito também é algo que faz uma edição modernizada perder seu valor filológico, mesmo quando a intenção é fazer uma edição semidiplomática ou crítica.

A ocorrência de <baldroega> em Cardoso não pode ser uma abonação para um verbete ***beldroega**, inexistente no *corpus*. No entanto, o dicionário Houaiss aponta <baldroega> como *terminus a quo* de **beldroega** (HOUAISS; VILLAR, 2001, p. 426), o que se revela uma informação errônea. Nas suas “formas históricas”, Houaiss e Villar apresentam a grafia <beldroegas> apenas para o final do século XIX; Cunha (1982, p. 104), de fato, apresenta o *terminus a quo* de **beldroega** em uma sincronia bem posterior à de Cardoso, a saber, o século XVIII. As informações de Machado (1967, p. 411) são inconclusivas sobre a possibilidade de retrodatação de **beldroega**, uma vez que a forma apresentada é outra, a saber, o moçárabe <berdilaca>, do século XII, mais próximo de um étimo do que de uma variante do item lexical. De qualquer forma, com essa pequena amostra, pode-se chegar à seguinte conclusão: o *terminus a quo* do hiperlema de Houaiss e Villar refere-se ao dado histórico **baldroega** e não a **beldroega**: a atribuição do século XVI vale apenas para uma entidade lexicográfica (o hiperlema **beldroega**) mas não para uma realidade lexical e etimológica que seja útil para a Linguística Histórica e para a Filologia (ou seja, o lema **beldroega**), que inexistente no *corpus* de Cardoso (e talvez em todos os *corpora* do século XVI, até comprovação contrária) e é, segundo os dados apresentados, pertencente a uma sincronia posterior (século XVIII ou XIX).

A existência de <beldroega> no *corpus* estudado de ***beldroega** e a inexistência absoluta de **★beldroega** no século XVI, a julgar pelo *corpus* cardosiano e pelas informações dos dicionários etimológicos, valeria o mesmo para a existência de <balea> em relação a uma hipótese da inexistência de **★baleia**, de <bataria> à de **★bateria**, de <bisconde> à de **★visconde**, de <biscouto> à de **★biscoito**, de <bisouro> à de **★besouro**, de <bespara> à de **★véspera**, de <boluer> à de **★volver**, de <bozina> à de **★buzina**, de <cangrejo> e <cranguejo> à de **★caranguejo** etc. Do mesmo modo não se unifica, em um dicionário com preocupações etimológicas e históricas, sob um único hiperlema itens lexicais com diferenças fonéticas, tais como <bandouua> e <bandouba>, <barboleta> e <borboleta>, <baxo> e <baixo>, <salairo> e <salario>, <barruga>, <berruga> e <verruga>. Cada uma das variantes existentes (que não são meras variantes gráficas) deve ter lemas independentes, pois os percursos fonológicos diacrônicos foram distintos e, conseqüentemente, o *terminus*

a quo de cada variante também será distinto. O máximo que pode ocorrer, neste caso, é a remissão de um verbete para outro, cuja distinção entre principal e secundário é determinada, como visto, um tanto arbitrariamente (normalmente, nesses casos, um critério anacrônico da maior frequência de uso na sincronia atual costuma prevalecer, quando o mais coerente seria a maior frequência de ocorrências no *corpus*). Em dicionários etimológicos, como vimos, essa remissão também equivale a uma única explicação etimológica no verbete principal.

Para evitar perda de informação etimológica, uma modernização de verbetes em Lexicografia pode optar, portanto, somente pela ortografia mais próxima da moderna quando isso não significa perda de informação fonética, por exemplo, a opção do lema **Coimbra** entre as variantes <Coimbra>, <Coïbra> e <Coymbra> no mesmo *corpus*. O mesmo verbete **campo** pode ser igualmente utilizado para variantes <campo> e <cãpo>. Nesse processo, contudo, podem surgir **modernizações parciais**. Seguindo a lógica aqui apresentada, <bautizar>, <boutizar>, <bautejar> não sustentam a existência de um verbete ★**batizar**, mas três distintos: **bautizar**, **boutizar** e **batejar**. No entanto, <baptismo> poderia gerar um verbete **batismo**, se é entendido, na modernização, que <pt> é uma representação de */t/. Nesse caso, forma-se uma unidade real, mas inexistente. A solução alternativa seria um verbete **baptismo**, que revelaria ou uma ausência de modernização ou uma interpretação de que o dado existente <pt> corresponde ao dado reconstruído */pt/ (que se torna real para a discussão ulterior da Linguística Histórica).

Emerge, no último exemplo, a estranha situação, já prevista antes, de que, pela exclusão dos dados não-reais (entendidos como impossíveis de se reconstruir), haveria apenas dados existentes e reais, assim como inexistentes e reais. No caso do dado **batismo**, a Filologia defenderia a necessidade de apontar esse problema de alguma forma, por exemplo, por meio de colchetes quadrados: **ba[p]tismo**. Um outro recurso, proposto pela Lexicografia, seria colocar o dado real como verbete, independentemente do *corpus*, e apresentar os dados existentes depois, por exemplo, um verbete como:

batismo: sb masc || sg <baptismo> Lustricus dies. Ho dia do baptismo. [1570¹, 127v, A, 26]. ◊ Var: *bautismo*.

bautismo: sb masc || sg <bautismo> Bautismo. Baptisma, atis. [1562-1563¹, 24v, B, 24]; Bautismo. Baptisma, atis. [1570², 17v, B, 3]; Baptisma, atis. Ho bautismo. [1570¹, 36v, A, 33]; Baptisterium, ij. A pia do bautismo. [1570¹, 36v, A, 34]; Pater spiritalis . Ho padrinho do bau/tismo. [1570¹, 175r, A, 9-10]. ◊ Var: *batismo*.

O verbete, em **negrito**, sendo o dado real, mas não o existente (que estaria em **itálico**), somente sofreu modernização para adequar-se à representação fonologicamente sustentada */ba'tismo/. Uma representação filológica como **ba[p]tismo** teria o mesmo intuito e talvez induza menos a erro, por ser mais redundante. Obviamente nenhum desses problemas surgiria se o verbete, em vez de **batismo**, **baptismo** ou **ba[p]tismo** fosse */ba'tismo/, mas apenas a tradição lexicográfica parece impedir esse tipo de verbete. O problema de **batismo** é o mesmo de outros lemas, que não se apoiam diretamente dados, por exemplo, **birinjela** para <biringela>, **bisseisto** para <biseisto>, **Biscaia** para <Bizcaya>, **buticairo** para <buticayro>, **cogombro** para <cogõbro>, **breve** para <breue>, **búzio** para <buzio>, **cácere** para <caçere>, **Campânia** para <cãpania> etc. Em uma solução sustentada filologicamente, seriam necessárias indicações especiais, por exemplo, com colchetes uncinados (para acréscimos) ou letras sublinhadas (para modernizações): **birinjela**, **bí<s>seisto**, **Biscaia**, **buticairo**, **cogombro**, **breve**, **búzio**, **cácere**, **Campânia**.

Como se pode observar, a tonicidade se evidencia ao se atualizar a ortografia. No entanto, isso pode gerar dúvidas: <bençam> é paroxítona como determina a norma culta atual ou ainda era oxítona na sincronia pretérita da qual o dado foi coletado? Havia ditongação ou não? Tudo isso pode pesar na construção de lemas híbridos, que não se apoiam diretamente nos dados, como **bênção**, **benção**, **bençã** ou **bênçã**. Os dados relativos à tonicidade e à ditongação que emergem dessas decisões devem ser argumentativamente justificados como casos especiais num prefácio, pois haverá consequências quantitativas e qualitativas por causa delas.

Por fim, a necessidade de se reconstruir verbetes a partir de dados inexistentes se dá quando se supõe haver **erro tipográfico** no dado: a partir de <bargantatia> pode-se deduzir um lema corrigido ***bargantaria**, assim como de <ciscouto> se reconstrói ***biscouto**. O lema pode perder o asterisco ao final do levantamento do *corpus* por causa de outras ocorrências sem erros. Nesse caso, o dado, proveniente de um suposto erro tipográfico, é que posto em dúvida, marcado, por exemplo, usando-se um ponto de interrogação superscrito no início da abonação (?). Cf.:

bargantaria: *sb fem* || *sg* <bargantaria, bargantatia> Flagitium, ij. A bargantaria. [1570¹, 84r, B, 40]; Libera corcyra, caca vbi libet. (...); [?]Sybaritica calamitas . Dirseha da/ grãde destruição como foy a dos Sy/baritas cuja bargantatia foy causa de/ serem postos por terra. [1570¹, 245v, A, 28-31].

Reconstruções desse tipo surgem em desenvolvimentos de abreviaturas, por exemplo, <busq̃mos> é única ocorrência da flexão ***busquemos** no *corpus* de Cardoso, no entanto, esse asterisco pode ter um grau de certeza menor do que o uso do asterisco à direita. Sabe-se que em verbos irregulares atua, por vezes, uma interferência analógica regularizante, mas um verbo regular numa sincronia anterior à sincronia em análise não teria razões para ter sido irregular se na sincronia posterior a ela também há regularidade. A formação das irregularidades nasce sobretudo das grandes transformações fonéticas. Quando não há razões para pensar isso, pressupõe-se uma certa inércia, em que os regulares de outrora que são regulares posteriormente continuaram sendo regulares no *corpus* que não o documenta ou no que o documenta parcialmente. Em Filologia, esse dado necessariamente deve ser indicado com o auxílio de parênteses ou de desenvolvimento em itálico: **busq(ue)mos** ou **busquemos**.

8 Conclusão

Em suma, de tudo que foi apresentado, é possível determinar ao menos oito situações relativas à existência de um dado linguístico real:

- 1) O dado é testemunhado, portanto, existente na sua materialidade, sem qualquer necessidade de reconstrução. Essa é a única situação em que a existência e a realidade do dado se equivalem. Para um dado imaterial, por exemplo, um dado semântico, a isso quase nunca se atribuirá a mesma certeza.
- 2) O dado não é testemunhado, mas é possível fazer uma reconstrução dele em consonância com outros dados obtidos comparativamente. Nesse caso, nenhum dado existente sobreviveu na sincronia estudada, por alguma razão (por exemplo, expressão linguística dessa sincronia era puramente ágrafa). A proposição de sua realidade deve respeitar regras dedutivas em que se baseia a argumentação a seu favor, como leis fonéticas e outros dados da Linguística Comparativa. Esse é o caso do dado antecedido por um asterisco ***x** (ou às vezes por dois asteriscos ****x** em algumas situações), por exemplo, a reconstrução de um item lexical em indo-europeu ou em prototupi.
- 3) O dado não é testemunhado, mas é possível fazer uma reconstrução dele em consonância com outros dados materialmente existentes na mesma sincronia pretérita. Nesse caso, a expressão escrita faz parte da realidade dessa sincronia. Sua inexistência é atribuída ao acaso ou ao tipo de registro

parcamente documentado (por exemplo, dados pertencentes ao uso coloquial em uma sincronia pretérita). No entanto, a sua realidade é defendida por haver **consequências** de sua existência. É o caso, por exemplo, de um étimo como ***excadescere** para o item lexical **esquecer**, cujo grau de certeza de sua realidade é considerado possível ou provável (VIARO; BIZZOCCHI, 2016), uma vez que há não só documentação de formas subseqüentes (cf. **escaecer**, **esqueecer**), com consequências na sincronia atual (cf. a abertura da pretônica [ɛ] no português lusitano atual), mas também paralelos semânticos obtidos pela Linguística Comparativa, que apontam para metáforas similares (cf. o italiano **dimenticare** traz a metáfora conceptual semelhante de “sair da mente” assim como ***excadescere** seria literalmente “cair para fora”). Como em (2), é imprescindível o uso do asterisco schleicheriano (à esquerda) neste caso.

- 4) O dado não é testemunhado no *corpus* estudado de uma determinada sincronia pretérita, mas existe em outros *corpora* da mesma sincronia. Por exemplo, no *corpus* cardosiano não há ***búfalo**, mas apenas <bufaro>, que conduz ao verbete **búfaro**, no entanto, **búfalo** é atestado em outros *corpora* do século XVI. O emprego do asterisco à esquerda aqui depende da perspectiva adotada: usa-se o asterisco se o que o dado se refere ao *corpus*, mas descarta-se no caso de ser a descrição de toda uma sincronia pretérita. A única exceção seria a situação (8) abaixo. Enquadram-se aqui também dados que aparecem em *corpora* anteriores e posteriores à sincronia estudada, onde não foi encontrado, formando um intervalo de ocorrências de ocorrências que permita a reconstrução. Esse recurso se emprega cautelosamente quando não se trata de dado de origem culta, sobre o qual se pode aventar uma ressurreição por meio da escrita. Não se trata do caso de **★beldroega** em Cardoso, cuja existência em Houaiss e Villar (2001) não se sustenta e parece apoiar-se na idealização normativa de estilo vaugeoisiano, de modo que se considera “existente” exclusivamente a norma culta atual, mas isso obviamente é um emprego abusivo do conceito de existência, que deveria ser rejeitado pela Linguística que se pretende científica, embora seja comum na prática lexicográfica tradicional.
- 5) O dado existe parcialmente, ou seja, pode aparecer no *corpus* apenas em algumas flexões, mas não em todas. Nesse caso, as flexões que não são abonadas podem ser distinguidas por um asterisco à direita e é importante

para as flexões tradicionalmente utilizadas como lemas (infinitivo para verbos, singular para substantivos, masculino singular para adjetivos etc.): no **corpus** cardosiano há **buscar, buscarem, buscas, busca, buscam, buscavas, buscara, busque, busquemos, buscada, buscando** mas não **busco***, **busquei***, **buscávamos***, **buscado***, **buscássemos*** etc. A indicação da realidade de uma flexão inexistente é importante para muitas argumentações em Linguística Histórica. O uso do asterisco à direita descarta a possibilidade redundante de um emprego antes e depois do item lexical (***x***), uma vez que um dado assim seria apenas algo reconstruído (***x**), que inexistiria numa determinada flexão, como nas quatro primeiras situações acima apresentadas. Dito de outro modo, **x*** só seria empregável para dados dos *corpora* e não para dados de uma sincronia.

- 6) O dado pode existir parcialmente, ou seja, sua realidade requer reconstrução. Nesse sentido, é real graficamente, mas não é um dado com realidade fonética, por se tratar de um erro gráfico, como <amosa> no dado “Amosa. Morum, i. [1570², 9v, B, 30]” que nada mais é que ***amora** (ou, filologicamente **amora**), ou <armae> ou <cousu>, que contextualmente em “Armiger. a. um. Cousu ã traz armae. [1570¹, 32r, B, 18]” equivalem respectivamente a ***armas** e ***cousa** (ou, filologicamente, **armas** e **cousa**). Dados como este às vezes têm algum valor filológico como “erros”, mas não são valorizados, de modo geral, pela Linguística, a ponto de serem ignorados. A afirmação, contudo, que um dado é apenas um aparente erro gráfico pressupõe argumentos robustos e promovem por vezes revisão dos dados, podendo ter certo impacto na teorização da Linguística Histórica.
- 7) O dado não existe graficamente no *corpus*, mas sua realidade é linguisticamente relevante: o dado existente no *corpus* cardosiano é <Aristophanes>, mas o dado real é **Aristófanes**. Nesse caso, prescinde-se do asterisco embora seja **Aristófanes** seja um dado inexistente para a Filologia, por exemplo, ao se fazer uma edição diplomática ou em algumas semidiplomáticas (para as quais só existiria **Aristophanes**). Apesar de ser uma reconstrução, o dado **Aristófanes** não ser marcado com nenhum símbolo, em Lexicografia, como no caso (6). O valor dessa prática é pequeno para a Filologia, mas grande para a Linguística Histórica, por exemplo, numa quantificação de palavras proparoxítonas em um *corpus* ou na descrição de uma sincronia pretérita.

- 8) Diferentemente dos sete situações acima, o dado pode ser considerado **irreal**, i.e. inexistente de maneira absoluta numa sincronia e não apenas num *corpus*. A essa conclusão se chega dedutivamente. A afirmação da inexistência absoluta, isso é, de sua imaterialidade dotada de uma realidade que beira o irreconhecível ou o incognoscível (apenas atingível por meio da imaginação e não por algum grau de certeza que envolva uma aposta da reconstrução) é muito comum em estudos sincrônicos de Linguística, embora bem mais ousada. Como dito, conviria que fosse indicada por outro símbolo distinto do asterisco à esquerda quando trabalhada juntamente com dados passíveis de reconstrução. Propôs-se, para tal, o uso da estrela (★).

Em todas as oito situações, como atitude de honestidade científica – incluindo a situação (8) da inexistência absoluta - eliminam-se os símbolos relativos à falta de existência do dado supostamente inexistente quando o mesmo dado é finalmente encontrado.

O celacanto era considerado um fóssil ou mesmo um ser fictício até que um deles foi pescado em 1938 e descrito com os rigores científicos em 1939. Apesar de celacantos linguísticos surgirem o tempo todo, nem sempre se obedece da mesma maneira rigorosa, organizada e sistemática a transformação dados irreais em reais, ou de dados inexistentes em existentes, para prejuízo do Conhecimento na Ciência linguística.

Referências

- BRANQUINHO, J.; MURCHO, D.; GOMES, N. G. **Enciclopédia de termos lógico-filosóficos**. São Paulo: Martins Fontes, 2006.
- BLOOMFIELD, L. **Language**. New York/ Chicago/ San Francisco: Holt, Rinehart and Winston, 1966 [1933¹].
- BUCHI, É.; SCHWEICKARD, W. (ed). **Dictionnaire étymologique roman (DÉRom): genèse, méthodes et résultats**. Berlin/ München/ Boston: Walter de Gruyter, 2014.
- CAMBRAIA, C. N. **Introdução à crítica textual**. São Paulo: Martins Fontes, 2005.
- BUNGE, M. **Dicionário de filosofia**. Trad. G. K. Guinsburg. São Paulo: Perspectiva, 2002.
- CARDOSO, J. **Hieronimi Cardosi Lamacensis dictionarium ex lusitanico in latinum sermonem**. Lisboa: Ex officina Ioannis Aluari typographi Regij, 1562-1563. Disponível em: <https://purl.pt/15192>. Acesso em: 25 jul. 2022.

CARDOSO, J. **Dictionarium latinolusitanicum & vice versa Lusitanico latinũ.** Coimbra: Ioan. Barrerius, 1570. Disponível em: <https://purl.pt/14265>. Acesso em: 25 jul. 2022.

CHOMSKY, N.; HALLE, C. **The sound pattern of English.** Cambridge/ London: The MIT Press, 1968.

CUNHA, A. G. da. **Dicionário etimológico Nova Fronteira da língua portuguesa.** 2ª ed., 6ª impressão. Rio de Janeiro: Nova Fronteira, 1982.

GRAFF, G. The asterisk from historical to descriptive and theoretical linguistics: an historical note. **Historiographia Linguística**, Amsterdam, v. 29, n. 3, p. 329-340, 2002. Disponível em: https://www.researchgate.net/publication/233564347_The_Asterisk_from_Historical_to_Descriptive_and_Theoretical_Linguistics_An_historical_note. Acesso em 26 jul. 2022.

HEGENBERG, L. **Definições:** termos teóricos e significado. São Paulo: Cultrix/EDUSP, 1974.

HEGENBERG, L. **Explicações científicas:** introdução à Filosofia da Ciência. São Paulo: Herder/EDUSP, 1969.

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

KUHN, T. S. **The structure of scientific revolutions.** 3ª ed. Chicago: University of Chicago, 1996.

MACHADO, J. P. **Dicionário etimológico da língua portuguesa.** 2ª ed. Lisboa: Confluência/ Horizonte, 1967.

MAIA, C. de A. **História do galego português:** estado linguístico da Galiza e do Nordeste de Portugal desde o século XIII ao século XVI, com referência à situação do galego moderno. Lisboa: Fundação Calouste Gulbenkian/ Junta Nacional de Investigação Científica e Tecnológica, 1986.

MAURER JR, T. H. **A unidade da România Ocidental.** São Paulo: s/ed, 1951.

MAURER JR, T. H. **Gramática do latim vulgar.** Rio de Janeiro: Acadêmica, 1959.

MAURER JR, T. H. **O problema do latim vulgar.** Rio de Janeiro: Acadêmica, 1962.

MILLER, G. **The Oxford Gothic grammar.** Oxford: OUP, 2019.

RICKEN, F. **Dicionário de Teoria do Conhecimento e Metafísica.** Trad. I. Kayser. São Leopoldo: Unisinos, 2002.

SAUSSURE, F. de. **Cours de linguistique générale.** Publié par C. Bailly et A. Séchehaye avec la collaboration de A. Riedlinger. Édition critique préparée par T. de Mauro. Paris: Payo, 1967 [1916¹].

SILVA NETO, S. da. **Fontes do latim vulgar: o Appendix Probi.** 3ª ed. Rio de Janeiro: Acadêmica, 1956.

SILVA NETO, S. da. **História do latim vulgar.** Rio de Janeiro: Ao Livro Técnico, 1977.

SPAGGIARI, B.; PERUGI, M. **Fundamentos de Crítica Textual**. Rio de Janeiro: Lucerna, 2004.

SPINA, S. **Introdução à Edótica**. 2ª ed. São Paulo: Ars Poetica/ EDUSP, 1994.

TEYSSIER, P. **História da língua portuguesa**. Trad. Celso Cunha. Lisboa: Sá da Costa, 1990 [Histoire de la langue portugaise. Paris: PUF, 1980¹].

VASCONCELOS, J. L. **Opúsculos: Dialetoлогия** (parte 1). v.2. Coimbra: Imprensa da Universidade, 1928.

VASCONCELOS, J. L. **Opúsculos: Dialetoлогия** (parte 2), v.6. Coimbra: Imprensa Nacional/ Casa da Moeda, 1985.

VIARO, M. E. Linguística da comunicação e linguística descritiva: os eixos sincrônico e diacrônico nos atuais modelos de Morfologia. **Estudos linguísticos**, São Paulo, v. 41, n.1, p. 277-290, jan-abr, 2012. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/1232/785>. Acesso em 25 jul. 2022.

VIARO, M. E. **Etimologia**. São Paulo: Globo, 2014.

VIARO, M. E.; BIZZOCCHI, A. L. Proposta de novos conceitos e uma nova notação na formulação de proposições e discussões etimológicas. **Alfa: revista de linguística**, São Paulo, v. 60, n. 3, p. 579-601, 2016. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/8453>. Acesso em: 25 jul. 2022.

VIARO, M. E. O Dicionário Etimológico da Língua Portuguesa (DELPo): conceitos de metalema, hemilema, hiperlema e ultralema. In: DE ROSA, G. L.; DEGLI ATTI, F.; CHULATA, K. de A.; MORLEO, F. (org.). **De volta ao futuro da língua portuguesa: Atas do VSIMELP – Simpósio Mundial de Estudos de Língua Portuguesa**. Lecce: Università di Salento, p. 143-156, 2017. Disponível em: <http://sibaese.unisalento.it/index.php/dvaf/article/view/17775/15134>. Acesso em: 26 jul. 2022.



On the problem of existence of linguistic data

ABSTRACT:

There are several ways to understand the existence of linguistic data. Their relevance cannot be measured solely by their materiality, once Historical Linguistics, Philology and Etymology also recognize the reality of various non-material data as reconstructions. These data are also used by General Linguistics and by Lexicography in particular. There are constantly some philological problems related to the updating of their graphemes, which is traditionally done in order to avoid anachronism, that can confer a lack of scientificity to the linguistic data and the analyses based on it. Several situations involving the question of the existence and reality of linguistic data obtained from three works by Jerónimo Cardoso (1508-1569) are presented in this paper.

KEYWORDS:

Linguistic data;
Existence;
Historical Linguistics;
Lexicography;
Philology.