

ENTREVISTA

O Dicionário Houaiss da Língua Portuguesa foi publicado em 2001

Manoel Pinto Ribeiro entrevista Mauro Villar

Há dez anos vem sendo elaborado o *Dicionário Houaiss da Língua Portuguesa*. Trata-se de obra monumental, cuja publicação foi lançada em 2001 – pela Editora Objetiva, no Brasil, e pelo Círculo de Leitores, em Portugal. Para se ter uma idéia desse trabalho, que, após o falecimento de Antônio Houaiss, em 7 de março de 1998, ficou sob a direção técnica do seu co-autor, Mauro de Salles Villar, a Academia Brasileira de Filologia publica, neste número, uma entrevista com esse ilustre lexicógrafo.

Manoel Pinto Ribeiro – Como se faz um dicionário?

Mauro de Salles Villar – Muita gente se pergunta como são eles feitos. A crença popular é a de que quem neles trabalha *sabe tudo*. Obviamente, porém, tal coisa nem possível seria. A lexicografia é um trabalho de gente como qualquer outra, não necessariamente super-homens intelectuais. É, por outro lado, uma técnica desenvolvida em muitos anos de labor humano. Bastaria referir que o primeiro dicionário conhecido foi produzido em Ebla, um dos grandes centros urbanos da cultura proto-síria, há no mínimo 4.300 anos.

E o que são os lexicógrafos? Basicamente, redatores generalistas que dominam as regras da metalinguagem empregada no fazer de dicionários, enciclopédias e afins. É por essa razão que, para elaborarem tais obras de referência, carecem necessariamente da contribuição de um conjunto de especialistas das mais diversas áreas do saber e do fazer humanos, contingente capaz de suprir as lacunas de informação técnica que ocorram nas definições.

Quantas pessoas trabalham na elaboração de um grande dicionário?

As enciclopédias, os glossários, os dicionários, tais como existem hoje, são uma atividade de complexa estrutura, não podendo mais ser obra de um grupo singelo de redatores auto-suficientes, por mais abarcantes que sejam os seus conhecimentos. Humanistas desse tipo não mais existem em nosso mundo, um tempo em que a realidade se tornou de tal maneira multímoda, no que respeita à sua variedade e extensão, e em que as suas prospecções verticais multiplica-

ram-se tanto e atingiram tamanhas profundidades, especialmente no campo das ciências e da tecnologia, que uma visão abarcante e absoluta é impossível, a não ser, de modo não mais que aproximativo, como fruto de um vigoroso trabalho colegiado.

Essa é a razão por que um grande dicionário, como o *Third new international Webster*, lançado nos Estados Unidos em 1961 com 460.000 entradas, contou com uma força-tarefa de 1 editor-chefe, 13 redatores associados, 58 redatores assistentes, 66 assistentes editoriais, 31 assistentes de secretaria e 343 especialistas externos utilizados como consultores. Em outras palavras, o numeroso grupo editorial interno teve como adjutório forçoso o trabalho conexo de centenas de especialistas, dentro e fora dos Estados Unidos, para levar a cabo a sua missão.

De que maneira os dicionários ditos de língua geral escolhem suas nominatas?

Nos dicionários ditos de língua geral, a base das averbações é de palavras da língua de *nível zero* – a não-científica ou técnica, a língua comum do dia-a-dia e a literária. Nesse campo, a competência básica é dos lexicógrafos – lingüistas, gramáticos e filólogos. Mas o contingente numericamente fundamental das línguas modernas é composto pela terminologia, sistema de palavras empregadas numa disciplina particular (por exemplo, a terminologia botânica, ou da marinharia, ou da matemática). Foi nessa área que explodiram quantitativamente as línguas de cultura no século XIX e no XX, especialmente aquelas que contam com grande desenvolvimento nas tecnologias de ponta. Todavia, como a nomenclatura técnica é crescentemente universal, cada neologismo criado por força de nova tecnologia ou descoberta é também, pelo menos potencialmente, palavra nova de qualquer outra língua cujos falantes dela venham a se utilizar.

Poderia dar-nos uma idéia da expansão lexical do português?

Para se ter uma idéia desse fenômeno de expansão lexical, comparo aqui os números da língua portuguesa na medievalidade com aqueles de hoje. Calcula-se que o português medieval contasse com cerca de 15 mil vocábulos. Em meados do século XVI, com a expansão marítima, esse total estaria entre 30 e 40 mil. No final do século XIX, o registro em dicionários atingia a casa das 80-90 mil unidades léxicas. Saltando daí para o início da década de 1980, o *corpus* do *Vocabulário ortográfico* da Academia Brasileira de Letras já registra cerca de 360 mil entradas.

A dicionarização representa o número exato de palavras de uma língua?

De modo algum. Se o *Oxford english dictionary*, essa catedral da lexicografia internacional, já conseguiu registrar mais de 700 mil unidades da Língua Inglesa, não se pense que o seu léxico se esgote em tal quantitativo. Muito longe disto.

Os projetos de lexicografia, para se comporem, fazem uma escolha dentre o extensíssimo material com que conta a língua, seja determinando no tempo o período fechado sobre o qual focalizarão o seu esforço de pesquisa, seja prefixando o número de entradas sobre as quais irão versar. Os dicionários, especialmente os de língua geral, jamais se utilizam de modo exaustivo das nominatas terminológicas. Nem poderiam fazê-lo.

Qualquer língua moderna conta, não com centenas de milhares de vocábulos, mas com milhões de palavras e sintagmas potencialmente dicionarizáveis. Calcula-se, por exemplo, que só os vocábulos e locuções ligados à medicina girem em torno de um total de 600 mil unidades. A terminologia química e farmacológica atinge quase 2 milhões de vocábulos, total excedido pela zoologia, pois só a entomologia tem 2 milhões de insetos classificados. Um registro abarcante da botânica universal não seria menos titânico. E se é verdade que a maior parte das espécies, gêneros, classes, tribos etc. dessas classificações tenha denominação apenas no latim científico, seus nomes são todos vernacularizáveis e, portanto, registráveis em dicionários, mesmo naqueles de língua geral. E pensar que apenas 13% das espécies da biodiversidade da Terra foram classificados até agora...

Que podem fazer os lexicógrafos diante de tal complexidade?

Quanto à linguagem especial do tecnoleto, os lexicógrafos têm mesmo de se valer do conhecimento de profissionais de cada área. Quem trabalha em lexicografia saberá padronizar, no estilo do dicionário que estiver fazendo, o material técnico recebido, mas fica-se dependente da qualidade, exatidão e utilidade das informações prestadas pelo colaborador especialista.

Que outras obrigações têm os lexicógrafos nos dicionários?

Em línguas bem estudadas, o seu papel é aprofundar a pesquisa; é procurar exceler na definição dos conceitos das unidades léxicas, tentando levá-las a uma eficácia e precisão cada vez maiores. É também incluir, com critério, entradas neológicas e a neologia ligada às acepções. Se a questão for de etimologia ou sobre a história de um vocábulo, é preciso ir além do que antes havia sido feito.

A lexicografia como registro da extensão horizontal e vertical das vivências e cogitações de um povo através de sua língua é um processo de agregação, no tempo, de esforços, pesquisas e resultados, um somatório de conquistas que se acumulam umas sobre as outras. Um grande dicionário espelha e baliza a expansão e os limites do pensamento e das conquistas materiais dos seus falantes. Por esta razão, tais obras são tão importantes para a política cultural dos povos.

Há países que vivenciam com mais empenho que nós a política da língua?

Por certo. A Espanha, por exemplo, cedo adiantou-se às demais línguas

européias na intuição da importância dos dicionários com o seu *Tesoro de la Lengua Castellana y Española*, de Sebastián de Covarrubias. Publicado em 1611, ele precedeu de um ano o célebre *Vocabolario degli Accademici della Crusca*, de Florença, editado em Veneza. No século seguinte, a Espanha foi além com o primeiro Borbón, Filipe IV, monarca esclarecido que em 1723 destinou por decreto uma renda anual, obtida por um imposto sobre o tabaco, para a publicação do *Diccionario de autoridades* da Real Academia Española, monumento da lexicografia desse país.

E na França?

Na França, a consciência da importância decisiva do idioma como uma espécie de “cola social” e ferramenta política de hegemonia cultural é patente tanto historicamente quanto no conjunto de meios e financiamentos postos à disposição de lexicógrafos para a elaboração do *Trésor de la Langue Française* em 16 alentados volumes - e isso numa língua que já contava com excelentes dicionários, como o *Littré*, o *Larousse*, o *Robert* e os da Editora Hachette, para citar apenas quatro bons filios lexicográficos.

Mas, voltando à nossa realidade, que podem fazer os lexicógrafos quando aspectos da língua sobre a qual um dicionário versa se encontram mal estudados, quando houve interrupção nos esforços de compreendê-los, pesquisá-los e registrá-los? Que fazer quando as informações especializadas sobre que basear os esforços do registro lexicográfico são por vezes pouco confiáveis, confusas, a exigir estudos que as retifiquem?

Os lexicógrafos sofrem algum tipo de pressão em seu trabalho?

Os dicionários e enciclopédias congeminam em si duas naturezas: são produtos culturais e bens de comércio. Por esse segundo motivo, sobre eles incidem as ubíquas pressões por produção, prazos, contenção de custos, redução de grupos de trabalho, moderação de pesquisa. Em outras palavras, acaba por faltar tempo para a reflexão mais aprofundada, para as verificações mais minuciosas, para as leituras superpostas do material trabalhado, que é o meio que se tem de realizar uma filtragem dos deslizes advenientes do forte ritmo da produção. Os dicionários, gravados por tais exigências alheias ao seu processo, passam então a se apressar, repetindo-se, copiando definições, reproduzindo velhas soluções propostas para etimologias, replicando tradicionais procedimentos técnicos, sem ânimo ou possibilidade de sugerir avanços, estabelecer características mais pessoais, pesquisar desenvolvimentos.

O perigo disso é o de se acabar fazendo sempre o mesmo dicionário, apenas cosmeticamente alterado, com permanência de equívocos definitórios e reinclusão das nefastas *palavras de papel*, que é como chamamos aquelas que se encontram em vocabulários, léxicos e em outros livros de autoridade similar, mas sem curso real na língua – palavras originárias de erros ortográficos, tipográficos, ou resultantes de má leitura de originais ou de pronúncia incorreta.

Como foi atacado o problema dos tupinismos no *Dicionário Houaiss*?

Há consideráveis divergências sobre o que se entende lingüisticamente por tupi e por guarani nas obras de referência e dicionários que tratam da questão no Brasil. Basta comparar o que dizem a *Enciclopédia Mirador internacional* (texto redigido na década de 1970 que retrata o pensamento da antropologia brasileira coeva), o professor sul-rio-grandense Frederico G. Edelweiss (no ensaio denominado “O indianismo ao tempo do império”, de janeiro de 1955, constante da abertura da 4ª edição de *O tupi na geografia nacional*, de Teodoro Sampaio), o próprio Teodoro Sampaio no capítulo 1 desse livro e o professor Eduardo de Almeida Navarro (em seu *Método moderno de Tupi antigo*). Considerem-se também as opiniões de Martius, Porto Seguro, Couto de Magalhães, Batista Caetano e Barbosa Rodrigues – só para citar uns quantos.

O quadro das lingüístico indígena do Brasil apresenta ainda hoje cerca de 200 línguas. Na época do descobrimento crê-se que houvesse talvez umas 500, faladas por 5 milhões de indígenas. (Hoje estes são cerca de 300 mil.)

Que, afinal, engloba a denominação tupi ou tupinismo?

Sob a denominação comum de tupi ou vagamente de tupinismo, têm entrado nos dicionários de língua portuguesa, do mesmo modo, o tupi antigo, o guarani, o nheengatu (também dito neotupi) e eventualmente vocábulos cujos étimos pertenceriam a outros troncos e a outras famílias lingüísticas. (O quadro lingüístico indígena do Brasil apresenta aproximadamente a metade do total de línguas faladas na época do descobrimento, mas ainda assim isso representa de 120 a 150.)

Além de os estudos etimológicos estarem consideravelmente atrasados quanto à lingüística indigenista, a utilização do rótulo *tupi* num sentido hiperonímico de “empréstimo de alguma vaga língua indígena do continente americano” continua a ocorrer e está longe de ser ideal, mas os dicionários persistem em tal simplificação por falta de fontes sobre que debruçar os seus levantamentos.

Como a equipe de lexicógrafos do *Dicionário Houaiss* tem trabalhado com esse óbice de fontes?

Também nos ressentimos dessa limitação endêmica, mas empenhamo-nos ferreamente em levar o mais longe possível as pesquisas que se podem fazer e as sugestões que se podem oferecer sobre a etimologia de tais empréstimos.

No que respeita à lingüística e à etnologia indigenista nacional, preferimos fazer tábua rasa do que os outros léxicos e enciclopédias registravam, por confuso, desatualizado e por vezes errôneo ou impossível de comprovar, e lançar-nos a um levantamento próprio, contemporâneo e histórico, de grupos e línguas indígenas baseado na documentação mais recente dessa área. Disto resultaram cerca de 650 verbetes que incluem informações sobre o indivíduo, o grupo com sua localização geográfica e a língua ou dialeto por eles falado, além

do registro de seu etnônimo brasílico, vale dizer, a denominação de cada grupo do território brasileiro, extinto ou existente, com sua transcrição segundo regras estabelecidas por antropólogos e lingüistas brasileiros.

Mas quantos seriam os indigenismos no português?

No que concerne à quantificação dos indigenismos lingüísticos no português, ou seja, das palavras, construções ou locuções de qualquer língua indígena americana tomadas de empréstimo por nossa língua, o cômputo existente é caótico, pois nunca foi levado a efeito com método. As opiniões divergem de meros 4.500 vocábulos aos 50 a 100 mil estimados por Silveira Bueno. Não creio haver dúvida de que as vozes amerindigenistas superam em número as dicções árabes conservadas no léxico português. Afinal, é patente a quantidade de topônimos, os numerosíssimos fitônimos, além de diversos etnônimos, litônimos, nosônimos, melônimos, mitônimos, meteorônimos, demonônimos, alguns antropônimos etc. — que colorem a nossa língua, especialmente a falada no Brasil, de belos sons e visões ameríndias. Todavia, no que respeita à sua quantificação real, não há rigorosamente mais que meras presunções.

E no *Dicionário Houaiss*?

No que respeita a nomes de animais e especialmente de plantas brasileiras, é bastante considerável o número de empréstimos indígenas, dos quais se registram também grande número de formas e variantes. Dou como exemplo o verbete *guapuruvu*, uma leguminosa que chega a atingir 30 metros de altura. Entre os 51 sinônimos e variantes que o dicionário refere, estão: *baageiro*, *bacumbu*, *bacuparu*, *bacurubu*, *bacurubu-ficheira*, *bacuruvu*, *bacuva*, *bageiro*, *bagiro*, *baquerubu*, *baqueruvu*, *beri*, *bucurubu*, *gabiruvu*, *gapuvuru*, *garapivu*, *garapuva*, *garapuvu*, *garipuvu*, *grapibu*, *guaburuvu*, *guaparuva*, *guaperubu*, *guaperuvu*, *guapiruva*, *guapiruvi*, *guapiruvu*, *guapivuçu*, *guaporuba*, *guapurubu*, *guapuruva*, *guarapuvu*, *guarupuvu*, *guavirova*, *guavirovo*, *paricá*, *paricá-grande* etc. — grande número dos quais, indigenismos.

Na medida em que tal registro é tão expressivo na obra, a fixação de suas etimologias constitui-se num ordálio para os especialistas, já pela divergência de opiniões sobre étimos, já pela infirmitade do conceito do que é o tupi nas obras de referência utilizáveis, já por serem o tronco Aruák e as famílias Karíb, Tukâno, Makú, Pâno etc. pouco estudadas, já — e pior que tudo — pela inexistência de estudos sobre grande parte de tais empréstimos, o que exigiu que o nosso grupo de etimologistas tivesse de ir além do que se costumava pedir à sua classe nos dicionários de língua portuguesa dita geral até agora realizados.

E quanto à fixação da grafia dos nomes indígenas?

Essa é realmente outra questão espinhosa. O item IV, sobre a hifenação, do Formulário Ortográfico da Academia Brasileira de Letras, aprovado em agosto de 1943, estatui, no inciso 4.º, que ele deve ser empregado “nos vocábulos formados por sufixos que representam formas adjetivas como *açu*, *guaçu* e

mirim, quando o exige a pronúncia e quando o primeiro elemento acaba em vogal acentuada graficamente”. Exemplificando tal regra, os vocábulos *andá-açu*, *amoré-guaçu*, *anajá-mirim* e *capim-açu*. Mas que dizer de *baiacumirim*, em que o u é semitônico ou de *carapicupeba*, onde o mesmo ocorre? O *Vocabulário Ortográfico* registra-as sem hífen.

O problema está em que *gwa'su* (>açu, uçu e guaçu) e *mi'ri* não passam de exemplos de elementos de composição tupis que sincretizam palavras no português. Há grande número de outros formantes com extensa representação em nossa língua, por exemplo, *pinima* “listrado, manchado”, *'pewa* ou *bewa* ‘achatado, liso’, *gwa'yã* ‘caranguejo’, *'taya* ou *'aya* ‘que arde, picante, ácido, azedo’, *pi'xuna* ‘negro, preto, escuro’ e diversos outros. As palavras compostas com tais elementos não são regidas por regras claras quanto ao modo de serem grafadas, o que gera considerável instabilidade ortográfica, que aliás se espelha na própria nominata do *Vocabulário ortográfico*. Eis aqui um pequeno grupo, meramente exemplificativo, de palavras formadas com elementos de composição pospositivos de origem tupi onde tal problema se torna patente: escrevem-se sem hífen (embora sempre o primeiro elemento formador seja oxítono e o segundo, adjetivante): *acaripixuna*, *acarapeba*, *aratupinima*, *acarauçu*, *amborepinima*, *amborepixuna*, *acaraia*, *carapicupeba*, *ipocupinima*, *itupinima*, *acaricuiara* etc. Em *bacabamirim*, entende-se a inutilidade do hífen, mas *baiacumirim*, que não tem hífen, poderia ser lido /baiàcumirim/.

Aliás, um bom exemplo conjunto da instabilidade observa-se em palavras compostas com esse ictiônimo *baiacu*. Escrevem-se sem hífen *baiacuguaima*, *baiacuguima*, *baiacuguimbá*, *baiacumirim*, mas grafam-se com hífen *baiacu-guaiama*, *baiacu-guaimá*, *baiacu-guarajuba*, *baiacu-arará*, *baiacu-ará*. Quanto a *baiacuarara*, o *Vocabulário ortográfico* registra-o com e sem hífen.

Em línguas de grafia estabilizada, o trabalho da lexicografia é reproduzir, na nominata de suas entradas, o adrede estabelecido. Na nossa, em casos como esse, o dicionário tem com frequência de absorver o papel, que não é realmente seu, de normatizador ortográfico, por se ver forçado a tomar um partido dadas as necessidades de padronização. Mesmo nesse caso, porém, a instabilidade ocorre nos registros, pois, não havendo uma regra una, oficial, ocorrem flutuações.

E como se processou o levantamento dos africanismos?

No que respeita aos africanismos – compreendidos nestes os empréstimos de línguas e dialetos africanos integrados no sistema de nossa língua e também vocábulos e expressões dessas línguas e dialetos registrados *in natura* em livros de literatura africana escritos em português – o *Dicionário Houaiss* conta com um levantamento próprio, realizado no Brasil, a partir de glossários apensos a obras literárias de e sobre a África lusofônica. Foram também utilizados alguns dicionários específicos publicados em Portugal e em outros países, como é o caso do *Dictionnaire bilingue portugais-français des particularités*

de la langue portugaise en Guinée-Bissau, do Prof. Jean-Michel Massa.

A flora e a fauna da região foram arroladas e pesquisadas através de manuais locais, entre eles, por exemplo, o excelente *Agronomia angolana*, edição da Repartição Central dos Serviços de Agricultura de Angola, Luanda, n.º 7, 1953 (que é uma publicação póstuma do último trabalho do botânico John Gossweiler). Notável é também *Plantas úteis da flora de S. Tomé e Príncipe – medicinais, industriais e ornamentais*, de Luís Lopes Roseira, de 1984. Isso gerou um número tão expressivo de averbações que nos vimos forçados a guardar material para outras edições ou outras obras, de modo a possibilitar que a massa de informações que constarão do dicionário nesta primeira edição pudesse ser processada pelos botânicos, redatores, datadores e etimologistas do nosso grupo dentro do tempo de trabalho que tinham.

Junto com isso, um grupo de professores africanos de Moçambique, Angola, Guiné-Bissau, São Tomé e Príncipe e Cabo Verde trabalhou em Lisboa, sob a supervisão do professor João Malaca Casteleiro, recolhendo material para outros verbetes regionais a serem introduzidos no *Dicionário Houaiss*.

Houve problemas aqui também com a ortografia?

No caso de africanismos de registro neológico em dicionarização e mesmo de palavras já dicionarizadas, ocorre duplicidade ou mesmo multiplicidade de grafias entre as de cunho aportuguesado e aquelas usadas nos países de África, ocasionando, também aqui, instabilidade de padrão ortográfico no registro. Os países africanos lusófonos, mergulhados em problemas sociais e econômicos, quando não em guerra, não tiveram tempo nem disponibilidades para encetar seus levantamentos lingüísticos e estabelecer dicionários e vocabulários ortográficos com o seu padrão ideal, donde a existência de uma flutuação de soluções, por vezes dentro de um mesmo país, flutuação essa que necessariamente se reflete em qualquer dicionário que pretenda integrar a sua contribuição vocabular sem impor-lhes regras. É o que ocorre com o *Dicionário Houaiss*.

Grande número de línguas e povos africanos da costa, contracosta e centro continental, especialmente os ligados à área de influência da expansão e colonização portuguesas, encontra-se registrado no dicionário com as informações mais atualizadas disponíveis sobre sua caracterização, localização, língua falada e eventualmente os seus traços culturais mais importantes.

E quanto aos asiaticismos?

Não é comum que dicionários elaborados no Brasil preocupem-se em averbar empréstimos de línguas orientais ao português. Basicamente eles são da Índia e de Macau, e em menor escala, por exemplo, de Timor, do *crístang* de Malaca, na Malásia etc. A vocação do *Dicionário Houaiss*, porém, era ir tão longe no registro lusofônico quanto o possível, razão por que, assim como foram registrados regionalismos portugueses e brasileirismos e algumas palavras das línguas e crioulos africanos da área lusófona, fizemo-lo também com

parte do material arrolado por monsenhor Sebastião Rodolfo Dalgado, professor de sânscrito na Faculdade de Letras da Universidade de Lisboa, em seu *Glossário luso-asiático*. Todavia, por ser histórica a maior parte desse registro, preocupamo-nos em completá-lo com informação coeva, viva. Entre as fontes com que contamos para tanto, cito o precioso levantamento do professor Raul Gaião, licenciado em Filosofia que viveu em Macau por nove anos e fez seu mestrado de Lingüística Portuguesa na Universidade de Macau, onde também lecionou português. Tal inventário, por ele mesmo extraído e refeito para nós a partir de sua tese de mestrado, constitui-se num minucioso glossário de palavras e locuções ainda vivas nos falares locais e/ou utilizadas em obras literárias de expressão portuguesa que integram elementos do chinês e do crioulo macaísta.

Registramos também um bom número de substantivos crioulos usados em Timor Leste, respeitando sua grafia.

Que definição genérica, para encerrar, poderia dar-nos do dicionário que dirige?

Cito aqui um verso do poeta português Manuel Alegre que diz: “Na minha língua... cada verso é uma outra geografia”. Com certa liberdade, atrevo-me a afinar tal pensamento com o que buscou Antônio Houaiss em seu dicionário, que sempre foi integrar o labor coletivo de gentes brasileiras e do resto do mundo lusofônico numa nominata comum, em pós da harmonia que as diversas realizações da Língua Portuguesa geram em seu vigoroso e variado conjunto.