

O CONHECIMENTO, A TERMINOLOGIA E O DICIONÁRIO

Maria Tereza Camargo Biderman

A NOMEAÇÃO DA REALIDADE E O LÉXICO DAS LÍNGUAS NATURAIS

É a partir da palavra que as entidades da realidade podem ser identificadas e nomeadas pelos seres humanos. A designação e a nomeação dessas realidades cria um universo significativo revelado pela linguagem.

A atividade de nomear resulta do processo de categorização. Por sua vez, a categorização fundamenta-se na capacidade de discriminação de traços distintivos entre os referentes percebidos ou apreendidos pelo aparato sensitivo e cognitivo do homem. A esse processo segue-se o ato de nomear. Por essa razão a categorização é o processo em que se baseia a semântica de uma língua natural, por meio do qual o homem desenvolveu a capacidade de associar palavras a conceitos.

O léxico de uma língua constitui, portanto, uma forma de registrar o conhecimento do universo. Ao dar nomes às entidades perceptíveis e apreendidas no universo cognoscível, o homem as classifica simultaneamente. Assim, a nomeação da realidade pode ser considerada como a etapa primeira no percurso científico do espírito humano de conhecimento do universo. Ao identificar semelhanças e, inversamente, discriminar os traços distintivos que individualizam estes referentes em entidades distintas, o homem foi estruturando o conhecimento do mundo que o cerca, dando nomes (palavras e termos) a essas entidades discriminadas. É esse processo de nomeação que gerou e gera o léxico das línguas naturais.

Por outro lado, e inversamente, esse processo está indissoluvelmente associado à cultura com que se conjuga uma língua natural. Daí resultam as disparidades vocabulares que opõem, muitas vezes, variedades de uma mesma língua como muito bem ilustram os evidentes contrastes entre o português do Brasil e o português europeu, sobretudo com respeito às terminologias técnico-científicas.

Outra questão importante: a expansão do universo cognoscível na contemporaneidade em virtude da ampliação *ad infinitum* desse universo, propiciada e provocada pelas mudanças contínuas emanadas das ciências e das técnicas. Estamos falando da explosão do léxico nas modernas culturas e civilizações, gerando um *motu continuo* de neologismos designadores de novos conceitos que se vão formando e criando. À medida que fabrica novas realidades, o homem cria novas palavras em um processo incessante. E o léxico vai assumindo dimensões gigantescas sendo praticamente impossível registrá-lo e descrevê-lo por meio de um dicionário.

A DISSEMINAÇÃO DO CONHECIMENTO E A DEMOCRATIZAÇÃO DO SABER Na contemporaneidade o conhecimento adquirido e produzido por alguns membros das diversas comunidades humanas tende a ser divulgado e conhecido por todo o planeta Terra por meio de algumas instituições típicas das sociedades modernas: 1. a mídia escrita, a eletrônica, a cinematográfica; 2. o jornalismo, sobretudo o científico; 3. e a escola, entendendo-

se por esta palavra toda e qualquer instituição dedicada à transmissão do conhecimento de maneira organizada, sistemática e institucionalizada. Assim, tais veículos garantem a disseminação do conhecimento junto a um público cada vez maior, levando juntamente com o conhecimento das realidades o vocabulário que as designa. Vamos exemplificar com duas áreas diretamente relacionadas com a vida dos seres humanos em nossos dias: a área das ciências biomédicas e a economia.

O VOCABULÁRIO DAS CIÊNCIAS DA SAÚDE

Na sua convivência freqüente com esse mundo o público leigo vem integrando em seu vocabulário muitos termos técnico-científicos que designam especialidades médicas, tipos de exames laboratoriais e de diagnóstico, técnicas de tratamento, drogas e medicamentos, recursos vários usados na cura das doenças, e assim por diante. Vamos ilustrar com alguns exemplos:

doenças e manifestações mórbidas da saúde: carcinoma, cisto, dengue, distúrbio neurológico, enfisema pulmonar, epidemiologia, gastroenterite/gastroenterite, glaucoma, hepatite, herpes, hipertensão arterial, hipoglicemia/hiperglicemia, hipertireoidismo, isquemia, isquemia cerebral, leishmaniose, mal degenerativo, neurose, osteoporose, patologia, paranóia, pólipos, sarcoma, septicemia, síndrome, síndrome do pânico, soropositivo, síndrome de Alzheimer, síndrome de Down;

técnicas, exames e tratamentos: análise do DNA, biologia molecular, biópsia/biopsy, densidade óssea, diagnóstico por imagem, exame parasitológico, gastrectomia, hemograma, informação genética, psicoterapia, laparoscopia, mamografia, neurocirurgia, raio X, ultra-sonografia;

profissionais da saúde e especialidades médicas: cardiologia, cardiologista, citologia, dermatologia, dermatologista, endocrinologia, endocrinologista, fisioterapia, fisioterapeuta, gastroenterologia, geriatra, geriatria, ginecologia, ginecologista, hematologia, imunologia,

microbiologia, neurologia, neurologista, oftalmologia, oftalmólogo, ortopedia, ortopedista, otorrinolaringologia, otorrinolaringologista, parasitologia, proctologia, pneumologia, toxicologia, urologia, urologista.

O VOCABULÁRIO DA ECONOMIA E DAS FINANÇAS

Nesse domínio terminológico muitos são os termos do complexo bancário e financeiro, das finanças públicas, do mercado *tout court* e do mercado de capitais com que os cidadãos vêm convivendo nos dias atuais e que incorporaram a seu vocabulário por viverem em uma sociedade capitalista dominada pelo mercado. Vejamos alguns exemplos dessa imensa pletora de termos: ágio, alíquota, amortização, apólice, auditoria, balança comercial, bens de consumo, capital de giro, cartão de crédito, cartel, cheque especial, contribuinte, data-base, déficit, deságio, dívida pública, dividendo, duplicata, endossar, endosso, *factoring*, fluxo de caixa, globalização, *holding*, imposto de renda, imposto territorial urbano (IPTU), imposto sobre a propriedade de veículos automotores (IPVA), inadimplência, *joint venture*, *leasing*, licitação, margem de lucro, monopólio, nota promissória, oligopólio, passivo, patrimônio, pessoa física, pessoa jurídica, preço de mercado, privatização, risco-Brasil, saldo médio, *spread*, superávit, taxa de câmbio, valor de mercado, valor venal, venda a prestação, etc.

**ESTAMOS
FALANDO DA
EXPLOÇÃO DO
LÉXICO NAS
MODERNAS
CULTURAS E
CIVILIZAÇÕES ...**

A CRIAÇÃO DE NOVOS TERMOS NA LÍNGUA Um problema teórico de grande relevância relativamente aos vocabulários científicos e técnicos diz respeito ao processo de criação das novas unidades léxicas que, como se disse, é fato que ocorre com enorme freqüência nesses domínios. Ora, sucede que as linguagens de especialidade geram novos termos com base no acervo que a língua já possui. De um lado, reutiliza palavras já existentes para criar outras, ou então, serve-se dos processos de formação de palavras que existem no sistema do português. Assim, por exemplo, os processos de derivação e composição são continuamente solicitados para gerar neologismos que se fazem necessários. Mas a língua se serve também, não raro, de empréstimos de outras línguas, sendo o inglês, na atualidade, o idioma que maior número de termos fornece ao português. Ver exemplos mencionados acima: *holding, leasing, spread* na economia. A derivação é acionada freqüentemente sendo que, nesse caso, a prefixação ocorre comumente. Aqui tem imensa importância o uso de prefixos ou elementos de composição gregos e latinos, como por exemplo: **aero-**, **agro-**, **anti-**, **arqui-**, **auto-**, **bio-**, **cine-**, **demo-**, **electro-**, **endo-**, **epi-**, **extra-**, **fisio-**, **fono-**, **foto-**, **geo-**, **hetero-**, **hidro-**, **hiper-**, **hipo-**, **inter-**, **macro-**, **maxi-**, **mega-**, **micro-**, **mini-**, **mono-**, **neuro-**, **multi-**, **pluri-**, **poli-**, **proto-**, **pseudo-**, **radio-**, **retro-**, **semi-**, **sub-**, **super-**, **supra-**, **tele-**, **ultra-**, etc. É verdade que o neologismo científico formado a partir de um desses formantes muitas vezes não foi gerado no português, mas em outro idioma (o inglês, o francês) e nossa língua adotou esse vocábulo, adaptando-o às características morfosintáticas do português, podendo-se considerar que se trata de empréstimo. De fato, o vocabulário técnico-científico tem caráter universal e, para todas as línguas, o grego e o latim constituem a fonte lexical onde todos vão beber na revitalização contínua de seus léxicos.

Mas também o processo de composição tem enorme vitalidade na geração dos vocabulários científicos e técnicos. Exemplos: **cardiovascular**, **aminoácido**, **carboidrato**, **ácidos graxos**, **ação oxidante**, **distúrbio neurológico**, **enfisema pulmonar**, **dermatologia**, **dermatologista**, **endocrinologia**, **endocrinologista**, **fisioterapia**, **fisioterapeuta**, **neurotransmissor**, **radical livre**, **síndrome do pânico**, **síndrome de Down**, **balança comercial**, **balanço de pagamentos**, **bens de consumo**, **capital de giro**, **cartão de crédito**, **imposto de renda**, **imposto predial ou imposto territorial urbano (IPTU)**, **margem de lucro**, **pessoa física**, **pessoa jurídica**, **preço de mercado**, **risco-Brasil**, **saldo médio**, etc.

Por outro lado, ao servir-se de seu patrimônio lexical, muitas vezes o português procede a uma ressemantização de unidades já comuns na língua. A reutilização do vocabulário já existente na língua, atribuindo-lhe novos sentidos, é um recurso usado freqüentemente nas terminologias. Esse procedimento evita a criação contínua de novas unidades até então inexistentes, fazendo-se economia de novas formas, o que expandiria *ad infinitum* o número de unidades diferentes do léxico. Assim, com base em traços semânticos que a significação nuclear da palavra já possui, ampliam-se seus valores de sentido para incluir novas denotações. É o que ocorreu recentemente com o vocabulário da informática. Vários termos foram criados a partir de outros já existentes no português. Contudo, a esses vocábulos foram agregadas características semânticas novas como por exemplo, ocorreu com **arquivo**, **barra**, **botão**, **carregar**, **comando**, **disco**, **janela**, **lixeira**, **memória**, **placa**, **porta**, **programa**, **rede**, **rodar**, **salvar**, **tecla**, **teclado**, **vírus**.

O REGISTRO DO Vocabulário Técnico-Científico em Dicionários Gerais da Língua Os dicionários constituem uma organização sistemática do léxico, por meio da qual os lexicógrafos tentam descrever o vocabulário dessa língua acumulado ao longo dos séculos. Assim, essas obras recolhem o patrimônio léxico da língua num dado momento da história da comunidade, visando descrever e documentar esse tesouro lexical que a tradição foi armazenando. De fato, esse ideal é intangível, já que o léxico cresce em progressão geométrica e esse processo nunca cessa enquanto a língua for viva. Até mesmo os maiores tesouros lexicográficos já compilados, como a edição de 1983 do *Webster* (500 mil verbetes) nem de longe registrou todo o acervo lexical do inglês naquela data.

No caso da língua portuguesa, nossos dicionários gerais são mais modestos. Nosso dicionário geral de maior porte é a 10ª edição do *Dicionário de Moraes*, reeditado por J.P. Machado para a Editorial Confluência em 12 volumes (1949-1959), que inclui uma nomenclatura de 306.949 verbetes. Para que se tenha um parâmetro comparativo da expansão do vocabulário português ao longo do tempo e da história da língua, vamos comparar essa 10ª edição com a edição considerada fundadora desse Dicionário de Antonio Moraes e Silva, a 2ª edição de 1813. Nessa edição de 1813, Moraes registrou 40 mil vocábulos (verbetes) tendo abarcado a língua dos séculos XVI, XVII e XVIII. As obras que lhe serviram de fonte de referência são desses três séculos. Examinando o vocabulário técnico-científico desse dicionário e comparando-o com o de nossos dias, podemos dar-nos conta da imensa evolução da terminologia científica.

Na contemporaneidade são dicionários gerais: o *Novo Aurélio - Dicionário da língua portuguesa - Século XXI* (1999) registrando uns 135 mil verbetes, o *Dicionário Houaiss da língua portuguesa* (2001), cerca de 228.500 verbetes e o *Dicionário Michaelis de português*, cerca de 200 mil verbetes. Ora, seja qual for a dimensão de um dicionário geral, o fato é que a imensa maioria desse acervo é constituída pela terminologia técnico-científica. O vocabulário comumente usado na língua não ultrapassa 10 mil unidades vocabulares, embora essas palavras sejam usadas com uma altíssima freqüência, além de serem reutilizadas continuamente com novas conotações, gerando o fenômeno da ressemantização acima referido, e gerando também o fenômeno da polissemia pelo qual os valores semânticos dos vocábulos se multiplicam. Inversamente, o vocabulário técnico-científico tem freqüência de uso muito baixa na língua como um todo, sendo usado apenas no âmbito de cada língua de especialidade pelos profissionais dessa área como é o caso dos dois domínios que comentamos acima: o das ciências biomédicas e o da economia. Entretanto, em face do fenômeno da democratização do saber, uma parcela desse vocabulário ultrapassa as fronteiras do domínio especializado e ingressa na grande corrente do uso geral.

Portanto, dado o número gigantesco de termos técnicos e científicos, e considerando-se o escasso emprego de muitos deles, o dicionarista terá uma tarefa complexa na seleção das unidades que integrarão a nomenclatura do dicionário geral. Uma vez que é impossível registrar tudo, ou quase tudo, e levando em consideração as limitações de uma obra impressa, quais critérios utilizará o lexicógrafo para selecionar a nomenclatura do dicionário? Na realidade mesmo os dicionários mais bem elaborados que utilizaram critérios lingüísticos e léxico-estatísticos não operaram de modo rigorosamente científico. Basta ler o prefácio do dicionário do francês *Petit Robert*, escrito pelo grande lexicógrafo e especialista nas ciências do léxico, Alan Rey. Os melhores dicionários da língua inglesa, tipo Oxford e Cambridge, também não

justificam de modo totalmente satisfatório suas escolhas. Rey explica que selecionou apenas os termos vulgarizados na língua comum, o que não esclarece muito. O acervo do *Petit Robert* é de 70 mil entradas. Aliás, não se trata de um dicionário geral como estamos aqui considerando, mas de um dicionário de porte relativamente grande para o uso do consulente letrado. Em qualquer hipótese, como proceder? Onde recolher esse vocabulário geral da língua inclusive um número suficientemente grande de termos técnico-científicos para atender às necessidades de um eventual consulente letrado, mas não especialista? A resposta é: partir de um gigantesco *corpus* de textos informatizados da língua, textos esses que sejam representativos de todos os gêneros, que incluam as duas modalidades da língua (falada e escrita) e que abranjam todos as áreas do conhecimento. Seria factível? Relativamente ao número de áreas técnico-científicas, não se sabe exatamente que número seria esse; talvez totalize uns 250, 300 ou mais domínios. Numa tentativa que estava sendo levada a efeito pela Editora Longman de Londres no início dos anos noventa, a partir do *corpus* do *Webster*, a equipe de especialistas elencou 200 áreas do conhecimento para trabalhar com elas, formulando a hipótese de que cobririam quase a totalidade das áreas do conhecimento. A evolução da ciência de lá para cá ampliou o número de domínios científicos, criando muitas especialidades novas. Parece-me impossível responder a essa questão.

Por outro lado, infelizmente, a confecção de dicionários gerais da língua é planejada e executada por editoras comerciais que não estão dispostas a investir milhões em pesquisa para investigar a complexidade do léxico geral de uma língua e produzir um dicionário segundo critérios que poderíamos chamar de científicos.

Atualmente existem dois *corpora* de grandes dimensões do português que foram criados com objetivos lexicográficos – um brasileiro e um europeu. O europeu *Corpus de Referência do Português Contemporâneo*, (CRPC), da Universidade de Lisboa tem cerca de 201,5 milhões de ocorrências. O brasileiro da FCL da Unesp, campus de Araraquara com 200 milhões de ocorrências de palavras. A partir do *corpus* do CRPC foi elaborado o *Dicionário da língua portuguesa contemporânea* (2001) da *Academia de Ciências de Lisboa*, contendo 70.000 verbetes. A partir do *corpus* de Araraquara foi elaborado o *Dicionário de usos* com 60 mil verbetes (Ática, 2002). Portanto, nenhum desses dois dicionários tem dimensões de um dicionário geral, embora ambos tenham a virtude de se terem baseado em um *corpus* de textos efetivamente produzidos, logo com usos reais da língua.

Um *corpus* de referência de grandes dimensões (digamos uns 200 milhões de ocorrências) com as características acima referidas, pode embasar o projeto de produção de um dicionário geral da língua. De fato, algumas línguas como o francês e o inglês já possuem corpora com o dobro dessas dimensões, ou seja, 400 milhões de palavras. No caso do francês, a partir desse *corpus*, produziu-se um dicionário monumental *Dictionnaire de la langue française des XIX^{ème}. e XX^{ème}. siècles* (1969-1994) em 16 volumes, embora contendo apenas 80 mil verbetes.

Um *corpus* ideal de uns 200 milhões de palavras, desenhado com a arquitetura acima referida, teria condições de embasar um dicionário geral do português. Dele seria extraída a **nomenclatura** (lista das palavras-entrada) usando-se critérios lexico-estatísticos e evitando-se soluções aleatórias. Programas adequados extrairiam do *corpus* as concordâncias de texto dos lemas dessa nomenclatura. As concordâncias de texto permitiriam identificar os significados (conceitos) e os usos dos vocábulos selecionados para a nomen-

clatura do dicionário, que é a espinha dorsal do dicionário. E haveria ainda todas as outras enormes complexidades postas pelas redes de significação em que se encadeiam os significados e conceitos no interior do léxico. Aqui dificilmente o computador poderia ajudar o lexicógrafo. Nessa etapa da elaboração do dicionário faz-se mister o engenho do dicionarista e um profundo conhecimento do idioma.

Infelizmente, creio que o Brasil ainda não possui esse tipo de obra, mas já tem condições de produzir tal dicionário.

Maria Tereza Camargo Biderman é lingüista e lexicógrafa, professora da Unesp é autora de dicionários: Dicionário ilustrado de português, Ática, 2005, Dicionário didático de português, Ática, 1998 e Dicionário do estudante, Globo, 2005.

BIBLIOGRAFIA CONSULTADA

Biderman, M.T.C. *Teoria lingüística*. São Paulo, Editora Martins Fontes, 2001.

Biderman, M.T.C. "Dimensões da palavra". In *Filologia e língua portuguesa*, S. Paulo, Humanitas Publicações/FFLCH/USP, nº 2, 1998, 81-118.

Cuenca, M. J. & Hilferty, J. *Introducción a la lingüística cognitiva*. Barcelona, Editorial Ariel, S.A., 1999.

Lenneberg, E.H. *Fundamentos biológicos del lenguaje*. Madrid, Alianza Universidad, 1975.

Robert, P. *Le Petit Robert*. Paris, Les dictionnaires-Robert, 1990.