

## **Lexicografia no início do século XXI** **– novas perspectivas, novos recursos e suas consequências**

Margarita Correia

Instituto de Linguística Teórica e Computacional / Faculdade de Letras da Universidade de Lisboa

margarita.correia@iltec.pt

**Keywords:** Lexicography; learning dictionary; general language dictionary.

### **Abstract**

Contemporary general monolingual dictionaries are much different from their ancestors: these were almost exclusively concerned with prescription, with the representation of the “correct” usages of lexical units, while present-day dictionaries are more concerned with the description of the effective usage of words. This fact is not always well understood by dictionary users, still acquainted with ancient models of the dictionary. This paper presents the main causes which gave rise to this change of paradigm, focusing recent approaches to lexical studies, *corpus* linguistics and the development of efficient databases, as well as the development of a learning lexicography. Main changes in the body of the dictionary will be referred, and some discussion about the role of dictionaries in contemporary Portuguese society, and about their prescriptive or descriptive character will be held.

### **1. Introdução**

No século XX, a linguística constitui-se como uma ciência e os dicionários, como repositórios de saber lexical, não poderiam ficar alheios a esse facto.

Nos últimos anos, tornou-se cada vez mais clara a noção de que o léxico de uma língua é um conjunto virtual de unidades lexicais e que, como tal, qualquer trabalho lexicográfico mais não é do que a descrição de um determinado vocabulário. Ora, sempre que a língua em observação é uma língua viva e em evolução, a selecção de palavras a incluir nesse vocabulário, assim como a informação a incluir sobre cada unidade lexical, acabam por constituir alguns dos tópicos mais interessantes e determinantes na realização de qualquer dicionário. A par desta concepção do léxico, a montante e a jusante dela, encontram-se os *corpora* linguísticos computadorizados, assim como as ferramentas

computacionais concebidas para o armazenamento e a exploração de dados lexicais. A linguística de *corpus* mudou em muito a forma de fazer linguística e, particularmente, a teoria e a técnica lexicográficas.

Com este trabalho pretende-se mostrar o impacto e as consequências que as novas concepções do léxico e da língua, por um lado, e os recursos computacionais disponíveis, por outro, têm para a lexicografia monolíngue contemporânea. Em suma, procurarei demonstrar como se passou de uma lexicografia geral de cariz mais prescritivo para uma lexicografia essencialmente descritiva e comunicativa. Referirei, por fim, algumas das implicações desta mudança no uso que o falante comum faz do seu dicionário monolíngue.

Neste trabalho, ocupar-me-ei apenas de dicionários monolíngues destinados a serem usados por humanos, deixando de lado o desenvolvimento da lexicografia bilingue e da lexicografia computacional.

## **2. Novas perspectivas: as concepções da língua e do léxico**

Com o dealbar da linguística como ciência, no início do século XX, graças sobretudo aos ensinamentos de Saussure e seus seguidores, a língua passou a ser encarada como um objecto de estudo por si própria e a ser alvo de um discurso que, como qualquer discurso científico, se quer descritivo, objectivo, distinto, portanto, do gramaticismo prescritivo predominante ao longo da história. Também o dicionário, repositório de saber lexical de uma língua, tem vindo a incorporar as realizações da linguística e o seu discurso científico próprio, predominantemente descritivo, tendo-se transformado, sobretudo ao longo da segunda metade do século passado, num objecto distinto daquele que conhecíamos e que consistia num instrumento por excelência de normalização linguística. Esta mudança de carácter do dicionário nem sempre tem sido entendida pela sociedade, alheia aos desenvolvimentos da própria linguística e da lexicografia.<sup>1</sup>

Ao longo do século XX, mesmo do ponto de vista científico, a concepção de língua foi-se enriquecendo. De uma concepção de língua como fenómeno social, introduzida por Saussure e seguida por grande parte dos linguistas da primeira metade do século, passou a ver-se a língua também como facto mental, graças, sobretudo, aos trabalhos de Chomsky e seus seguidores, já na segunda metade do século.

Além disso, os estudos em psicolinguística ao longo dos últimos 50 anos proporcionaram-nos um entendimento do modo como adquirimos a língua, materna ou segunda, e do modo como processamos o conhecimento linguístico a nível mental, tendo tido, genericamente, uma importância decisiva para a confirmação das propostas teóricas de Chomsky relativamente ao bioprograma humano para a linguagem. Actualmente as neurociências trazem-nos cada vez mais evidências sobre o modo como a linguagem é processada a nível cerebral.

A partir dos anos 60 do século XX, os trabalhos de Labov trouxeram, ainda, para a ordem do dia a questão da relação íntima entre a língua e a sociedade, tendo-se em conta, nos estudos linguísticos, parâmetros de variação linguística tais como a classe social, a geografia, o sexo ou a situação política dos Estados. Os trabalhos da sociolinguística levaram a que se se explicassem e se tomassem como igualmente válidas e dignas de estudo variedades linguísticas pouco reconhecidas socialmente.<sup>2</sup>

Ao nível dos estudos do léxico, esta mudança teve consequências, como seria de esperar. Uma das principais consequências é a concepção clara de que o léxico de uma língua é um conjunto apenas virtual de unidades lexicais e, portanto, impossível de descrever em extensão. A lexicologia ocupa-se de descrever a estrutura do léxico, o modo como se organiza, as regularidades apreensíveis no léxico, que têm vindo progressivamente a tornar-se mais claras para o observador: o léxico já não é hoje encarado meramente como o repósitório das unidades lexicais e suas respectivas idiossincrasias, mas antes como uma componente da gramática que, apesar das suas particularidades (tais como o facto de ser uma componente aberta e em expansão, de limites imprecisos, abrangendo todo o universo conceptual de uma língua<sup>3</sup>), apresenta as suas regularidades próprias e uma forma de estruturação específicas. Trabalhos extensivos em Morfologia, no quadro de modelos associativos como o de Danielle Corbin, contribuíram, também, para a percepção crescente das regularidades internas da componente lexical.<sup>4</sup>

O léxico, como parte de uma língua entendida como facto social, foi estudado durante anos sobretudo ao nível do que efectivamente ocorre na língua, sobretudo no discurso formal escrito e na literatura,<sup>5</sup> como de resto acontece nas chamadas gramáticas tradicionais, de carácter prescritivo. Com a mudança de paradigma, os estudos lexicais passaram também a incorporar a tentativa de entender o modo como o conhecimento lexical é

estruturado, adquirido e processado a nível mental e a linguística passou a ver o léxico como uma componente específica da gramática. Uma das consequências da concepção de léxico como módulo gramatical foi o reforço do entendimento das relações estreitas entre léxico e as restantes componentes da gramática e da necessidade de descrever a própria gramática das unidades lexicais como única forma de ter uma melhor apreensão do que é o conhecimento lexical. Tem-se hoje uma percepção mais clara dos diversos tipos de conhecimento associados às unidades lexicais, que vão da sua forma (oral e escrita), à sua estrutura interna, flexão e às relações com palavras da mesma família, aos seus significados e, no caso das unidades de significado lexical, à sua capacidade denominativa, às relações com outras unidades no plano sintáctico, às suas combinatórias, passando pelas relações semânticas que estabelecem com outras palavras. Estes conhecimentos viram-se reforçados também pelos dados da linguística de *corpus*, como bem realça Sinclair a propósito dos estudos sobre as formas inglesas *yield* e *decline* e sobre os “phrasal verbs”.<sup>6</sup>

Em paralelo, é ao nível da psicolinguística que encontramos os primeiros grandes avanços no sentido de compreender a natureza, a aquisição e o processamento do saber lexical. O léxico mental passa a ser alvo de estudos sistemáticos e é hoje praticamente consensual o facto de que, na estruturação deste léxico mental, são absolutamente essenciais as relações que as palavras estabelecem entre si<sup>7</sup> e que podem ser de natureza diversa – formal (homofonia, homografia, homonímia, paronímia), morfológica (palavras da mesma família), semântico-conceptual (sinonímia, oposição, hiponímia / hiperonímia, meronímia / holonímia e, sobretudo mais recentemente, em parte graças às propostas de Pustejovski, outro tipo de relações como causa-efeito, matéria-produto, etc.).<sup>8</sup> O léxico passou, então, a ser entendido como uma rede de itens lexicais (daí o conceito de *wordnet*), ligados por nexos semânticos e conceptuais, indispensáveis à estruturação do léxico.

### **3. Novos recursos: o desenvolvimento da informática**

Ao longo da segunda metade do século XX, o progresso da micro-informática desencadeou, de acordo com linguistas como Sardinha, “uma verdadeira revolução no pensamento linguístico, com implicações sérias sobre como respondemos a questões fundamentais, tais como é a língua, como ela é organizada, como deve ser estudada, como deve ser ensinada. A mola propulsora dessa revolução é a tecnologia, mais

especificamente o computador.”<sup>9</sup> O desenvolvimento das capacidades de armazenamento e de gestão de grandes quantidades de informação, assim como as crescentes capacidades de pesquisa (semi-)automatizada cada vez mais avançadas, tiveram consequências directas para a linguística e, conseqüentemente, para o trabalho lexicográfico. No âmbito deste trabalho, iremos deter-nos em duas linhas de desenvolvimento fundamentais: o desenvolvimento de *corpora* e o incremento do armazenamento de informação em bases de dados.

#### *As fontes: os corpora*

*Corpus* linguístico não é um conceito novo em linguística, embora alguns muitas vezes o esqueçam). Desde sempre, os linguistas sentiram necessidade de coligir dados linguísticos que servissem de ponto de partida para as suas pesquisas, numa perspectiva empírico-indutiva, ou que pudessem confirmar as suas hipóteses de trabalho, quando adoptavam uma perspectiva hipotético-dedutiva. Dubois *et al.* incluem a entrada “*corpus*” no seu dicionário, explicitando que este é constituído pelos dados (passagens, frases, conjuntos de frases, palavras, ...), tomados de um universo, que serão submetidos a análise linguística.<sup>10</sup>

Quando, porém, se fala em *corpus* actualmente, pensa-se em *corpus* computadorizado. Ainda que sejam diversas as suas actuais definições, podemos definir *corpus*, no âmbito deste trabalho, seguindo Sardinha que, por seu turno, cita Sanchez e Cantos, como “Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum dos seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.”<sup>11</sup>

Desde 1964, quando foi lançado o primeiro *corpus* linguístico electrónico, o Brown University Standard Corpus of Present-day American English, que continha um milhão de ocorrências, quantidade notável para a época,<sup>12</sup> até aos nossos dias, as capacidades de armazenamento de dados têm possibilitado a compilação de *corpora* progressivamente mais vastos e mais representativos da língua. Tal facto, tem contribuído decisivamente para o desenvolvimento dos estudos linguísticos, dando origem àquilo que hoje alguns

Correia, Margarita 2008. Lexicografia no início do século XXI – novas perspectivas, novos recursos e suas consequências. In: Júnior, Manuel Alexandre (coord.) *Lexicon – Dicionário de Grego-Português, Actas de Colóquio*. Lisboa: Centro de estudos Clássicos / FLUL, pp. 73-85. (versão entregue para publicação)

defendem ser uma disciplina específica – a linguística de *corpus* – e, indiscutivelmente, para o desenvolvimento da lexicografia.

Para a língua portuguesa, destaquem-se três *corpora* com impacto na lexicografia desta língua: a) o *Corpus* de Referência do Português Contemporâneo, desenvolvido no Centro de Linguística da Universidade de Lisboa, sob a coordenação de Maria Fernanda Bacelar do Nascimento, do qual dados foram usados na confecção do *Dicionário da Língua Portuguesa Contemporânea*; b) o *corpus* do Laboratório de Lexicografia da Faculdade de Ciências e Letras de Araraquara (UNESP), sob a coordenação de Francisco da Silva Borba, que serviu de base à elaboração do *Dicionário de Usos do Português do Brasil*; c) o *corpus* do projecto *Dicionário Histórico do Português do Brasil – séculos XVI a XIX*, ainda em desenvolvimento, sob a coordenação de Maria Tereza Biderman, também da UNESP de Araraquara.

A compilação de vastos *corpora* textuais como base para o trabalho lexicográfico permitiu o desenvolvimento de toda uma lexicografia baseada em *corpus*, que, de acordo com Biber, Conrad e Reppen, permite abordar seis questões fundamentais para a lexicografia: 1) Quais os significados associados a uma determinada palavra?; 2) Qual a frequência de uma palavra relativamente a outras palavras relacionadas?; 3) Que padrões de associação não-linguísticos tem uma determinada palavra (ex.: registos, períodos históricos, dialectos)?; 4) Que palavras co-ocorrem vulgarmente com uma palavra particular e qual a distribuição dessas sequências combinatórias (“collocational”) ao longo dos diferentes registos?; 5) Como se distribuem os sentidos e usos de uma determinada palavra?; 6) Como são usadas e qual a distribuição de palavras aparentemente sinónimas?<sup>13</sup>

A maioria dos estudos que se têm desenvolvido nos últimos anos sobre combinatórias (entendidas aqui em sentido amplo como associações frequentes de palavras, que vão desde os compostos sintágmáticos às fraseologias que admitem maior ou menor grau de variação interna) têm sido desenvolvidos com base na análise de *corpora*. Muita da lexicografia que hoje se pratica a nível internacional é uma lexicografia baseada em *corpus*, o que tem permitido uma descrição mais pormenorizada e próxima do uso efectivo das unidades lexicais em situações concretas de uso.

*O armazenamento: o incremento da memória dos computadores e o desenvolvimento das bases de dados*

O desenvolvimento da informática abriu, ainda, a possibilidade de armazenar a informação em bases de dados digitais, não apenas tornando possível armazenar de modo estruturado quantidades de informação antes inimagináveis, como, sobretudo, de levar a cabo pesquisas rápidas e eficientes nessas bases de dados, agilizando significativamente o trabalho do lexicógrafo. Hoje em dia, o recurso a bases de dados torna muito mais fácil obedecer, por exemplo, a critérios de sistematicidade da obra lexicográfica.

Além disso, as bases de dados apresentam uma característica não negligenciável, que é a reusabilidade dos dados, isto é, a possibilidade de reaproveitar, corrigir e/ou adicionar constantemente os dados contidos nas bases de dados, dando azo à produção, em tempo reduzido, de edições específicas de dicionários com diferentes tamanhos, formatos e visando públicos-alvo distintos (ex.: uma editora de dicionários como a casa Le Robert, em Paris, pôde com relativa facilidade, a partir da sua base de dados, produzir, o *Petit Robert*, o *Grand Robert*, o *Robert Méthodique*, o *Robert Junior*, o *Robert Micro*, o *Robert de Poche*, o *Robert pour tous*, etc., para falarmos apenas dos dicionários monolíngues desta editora). A reusabilidade dos dados permite, ainda, a produção frequente de edições revistas e actualizadas do mesmo dicionário, como tem sido o caso, nos últimos anos, do dicionário geral monolíngue da Porto Editora.

Ainda que a produção de um dicionário de raiz seja um projecto extremamente caro e moroso (daí que tantos dicionários existentes no mercado mais não sejam do que antigos dicionários com novas roupagens), a possibilidade de reusar os dados do modo descrito faz com que um projecto deste tipo possa hoje ser encarado como um investimento a longo prazo e que pode ser constantemente rentabilizado. É de lamentar, portanto, que em projectos de dicionário com estas características, produzidos em contexto não-empresarial, esta filosofia não tenha ainda sido incorporada e o investimento feito para a produção de um dicionário não seja rentabilizado para a produção de outros, corrigidos, mais actuais e melhores.

#### **4. Breve panorâmica da lexicografia portuguesa no início do século XXI**

Ao desenvolvimento da lexicografia nos últimos anos não foi indiferente o desenvolvimento do ensino de línguas estrangeiras. Mais do que os grandes projectos de lexicografia baseada em *corpora* iniciados na segunda metade do século XX, de que o projecto do *Trésor de La Langue Française* (1971-1994) é significativo exemplo, aquilo que mais caracteriza a lexicografia moderna é o desenvolvimento de uma quantidade significativa de dicionários monolíngues destinados à aprendizagem de grandes línguas internacionais, particularmente da língua inglesa.<sup>14</sup> Trata-se de dicionários de pendor predominantemente descritivo (e não normativo), dando conta do uso real que é feito da língua, visando a descrição do vocabulário da chamada língua corrente, fornecendo informação pertinente não apenas para a função de compreensão linguística, mas também, e de forma significativa, informação pertinente para a função de produção e para o alargamento do vocabulário do consulente (informação gramatical, relativa a relações entre palavras, etc.). São exemplos de dicionários deste tipo:

- o *Collins COBUILD English Dictionary* (1.<sup>a</sup> ed.: 1987) e o *Longman Dictionary of Contemporary English* (1.<sup>a</sup> ed.: 1978), para o Inglês,
- o *Dictionnaire du Français Contemporain* (1966) e o *Lexis* (1.<sup>a</sup> ed.: 1975), para o Francês;
- o *Diccionario de Uso del Español* (1.<sup>a</sup> ed. 1967), de María Moliner, para o Espanhol;

Para a língua portuguesa, o único dicionário inteiramente baseado em *corpus* de que temos conhecimento é o *Dicionário de Usos do Português do Brasil*, de Francisco da Silva Borba, com primeira edição em 2002. Dicionários modernos, em termos de concepção e metodologia destinados à aprendizagem da língua portuguesa, são o *Dicionário do Português Básico*, de Mário Vilela (Porto, 1990), o *Dicionário Contemporâneo do Português*, de Maria Tereza Biderman (São Paulo; 1992; reeditado em 1998, com o título *Dicionário Didático de Português*) e o *Dicionário da Língua Portuguesa – Verbo* (Lisboa, 2006), embora nalguns casos não sejam completamente explícitos quanto ao uso que fazem de *corpora* textuais para a sua elaboração ou, noutros, não os tenham mesmo usado. Como pode verificar-se, a lexicografia baseada em *corpus* e a lexicografia de aprendizagem levaram algum tempo a desenvolver-se em Portugal e no Brasil.

Nos primeiros anos do presente século, Portugal conheceu a publicação de três dicionários gerais de língua que enriqueceram decisivamente a oferta lexicográfica do nosso país: o *Dicionário da Língua Portuguesa Contemporânea* (DLPC), em 2001, a versão *Portuguesa do Dicionário Houaiss da Língua Portuguesa*, em 2002-2003, e o *Grande Dicionário da Língua Portuguesa*,<sup>15</sup> em 2004. Tendo sido elaborado de raiz, o DLPC, com todos os seus erros, as suas faltas de sistematicidade e as escolhas discutíveis que apresenta, é aquele que, como projecto de dicionário, mais se aproxima daquilo que pode ser considerado um dicionário concebido de acordo com os recentes avanços da lexicografia: ele incorpora avanços da lexicografia de aprendizagem desenvolvida nos últimos anos. Assumindo-se como um dicionário geral de língua, ele funciona, também, como um dicionário apto para a compreensão e a produção em português europeu, incorporando um acervo importantíssimo de informação gramatical sobre cada uma das entradas que descreve e uma micro-estrutura rica e bem estruturada, contemplando uma quantidade de remissões que possibilitam o enriquecimento vocabular dos seus consulentes.

## **5. Consequências para a lexicografia contemporânea**

Os dicionários que hoje se publicam não poderiam ser indiferentes a todas as mudanças de perspectiva e ao surgimento de novos recursos descritos anteriormente. Os dicionários contemporâneos são necessariamente diferentes dos que conhecemos no passado, tendo incorporado os efeitos dessas mudanças, embora nem sempre com os efeitos mais desejados, nem contando com a aceitação do público geral. Nas linhas que se seguem, referiremos algumas das características dos dicionários gerais de língua modernos que os distinguem dos seus antepassados.

A primeira grande mudança do dicionário geral de língua tem a ver a sua representatividade. Numa lexicografia baseada em *corpora*, as palavras não podem ser inseridas na nomenclatura do dicionário apenas com base na intuição do lexicógrafo, mas sim com base em critérios rigorosos que vão da frequência de ocorrência de um item no *corpus*, à sua representatividade no tipo de discurso cujo vocabulário se pretende retratar. Deste modo e dado que o léxico é hoje claramente entendido como sendo impossível de descrever em extensão, um dicionário contemporâneo, por muito extenso que seja, não pode arrogar-se a capacidade de representar todo o léxico da língua, como no passado,

quando implícita ou explicitamente tinha essa pretensão.<sup>16</sup> Os bons dicionários de hoje são representativos de fatias bem delimitadas do léxico de uma língua, de vocabulários claramente delimitados em função de critérios como a frequência de ocorrência das palavras e o seu interesse para o público-alvo visado pelo dicionário.

Outra das consequências do uso de *corpora* prende-se com o tipo de exemplificação que é feita do uso das unidades lexicais. Não pretendemos afirmar, neste trabalho, que até à actualidade não tenham sido usados *corpora*, no sentido mais lato deste termo, para a confecção de dicionários: já no seu *Vocabulario Portuguez e Latino* (1712-1718), Raphael Bluteau destaca a importância de um *corpus* de referência formado por autores portugueses para a confecção do dicionário.<sup>17</sup> Também a descrição lexicográfica de grandes projectos clássicos de dicionários de língua inglesa como o *Dictionary of English Language*, de Samuel Johnson (1755) e o *Oxford English Dictionary* (elaborado a partir dos anos 60 do século XIX) se basearam na compilação de vastos conjuntos de excertos de autores literários.<sup>18</sup> A diferença entre os dicionários modernos e os seus antepassados é que as fontes usadas não se limitam a ser textos de escritores consagrados pela sociedade, mas são constituídas por textos representativos dos mais diversos registos linguísticos (formal e informal, escrito e oral, científico, jornalístico, pedagógico, etc.). Como consequência, os dicionários modernos não apresentam apenas abonações (excertos retirados de obras do cânone literário destinados a abonar, a validar, a indicar o “bom uso” das palavras), mas, sim, exemplos retirados das mais diversas fontes e destinados, sobretudo, a dar conta do uso efectivo, real da língua.

Se o dicionário se propõe descrever o uso efectivo dos itens lexicais por parte dos falantes, podemos entender que acabem por registar usos que, de acordo com a norma vigente são usos indevidos, corruptelas, mas que são efectivamente frequentes no uso quotidiano. Entende-se, assim, que o DLPC, por exemplo, tenha registado o uso do adjectivo *solarengo* com o sentido de “banhado pelo sol” – o que não pode aceitar-se é que essa acepção não tenha sido alvo de uma marca de uso que permita ao consulente saber que, embora se diga, a norma ainda considera esse uso errado e que, portanto, deverá recorrer ao adjectivo *soalheiro*.

Dada esta preocupação em retratar o uso efectivo da língua, os dicionários de hoje têm a preocupação de retratar a variação lexical, não apenas a variação diatópica e a

diacrónica – como já ocorria, de resto, ainda que de forma assistemática –, mas também a variação diastrática e a diafásica. Justifica-se, portanto, que o DLPC tenha inserido na sua nomenclatura a forma *bué*, dado tratar-se de uma forma de alta frequência no português informal contemporâneo dos estratos mais jovens da população – o que não se pode aceitar é que as entradas propostas para *bué* não tenham sido alvo de nenhuma marca de uso que as assinale como formas próprias de um registo muito informal da linguagem, próprio de adolescentes e jovens, não se tratando, portanto, de unidades passíveis de ser usadas em qualquer contexto, como o seu registo não marcado deixa supor.

Os exemplos de uso são um complemento da informação gramatical, que passa a ocupar um espaço cada vez mais alargado não apenas no interior do artigo lexicográfico, como em apêndices constituintes da macro-estrutura do dicionário. Além da categoria, informação gramatical clássica que qualquer dicionário de língua comporta, o dicionário passou a incorporar informação relativa a subcategoria (ex.: nome contável *vs.* nome não-contável), flexão (especialmente a não regular), estrutura argumental de verbos, regências, combinatórias (co-ocorrências privilegiadas), palavras derivadas (por exemplo, indicação de advérbios em *–mente* derivados de adjetivos que constituem lemas), etc. A incorporação crescente de informação gramatical na micro-estrutura do dicionário tem vindo a esbater as tradicionais fronteiras entre a descrição própria da gramática (a das regularidades da língua), por oposição ao dicionário (commumente considerado repositório das idiossincrasias próprias de cada item lexical).

Como consequência da visão do léxico como “rede” de itens, os dicionários de hoje apresentam um número cada vez mais significativo de remissões para outros artigos, relativos a unidades relacionadas com aquela em análise. A remissão é o meio disponível, nos dicionários impressos, necessariamente organizados por ordem alfabética, para o estabelecimento dessas ligações entre artigos impressos necessariamente em pontos distintos da obra. Já nos dicionários publicados em suporte digital (CD-Rom), mais leves, de consulta mais fácil e mais baratos, a representação destas relações é muitíssimo mais eficiente: graças ao recurso a hiperligações, o consulente pode navegar no seu dicionário de forma eficiente e agradável. Infelizmente, porém, em Portugal as entidades responsáveis pela publicação de dicionários (salvo honrosas exceções) continuam a considerá-los produtos de luxo, destinados apenas a pequenas elites, continuando, portanto, a apostar em

edições impressas desnecessariamente luxuosas e inaceitavelmente caras, e não investindo, por exemplo, na disponibilização de edições em CD-Rom (para além de edições impressas económicas, acessíveis a todos os públicos, como acontece com os grandes dicionários das grandes línguas mundiais).

Os dicionários em suporte digital permitem também facilidades de consulta antes inimagináveis. Com um bom dicionário em CD-Rom é possível, por exemplo, encontrar palavras com determinadas terminações, entradas que contenham determinada palavra na definição. Para quem trabalha em lexicologia ou em metalexicografia, por exemplo, estas possibilidades são de extrema importância.

Os dicionários incorporam uma quantidade cada vez maior de informações distintas sobre cada unidade lexical. Esta incorporação tem conduzido a uma complexificação crescente da microestrutura lexicográfica – é cada vez mais complexo o sistema de representação e diferenciação gráfica dos diferentes tipos de informação nela incorporados (fontes tipográficas diferentes, maiúsculas, negros, itálico, símbolos diversos, etc.). Esta complexificação é de tal ordem que acreditamos que a maioria dos consulentes não é capaz de a descodificar (apesar de os bons dicionários incorporarem actualmente “guias do dicionário”, onde é explicitada a estrutura dos artigos) e, infelizmente, os professores de língua não estão geralmente alertados para esta questão, não ensinando os estudantes a tirar o melhor partido do dicionário, pressupondo um conhecimento que eles efectivamente não têm.

De tudo o que temos vindo a expor, verificamos que a lexicografia deixou de privilegiar a faceta prescritiva que a caracterizou durante séculos, para centrar os seus esforços numa tentativa de descrição o mais eficiente possível das unidades lexicais tal como são efectivamente usadas e de tudo o que sobre elas sabemos actualmente.

Chamaremos a esta lexicografia de pendor descritivo “lexicografia comunicativa”, por duas razões: em primeiro lugar, porque se trata efectivamente de uma lexicografia centrada no consulente, visando incrementar o seu domínio da língua, a sua produção linguística (para além da mera compreensão) e a sua capacidade comunicativa; em segundo lugar, para retomar a designação de “Teoria Comunicativa da Terminologia” proposta por Maria Teresa Cabré para a sua abordagem teórica do trabalho terminológico,<sup>19</sup> e que no fundo se apresenta, na minha opinião, como uma tentativa de trazer para o âmbito do

estudo e da descrição das terminologias científicas e técnicas as práticas que a lexicografia de pendor descritivo desenvolveu sobretudo a partir das décadas de 60 e 70 do século XX, com resultados muito profícuos para ambas as disciplinas.

## **6. Uma lexicografia meramente descritiva tem valor social?**

Ao longo deste trabalho temos vindo a mostrar que, se os dicionários que hoje conhecemos são diferentes dos dicionários do passado e fonte de muitas perplexidades, tal não acontece por acaso e é fruto do próprio desenvolvimento da ciência linguística e da lexicologia, assim como do desenvolvimento tecnológico.

Vimos que os dicionários hoje se constituem como instrumentos privilegiados de promoção da comunicação linguística, em língua materna, segunda ou estrangeira, nas suas vertentes de compreensão, mas também de produção. Vimos que os dicionários perderam, portanto, o carácter fundamentalmente prescritivo que os caracterizou durante séculos.

Mas a verdade é que, se os dicionários mudaram, o seu público não mudou ao mesmo ritmo e, hoje como antes, o consulente continua a ver no dicionário um instrumento normalizador, um repositório de usos correctos, uma resposta para as suas dúvidas, enfim, o “manual de tira-teimas” que às vezes se refere. Mais do que explicações, o consulente do dicionário procura respostas concretas para perguntas do tipo: Posso dizer isto? Esta expressão é correcta? Farei boa figura se usar esta palavra? Em suma, o dicionário é, ainda, para o consulente, um orientador dos usos.

Além disso, o dicionário é (ou deveria ser) um instrumento pedagógico, útil para a aprendizagem de uma língua estrangeira, mas também para o estudo da língua materna, na escola. Ora, o ambiente linguístico em que vivemos em Portugal, a nível escolar, pode caracterizar-se por ser um ambiente em que as atitudes dominantes oscilam entre a mera imposição da norma culta e o bidialectalismo (aceitação do dialecto do aluno como objecto de estudo, mas imposição do uso da norma linguística).<sup>20</sup> Por outro lado, o mercado de trabalho é muitas vezes implacável relativamente aos usos “incorrectos” da língua, aos usos violadores da norma instituída. Deste modo, é desejável que o falante tenha ao seu dispor obras que lhe indiquem o modo “correcto” de falar, porque vai precisar de falar desse modo em diversos contextos de comunicação ao longo da sua vida.

Ora, o modo de conciliar uma lexicografia comunicativa moderna com estas prerrogativas passa, a meu ver, por um tratamento muito cuidadoso da variação, particularmente da diastrática e da diafásica, assim como dos usos não canónicos mas frequentes de determinadas unidades lexicais (muitas vezes indícios de mudanças em curso na língua), e do uso coerente e sistemático de marcas de uso no interior da microestrutura que possam tornar o dicionário capaz de conciliar as suas duas perspectivas fundamentais, a descritiva e a prescritiva, que, afinal, não são, nem podem ser, incompatíveis, mas devem, isso sim, ir de mãos dadas na produção lexicográfica.<sup>21</sup>

Assim, ainda que a lexicografia portuguesa muito se tenha desenvolvido nos últimos anos, muitos desafios novos se lhe colocam neste início de século e de milénio e muito trabalho se encontra ainda por realizar. Assim a sociedade portuguesa possa ultrapassar alguns preconceitos e traumas sentidos nos últimos anos e continuar a apostar na produção de uma lexicografia de qualidade, moderna, bem como no estudo do dicionário, na metalexicografia, como modo privilegiado de contribuir para o seu aperfeiçoamento.

---

<sup>1</sup> As críticas, tantas vezes injustas, e o debate levado a cabo na imprensa aquando da publicação, em 2001, do *Dicionário da Língua Portuguesa Contemporânea* (vulgarmente conhecido como “o dicionário da Academia”) são a prova cabal da incompreensão da sociedade em relação a muitas das características da moderna lexicografia.

<sup>2</sup> Cf. TRUDGILL, *Sociolinguistics – An introduction to language and society*. 4.<sup>a</sup> ed. Londres, 2000.

<sup>3</sup> Cf. BIDERMAN, Maria Tereza Camargo, *Teoria lingüística*, São Paulo, 2001, p. 179 e *passim*.

<sup>4</sup> Cf. por exemplo, CORBIN, Danielle, *Morphologie dérivationnelle et structuration du lexique*, 2 vols., Tubinga, 1987, e restantes trabalhos desta autora, os trabalhos publicados volume da revista *Lexique*, n.º 10 (Villeneuve d’Ascq, 1991), ou CORREIA, Margarita, *A denominação das qualidades em português – contributos para a compreensão da estrutura do léxico português*, diss. de Doutoramento, Lisboa, 1999.

<sup>5</sup> Mattoré, ao referir a confusão entre a lexicologia e outras disciplinas, refere o facto de, frequentemente, os estudos lexicológicos se confundirem com estudos de estilística, dada a confusão entre o estudo do vocabulário de uma época e de uma região e o estudo do estilo de um determinado autor. Estudava-se o vocabulário de autores consagrados, por se acreditar que estes teriam a prerrogativa de contribuir para a evolução da língua (cf. MATTORÉ, Georges, *La méthode en lexicologie – domaine français*, Paris, 1959, pp. 10-11).

<sup>6</sup> Cf. a obra clássica de SINCLAIR, John, *Corpus, Concordance, Collocation*, Oxford, 1991, particularmente os seus capítulos 4, 5 e 6. Cf. também toda a obra de BIBER, Douglas, Susan CONRAD e Randi REPPEN, *Corpus Linguistics – Investigating Language Structure and Use*, Cambridge, U.K., 1998, que, ao longo das suas 300 p., nos oferece uma panorâmica bastante informativa dos vários tipos de estudos realizados até à data em lingüística de corpus.

<sup>7</sup> Cf. AITCHINSON, Jean, *Words in the Mind*, Cambridge, Mass. e Oxford, U.K., 1987.

<sup>8</sup> Cf. PUSTEJOVSKI, James, *The Generative Lexicon*, Cambridge, Mass., 1995.

<sup>9</sup> SARDINHA, Tony Berber, *Lingüística de Corpus*, Barueri – SP, 2004, p. xvii (prefácio da obra).

<sup>10</sup> Cf. DUBOIS, Jean *et al.* *Dicionário de Lingüística*, 8.<sup>a</sup> ed. da tradução brasileira, 2001 (ed. original, *Dictionnaire de Linguistique*, Paris, 1973), s.v. corpus.

<sup>11</sup> SARDINHA, Tony Berber, *op. cit.*, p. 18.

Correia, Margarita 2008. Lexicografia no início do século XXI – novas perspectivas, novos recursos e suas consequências. In: Júnior, Manuel Alexandre (coord.) *Lexicon – Dicionário de Grego-Português, Actas de Colóquio*. Lisboa: Centro de estudos Clássicos / FLUL, pp. 73-85. (versão entregue para publicação)

<sup>12</sup> Cf. SARDINHA, Tony Berber, *op. cit.*, p. 1

<sup>13</sup> Cf. BIBER, Douglas, Susan CONRAD e Randi REPPEN, *op. cit.*, pp. 23-24 e ss.

<sup>14</sup> Também no passado, o desenvolvimento da lexicografia beneficiou do desenvolvimento do ensino de línguas não-maternas: “A lexicografia começou a estruturar-se como disciplina linguística desde a primeira metade do século XVI, em vários centros humanísticos europeus. Foi inicialmente motivada pelas solicitações do ensino do latim como língua não materna, e encontrou na técnica tipográfica uma condição determinante para a sua configuração e difusão.” (p. 15) – VERDELHO, Telmo, “Dicionários portugueses, breve história”, in NUNES, José Horta e Margarida PETTER (orgs.), *História do saber lexical e constituição de um léxico brasileiro*, São Paulo, pp. 15-64.

<sup>15</sup> Este dicionário, publicado pela Porto Editora, não apresenta exemplos de uso nem abonações e a sua informação gramatical limita-se à indicação da categoria ou categorias a que corresponde cada uma das suas entradas, características que fazem dele um dicionário apto apenas para a função de compreensão ou descodificação. Tem, no entanto, a virtude de ser o dicionário de língua portuguesa de nomenclatura mais extensa (cerca de 130 000 entradas) que apresenta transcrição fonética dos lemas.

<sup>16</sup> Torna-se, portanto, difícil de compreender que um dicionário publicado no início do século XXI, da dimensão do *Dicionário Houaiss da Língua Portuguesa*, apresente nas versões brasileira e portuguesa da sua “Introdução” afirmações do tipo das seguintes, sem fornecer qualquer pista concreta sobre as efectivas opções metodológicas assumidas na sua composição: «Do trabalho geral resultou uma obra de cerca de 228.500 unidades léxicas que não privilegia determinada faixa cronológica e geográfica da língua. Versa diacronicamente sobre fenómenos não apenas do português contemporâneo do Brasil e de Portugal, mas ainda, embora de forma selectiva, sobre vocábulos da língua antiga e da arcaica (...).

Outra característica da obra é a sua vocação lusófona, que a fez debruçar-se sobre dialectalismos tanto brasileiros como portugueses, assim como registar e definir na sua nomenclatura palavras e locuções dos crioulos orientais e africanos de origem portuguesa, além de vocábulos de outros idiomas – por exemplo do chinês e de algumas línguas africanas -, incorporados ao nosso léxico por se registarem em obras literárias cujo meio de expressão foi o português.» (VILLAR, Mauro, “Introdução”, *Dicionário Houaiss da Língua Portuguesa*, versão portuguesa, Lisboa, 2002-2003, p. xv – sublinhados meus).

Poucos estudos metalexigráficos foram, ainda levados a cabo sobre o *Dicionário Houaiss*, que, a meu ver, foi aceite acriticamente pela sociedade de fala portuguesa como a obra lexicográfica de referência contemporânea, mas que, no caso da versão portuguesa, provavelmente pelas limitações impostas à sua adaptação, acaba por se constituir como um objecto híbrido, assistemático e incoerente, que não representa já o léxico da variedade brasileira, mas que está muito longe de ser representativo do léxico português. Refiram-se BIDERMAN, Maria Tereza Camargo, “Análise de dois dicionários gerais do português brasileiro contemporâneo: o Aurélio e o Houaiss”, in Isquierdo, Aparecida N. & M. da Graça Krieger (orgs.) *As ciências do léxico: lexicologia, lexicografia e terminologia*. Vol. II. Campo Grande (MS-Brasil), 2004, pp. 185-200 e CORREIA, Margarita, “Versão portuguesa *Dicionário Houaiss da Língua Portuguesa*: um dicionário do português europeu?”, in Miyares, Leonel Ruiz, Alex Muñoz Alvarado & Célia Alvarez Moreno (eds.) *Actas – X Simpósio Internacional de Comunicación Social Santiago de Cuba*, Santiago de Cuba, pp. 270-274.

<sup>17</sup> *Apud* MURAKAWA, Clotilde de Almeida Azevedo, “D. Raphael Bluteau: marco na lexicografia portuguesa de setecentos”, in MURAKAWA, C. A. A. e Filomena GONÇALVES (orgs.), *Novas contribuições para o estudo da história e da historiografia da língua portuguesa*, Araraquara – SP, 2007, pp. 159-188 (p. 165).

<sup>18</sup> Cf. BIBER, Douglas, Susan CONRAD e Randi REPPEN, *op. cit.*, 1998, p. 22.

<sup>19</sup> Cf. CABRÉ, Maria Teresa, “Una nueva teoría de la terminología: de la denominación a la comunicación”, in CORREIA, Margarita (org.). *Terminologia Desenvolvimento e Identidade Nacional – Actas do VI Simpósio Ibero-Americano de Terminologia* [comunicação apresentada em 1998], Lisboa, pp. 41-60.

<sup>20</sup> Cf. TRUDGILL, Peter, *op. cit.*, p. 198-203.

<sup>21</sup> Um trabalho interessante a realizar é verificar como o *Dicionário de Língua Portuguesa*, da Editorial Verbo, editado em 2006 com apoio do Ministério da Educação e assumidamente destinado ao uso em contexto de estudo da língua materna, trata a variação e até que ponto usa de forma eficiente as marcas de uso.