

RESENHA/REVIEW

HALLIDAY, MICHAEL A. K.; Wolfgang Teubert; Colin YALLOP e Anna CERMÁKOVÁ (2004). *Lexicology and corpus linguistics – an introduction*. Londres e Nova York: Continuum, 184 págs.*

Resenhado por Tony BERBER SARDINHA
(Pontifícia Universidade Católica de São Paulo)**

Key-words: *corpus linguistics; lexicology; linguistic theory.*

Palavras-chave: *lingüística de corpus; lexicologia; teoria lingüística.*

1. Introdução

Lexicology and Corpus Linguistics é uma coletânea de quatro artigos, todos inéditos, escritos por Michael Halliday, Wolfgang Teubert, Colin Yallop e Anna Cermáková. Michael Halliday, professor emérito da Universidade de Sydney, Austrália, é um dos maiores lingüistas de todos os tempos, fundador da Lingüística Sistêmico-Funcional. Wolfgang Teubert é titular da cadeira de Lingüística de Corpus da Universidade de Birmingham, Grã-Bretanha e editor-chefe do *International Journal of Applied Linguistics (IJCL)*. Colin Yallop é professor adjunto da Universidade Macquarie, na Austrália. E Anna Cermáková é professora da Universidade Carolina, da República Tcheca e editora assistente do *IJCL*.

2. Síntese do conteúdo

O primeiro artigo é de Michael Halliday e se intitula 'Lexicology'. É o menor de todos, com apenas 22 páginas. Nele, Halliday oferece um pano-

* Agradeço os comentários dos pareceristas.

** Agradeço ao apoio do CNPq, mediante a bolsa de número 350455/03-1.

rama de assuntos relacionados ao estudo, pesquisa, teoria e análise da lexicologia. Ele começa problematizando a noção de palavra. Segundo ele, a palavra é uma das entidades mais tangíveis da língua, pois está presente em jogos (palavras cruzadas, etc), dicionários e na escola, desde a alfabetização. A palavra, contudo, é também um dos conceitos mais difíceis de definir universalmente, pois está intimamente ligado a aspectos do desenvolvimento da escrita no Ocidente. A palavra, no Ocidente, deriva da escrita alfabética e depende do espaço em branco, como separador de palavras; essa convenção demorou séculos para se estabelecer, como mostram os primeiros textos da Grécia Antiga, que eram escritos sem divisão entre as palavras. Já no Oriente, essa convenção não funciona, pois os caracteres identificam morfemas e não letras, e o espaço em branco não delimita palavras. Se deixarmos de lado a língua escrita e nos determos na língua falada, o problema da delimitação de palavras aumenta consideravelmente, pois na fala contínua as pausas e limites marcados fonética ou acusticamente não correspondem prontamente às unidades grafêmicas. Esse assunto é retomado por Teubert, no terceiro capítulo do livro.

Tendo em vista esse cenário, Halliday oferece o conceito teórico de 'item lexical'. Um item lexical é uma unidade da léxico-gramática da língua, que forma o seu vocabulário, ou léxico. O item lexical não está atrelado à forma como é grafado ou pronunciado; ele é uma unidade de sentido, que resulta de escolhas do falante ou escritor. Como cada sentido foi escolhido ante a uma série de outras opções disponíveis, cada item lexical constrói sentido não somente com as demais escolhas realizadas, mas também frente às escolhas não realizadas. Teubert, no terceiro artigo do livro, enfatiza o mesmo ponto, por meio da noção de 'unidade de significado' (vide mais abaixo).

Na parte seguinte, Halliday trata de duas maneiras culturalmente definidas de sistematizar o conhecimento das palavras: o dicionário e o 'thesaurus'. Aqui, ele introduz dois conceitos importantes: lema e taxonomia lexical. Um lema ('lemma') é a forma base de uma palavra; é aquela que encabeça um verbete, normalmente a forma morfológicamente mais simples (singular, infinitivo, etc.). O lema é o conceito organizador do dicionário. Uma taxonomia lexical, por sua vez, é uma organização de palavras em classes e sub-classes por meio das relações semânticas de hiperonímia ('x é um tipo de y') e de meronímia ('x é um tipo de y'). A taxonomia lexical é o princípio organizador do 'thesaurus'.

Em seguida, Halliday apresenta um histórico da lexicologia na Antigüidade no Ocidente e no Oriente. Halliday nos ensina que na Índia, nos séculos II e III antes de Cristo, foram compilados glossários das escrituras sagradas Hindus, e que na China, também no século III a.C., surgiu o precursor do 'thesaurus', o 'Er Ya'. Um dos dados mais impressionantes é a compilação de enciclopédias e dicionários gigantescos na China, culminando com a enciclopédia Yongle, que foi comissionada nas dinastias Ming e Qing e possuía mais de 10 mil volumes!

A seção histórica termina com o desenvolvimento de dicionários na Inglaterra e nos Estados Unidos. Os primeiros dicionários ingleses eram listas bilíngües do inglês e latim. O termo 'dictionary' apareceu na Inglaterra no século XIII, e o primeiro dicionário inglês monolíngüe foi o de Robert Cawdry, de 1604. O famoso dicionário de Samuel Johnson apareceu em 1755 e estabeleceu a tradição de os dicionários normatizarem a ortografia, pronúncia e definição das palavras. O dicionário de Noah Webster de 1828, nos Estados Unidos, é emblemático dessa tradição, assumindo para si a função de demarcar o inglês americano como uma língua própria: as modificações de ortografia do inglês introduzidas por ele tornaram-se norma nos Estados Unidos.

O artigo se encerra com considerações a respeito da lexicografia no final do século XX. Halliday argumenta que duas inovações revolucionaram a lexicologia e a lexicografia: o computador e o corpus eletrônico. Segundo ele, as duas tornaram muito mais rápido o trabalho de compilação de dicionários e glossários. Além disso, introduziram mudanças significativas na maneira pela qual os dicionários são feitos, pois o material para sua preparação deixou de ser coletado a partir da observação a olho nu e da memória do lexicógrafo, passando a ser levantado por meio do próprio computador. Isso por sua vez trouxe uma mudança quantitativa e qualitativa à empreitada: o lexicógrafo passou a dispor de mais informação e também de informações diferentes, como a frequência do uso. O próprio produto final também se alterou, devido às edições computadorizadas dos maiores dicionários que podem ser instaladas na máquina do consulente.

O segundo artigo é de autoria de Colin Yallop e se intitula 'words and meaning'. O artigo começa com o pressuposto de Halliday de que o significado é primordial no uso lingüístico; assim, usamos a língua primordialmente para trocar e construir sentidos:

“For meaning is ultimately shaped and determined by communal usage. A dictionary definition of a word’s meaning has authority only in so far as it reflects the way in which those who speak and write the language use that word in genuine communication. In this sense, meaning has a social quality, and while it is sometimes convenient to think of the meaning of a word as a concept, as ‘something stored in the mind’, this legitimate only to the extent that the concept is seen as an abstraction out of observable social behavior.” (:29)

Logo, o autor deduz, os dicionários deveriam privilegiar os significados em uso, mas isso nem sempre acontece. Ele cita o exemplo de ‘rose’ em um dicionário australiano de inglês (Macquarie Concise), em que o primeiro sentido é o técnico (‘espécie do gênero Rosa’). Porém, o sentido técnico não é o mais empregado na maioria das situações de produção e compreensão da língua. Por que então o sentido técnico aparece em primeiro lugar no dicionário? A resposta é que esse parece ser, na mente dos usuários, um sentido ‘básico’, ‘concreto’, ‘objetivo’ ou ‘nuclear’ (‘core’), que de algum modo subjaz aos demais e portanto deve anteceder os demais no dicionário. Yallop discorda dessa premissa, argumentando que não há sentido central e periférico de modo absoluto; há, sim, sentidos mais e menos presentes na língua em uso. A frequência do uso deveria ser o guia para a organização dos sentidos nos verbetes.

Os dicionários, entretanto, historicamente tomaram para si a função de registrar a origem das palavras, incluindo palavras e sentidos que estão em desuso. Essa função entra em conflito com a idéia de usar a frequência de uso corrente como parâmetro para a organização do dicionário. Yallop tenta resolver esse impasse reconhecendo o valor do registro da etimologia nos dicionários, tanto para satisfazer a curiosidade dos usuários como para manter viva a memória da língua.

A seguir, Yallop retoma a questão da prescrição, já introduzida por Halliday no artigo anterior, oferecendo um pequeno panorama histórico de tentativas bem e mal sucedidas de regulamentar a língua por meio de dicionários e de gramáticas. Ele menciona o caso do comparativo e superlativo no inglês, que pode ser formado atualmente tanto por sufixação (-er ou -est) ou parafrasticamente com os advérbios ‘more’ ou ‘most’. Entretanto, no período Elizabetano, havia a possibilidade de usar as duas formas conjuntamente, como em ‘more corrupter’ (o que hoje seria na norma culta ‘more corrupt’). Por volta do século XVIII, os gramáticos começaram a

criticar esse emprego e a exigir que se usasse um tipo apenas, o que permanece até hoje na escrita. Mas Yallop lembra que em países onde o inglês é primeira língua, ainda se usa a forma composta na fala, o que significa que mesmo séculos de prescrição não serviram para extinguir uma convenção de uso. Os puristas reclamam dessa forma combinada dizendo que é redundante, mas Yallop oferece muitos exemplos para demonstrar que a língua inglesa é repleta de redundâncias; por exemplo, a marca de plural no substantivo dentro de um grupo nominal pode ser vista como redundante: em ‘two pencils’, o –s é supérfluo dada a ocorrência de ‘two’¹.

Em seguida, o autor discute a questão da mudança do sentido típico de palavras ao longo do tempo, citando vários casos. Um deles é da palavra ‘gender’, do inglês, que passou do sentido de ‘gênero gramatical’ para ‘gênero humano’. Segundo ele, anos atrás essa palavra era comumente usada em expressões do tipo ‘the three genders in German’, para designar os gêneros masculino, feminino e neutro da língua alemã, tendo passado mais recentemente a se referir ao sexo das pessoas, substituindo a palavra ‘sex’, que passou a se associar ao ato sexual. Podemos verificar essa observação de Yallop fazendo uma busca na Web por cada uma das palavras. Enquanto ‘gender’ traz mais referências ao gênero biológico (inclusive em formas compostas como ‘gender roles’, ‘transgender’, ‘gender equity’, etc.), ‘sex’ recupera textos de conteúdo erótico. Yallop explica essas mudanças (‘shifts’) de sentido pelas mudanças na sociedade: as alterações do léxico refletem as transformações da sociedade.

No restante do artigo, Yallop trata de assuntos de natureza teórica, que compreendem três partes principais. A primeira se trata de um breve apanhado das idéias de Saussure. Entre elas, o autor destaca a da arbitrariedade do signo e a da descrição sincrônica da língua. Ambas são importantes na medida em que dão suporte ao argumento central que o autor tece ao longo do seu texto de que as línguas humanas são produto do meio social e portanto cada língua possui um vocabulário independente, que deve ser estudado sincronicamente, sem a pretensão de ser explicado por meio da etimologia. A segunda também é um apanhado, desta vez de Firth, de onde o autor recupera a noção de significado como função em

1. A eliminação dessa redundância, a propósito, é exatamente uma das marcas do português falado coloquial, em expressões como ‘duas caneta nova’, em que a marca de plural incide apenas no determinante.

contexto ('meaning as function in context', :49). Essa definição de significado nasceu do trabalho antropológico de Malinowski, que ressaltava a necessidade de estudar a linguagem em contexto. Para Firth, estudar a linguagem em contexto implicava estudar seus usos, quem a usava, quando e como, o que por sua vez demandava coletar muitos exemplos de fala ou escrita autênticas. Essa insistência no estudo da linguagem por meio de seus usos reavivaria, décadas mais tarde, o interesse na compilação de corpora, que acabariam se tornando eletrônicos, com a chegada do computador digital. Essa progressão histórica serve de ponto de partida para os capítulos seguintes do livro, que tratam justamente da Linguística de Corpus. A terceira parte diz respeito à obra de Chomsky na linguística. Os principais temas resgatados pelo autor são o mentalismo (uma reação à chamada 'abordagem taxonômica' da linguística que visa a explicar a língua por meio de processos mentais) e o universalismo (a existência de estruturas linguísticas profundas comuns a todas as línguas), ambos apresentados como contraponto à visão social de língua de Saussure e Firth. Essa contraposição pode ser resumida nesta passagem:

"If there are universals of language, they are best approached from the perspective of how language functions in human life and how it serves human purposes. All languages seem to be systems for making meanings (...). This is a quite different approach to universals from one which seeks to find a common core vocabulary or a universal set of concepts." (:68)

O terceiro artigo é de autoria de Wolfgang Teubert e se intitula 'language and Corpus Linguistics'. Ele se inicia com uma longa exposição acerca do desenvolvimento da linguística ao longo dos séculos. O ponto principal é as diversas escolas linguísticas aplicam princípios semelhantes, de um ponto de vista abstrato, que são a busca de leis, entidades e propriedades nas várias línguas descritas. Tomando dois casos particulares, o estruturalismo e o gerativismo, Teubert argumenta que esses princípios são empregados para intuídos diferentes. Enquanto no estruturalismo buscam-se regras para descrever como as línguas são escritas ou faladas, no gerativismo procuram-se essas regras para mostrar como as línguas poderiam ou não ser. No estruturalismo, 'rules were expressions of the linguist's ingenuity to make sense of the language evidence' (:78), ao passo que:

'Chomskyan linguistics thus changes the status of linguistic rules. Rather than being tools for language analysis, they now become the metaphysically real essence of language'. (:78)

Apesar dessas diferenças entre uma escola e outra, ou mesmo entre diferentes seguidores de uma mesma escola, há uma semelhança marcante entre todas, que é em relação à inabilidade geral do pensamento lingüístico em lidar com o vocabulário. Ou seja, as tentativas de aplicação de regras descritivas ou gerativas para sistematizar o léxico costumam falhar, seja porque as regras mais abstratas deixam de refletir as particularidades do vocabulário em uso, seja porque as regras mais específicas se multiplicam exponencialmente para dar conta de toda a variação do emprego de palavras na língua.

Teubert exemplifica esse problema citando casos de tradução de palavras de uma língua para outra, a partir do exame da informação fornecida por dicionários bilíngües. Uma dessas é 'grief', palavra inglesa cuja tradução para o francês seria 'chagrin' ou 'peine', segundo o Collins Robert French Dictionary. O problema nesse caso é saber se as duas opções são equivalentes ou se uma delas é preferível em dado contexto. Para tentar resolver essa questão, um consulente poderia procurar no mesmo dicionário cada uma das opções na língua de chegada. Fazendo isso, ele descobriria que para 'chagrin', o dicionário dá '(=affliction) grief, sorrow'. Essa informação poderia ser interpretada como uma confirmação de 'chagrin' como tradução de 'grief', até o mesmo consulente descobrir que 'peine' também traz 'grief' como sua tradução, o que significa que o consulente teria voltado ao ponto de partida, permanecendo a dúvida de qual opção seria a mais adequada. Para responder isso, é preciso saber o que cada opção significa e para tanto o dicionário apresenta outras palavras no verbete juntas podem elucidar a questão. Voltando ao verbete de 'chagrin', conforme vimos há ainda 'sorrow'. Um falante nativo do inglês provavelmente ficaria mais confuso, pois há uma diferença entre 'sorrow' e 'grief', e sendo assim ele não saberia se 'chagrin' quer dizer uma mistura de 'sorrow' e 'grief' ou apenas um deles. Enfim, a questão não tem solução com base na informação fornecida somente pelo dicionário, mesmo que seja uma obra de prestígio como o Collins Robert, pois as palavras isoladas são ambíguas por natureza. Elas adquirem sentido somente quando estão em uso (mesmo que esse uso tenha de ser imaginado, como em muitas vertentes da lingüística). Assim, Teubert conclui que as unidades de sentido não são cada palavra individualmente, mas sim conjuntos de palavras. A ambigüidade não existe nas unidades de sentido; ela existe nas palavras isoladas, justamente porque as palavras em si não são unidades de sentido:

'A unit of meaning is a word (often called the node or keyword) plus all those words within its textual context that are needed in order to disambiguate this word, to make it monosemous.' (:83)

Esse conceito é análogo ao de 'item lexical', proposto por Halliday no primeiro capítulo da obra, embora Halliday esteja se referindo de modo mais amplo também às escolhas de palavras não realizadas, paradigmaticamente, dentro das possibilidades do discurso.

Teubert, a seguir, se pergunta por que a palavra isolada continua exercendo fascínio na lingüística, tendo em vista esses problemas de ambigüidade? A resposta, segundo ele, tem a ver com a tradição lingüística de analisar a língua a partir da sintaxe. Essa primazia da sintaxe coloca o léxico em segundo plano, como um mero enchimento de casulos sintáticos; como esses casulos não colocam restrições de sentido para serem preenchidos, acabam sendo ocupados por palavras isoladas, gerando assim frases ambíguas. Isso leva muitos a supor que o léxico é por natureza ambíguo, quando na verdade ele o é em frases inventadas ou criadas especificamente com a finalidade de serem ambíguas (como no humor, ironia, etc.).

Se a ambigüidade deixa de existir normalmente quando se examina a linguagem em uso, por que ela continua sendo atrelada ao estudo do vocabulário? Segundo Teubert, basicamente porque as disciplinas tradicionais que se dedicam ao estudo do léxico terem falhado, cada uma a seu modo. Na semântica, devido ao viés cognitivo que a faz buscar uma língua universal no nível do pensamento ('the language of thought' :94). E na lexicologia, devido à insistência em decompor o significado em unidades abstratas, categóricas.

Uma saída para esses problemas, segundo Teubert, é a Lingüística de Corpus. A Lingüística de Corpus como tal não chega a ser definida no artigo, mas o leitor é informado de que ela representa uma nova maneira de fazer lingüística, que se iniciou com a disponibilidade de corpora, isto é, coleções de textos armazenadas em formato de computador. Teubert oferece uma pequena cronologia da área, colocando o Survey of English Usage (SEU), de Randolph Quirk, como o primeiro corpus, embora não fosse computadorizado. O SEU foi iniciado em 1959 e mais tarde teve seu componente oral (500 mil palavras) informatizado, formando assim o corpus London Lund. O primeiro corpus computadorizado foi o Brown, lançado em 1964. Atualmente há um número incontável de corpora em existência, para uma grande variedade das línguas do mundo.

Teubert lembra que o SEU não foi criado para estudar o significado, mas sim a sintaxe. Foi em Birmingham (Grã-Bretanha), contudo, que um tipo de Lingüística de Corpus que se preocupa com o estudo do significado por meio do léxico nasceu e se firmou, tendo em John Sinclair seu fundador. O restante do artigo equaciona a Lingüística de Corpus com essa vertente sinclairiana e nos lembra que seu conceito principal é a colocação, que surgiu nos anos de 1933 com Harold Palmer e A. S. Hornby (1933) e foi mais tarde retomada por John Firth (1957)). Firth foi professor de Sinclair em Londres e deve ter sido por meio desse contato que colocação entrou para a agenda de pesquisa de Sinclair. O primeiro estudo sistemático de colocações com corpora eletrônicos foi o *English Lexical Studies*, de John Sinclair, também conhecido como OSTI Report, distribuído precariamente em 1970. Esse estudo seria publicado definitivamente somente em 2003 (Sinclair, Jones, Daley, & Krishnamurthy, 2004). A ênfase da Lingüística de Corpus sinclairiana foi desde cedo o estudo do significado, reconhecendo que as palavras não possuem significado em si, mas em contexto. Isso, segundo Teubert, revela a verdadeira essência da Lingüística de Corpus, que deve ser social: 'Corpus Linguistics sees language as a social phenomenon. Meaning is, like language, a social phenomenon.' (:97)

Apesar disso, Teubert reconhece que a Lingüística de Corpus encampa outros interesses, o que fica patente na produção bibliográfica da área. Ele observa que os primeiros manuais de Lingüística de Corpus, de McEnery e Wilson (1996), Kennedy (1998) e Biber et. al. (1998), davam pouco ou nenhum espaço ao estudo do significado das palavras, preferindo enfocar questões como frequência, construção e exploração de corpora.

O quarto e último capítulo do livro intitula-se 'directions in Corpus Linguistics' e tem autoria conjunta de Wolfgang Teubert e Anna Cermáková. O primeiro assunto tratado por eles é a representatividade de corpus. Representatividade significa a propriedade de amostra de replicar, em escala menor, uma população. Um corpus é uma amostra de uma população. Essa população, Teubert e Cermáková chamam de discurso ou linguagem ('language'): 'a language, a discourse, consists of the totality of verbal interactions that have taken place in the community where this language is spoken' (:114).

A representatividade é um das questões mais sérias da compilação de corpus, pois ela impacta diretamente na capacidade de generalização dos

resultados advindos do próprio corpus. Se o analista descobre que uma determinada construção ou colocação ocorre x vezes no corpus, pode ele generalizar essa freqüência para a população de origem do corpus? A resposta será afirmativa somente se o corpus for representativo, isto é, se for um modelo em escala menor daquela população. Ou, nas palavras dos autores:

'(...) talking about the frequency of words, we just may be able to say that a corpus represents a discourse, inasmuch as the 10,000 most frequent words of the discourse are also the 10,000 most frequent words of the corpus.' (:116)

Caso contrário, a resposta é negativa, ou seja, o analista não terá lastro para extrapolar os resultados. O grande problema, porém, é saber se o corpus é representativo, e para isso os autores não oferecem solução. Eles argumentam que a única maneira de saber se um corpus é representativo é conhecendo a população fonte:

'We are only justified in claiming that a given corpus is representative of a discourse, however we have defined it, if we have at least in principle, access to all the texts the discourse consists of. (...) only then can we be sure that the corpus we compile as a sample of this discourse is representative.' (:117)

Como não temos acesso à população integral da maioria dos tipos de linguagem, conclui-se logicamente que não podemos saber, nesses casos, se um corpus é representativo: precisamos de uma amostra representativa (corpus representativo) para conhecer a população (linguagem), mas para obter essa amostra precisamos conhecer a população. Trata-se de um verdadeiro 'catch 22': cada etapa depende da outra ser realizada para se concretizar.

Dado que não se pode determinar a representatividade de um corpus, o melhor que se pode fazer, segundo o receituário tradicional, é compilar corpora de maneira criteriosa, especificando a fonte, época, gênero e demais características dos tipos de textos ou transcrição de fala que se deseja incorporar para se obter uma amostra equilibrada. Porém, com a maior diversidade e quantidade de arquivos de texto disponíveis, surgiu um outro tipo de corpus que abandona esse receituário: o corpus oportunista. Um corpus oportunista ('opportunistic') é aquele que consiste de vários outros corpora já existentes, tomando, dessa forma, um vulto incomum,

em termos de tamanho de palavras e de textos. O maior corpus oportunista da atualidade é o COSMAS, da língua alemã, que segundo estimativas já passou da casa do um bilhão de palavras. Com um corpus oportunista, ganha-se em escala, mas não em especificação; assim, os usuários não têm muito controle sobre seu conteúdo, visto que ele cresce muito rápido devido à incorporação de grandes quantidades de material coletado independentemente. Segundo Teubert e Čermáková, o corpus oportunista desautoriza de antemão o questionamento de sua representatividade, pois não faz sentido perguntar se ele oferece um modelo da população. A resposta é um patete ‘não’, visto que ele não foi coletado tendo em mãos uma descrição da população. Com um corpus oportunista, a representatividade, tal qual é conhecida, cai por terra.

Em seguida, os autores introduzem o conceito de conjunto de paráfrases como o as ocorrências de uma palavra de onde se determina seu sentido. Referindo-se ao sentido da palavra ‘unicorn’, os autores afirmam que:

‘this set of paraphrases then is the meaning of the lexical item *unicorn*. It cannot be reduced to a simple form. It is fuzzy, vague, full of contradictions; some of it may be true and some of it may be wrong. It is not the linguist’s task to filter out what they think is right. This is what the linguistic sign unicorn stands for: the set of paraphrases dealing with unicorns. This is what the word unicorn is about. Meaning as paraphrase thus shows us another way of identifying units of meaning. In this perspective, a unit of meaning is whatever we find paraphrases for in the discourse.’ (:132-133)

Para ilustrar, os autores apresentam o caso da palavra ‘globalisation’ (e de sua forma variante grafada com ‘z’, ‘globalization’). Com base em concordâncias, os autores mostram vários usos da palavra na mídia, que na sua totalidade parecem indicar um sentimento avesso a ela. Os autores contrastam isso com a definição dos dicionários e percebem que enquanto na língua os usos dessa palavra evocam muitas ressalvas, no dicionário a definição é neutra. Além disso, os autores salientam que o valor da descrição com corpus está na riqueza de detalhes que ela fornece, deixando entrever os diferentes meios de como a sociedade lida com esse conceito na prática, e como esse uso recorrente define o sentido da palavra pelo uso.

O livro termina com um glossário de quatro páginas e meia de termos de lingüística, seguido das referências (sete páginas) e do índice analítico (quatro páginas).

3. Apreciação geral da obra

Lexicology and Corpus Linguistics – An Introduction é uma obra de difícil compreensão, principalmente porque tenta cobrir áreas e temas muito vastos de modo inapropriado. Não há uma padronização da organização dos capítulos. Cada texto é escrito com uma estrutura particular. O leitor sente falta de um preâmbulo que norteie a leitura da obra, que nos informe qual o seu objetivo, público alvo e relação com demais trabalhos desses campos. A única informação contextual dessa natureza se encontra na capa, onde se lê que o livro é ‘a readable introduction to lexicology and corpus linguistics’ além de ‘an essential textbook for undergraduate students’.

Conforme se pode perceber pela síntese da seção anterior, ao contrário do que faz supor seu nome, *Lexicology and Corpus Linguistics – An Introduction* não é uma introdução. Todos os quatro capítulos exigem do leitor conhecimento prévio de lingüística, lexicologia e de Lingüística de Corpus. Conceitos e argumentos chave não são detalhados para o leitor, ao contrário do que se espera normalmente numa introdução. Por exemplo, não há uma definição clara de Lexicologia nem de Lingüística de Corpus em nenhum dos textos. Além disso, há uma confusão o tempo todo entre lexicologia e lexicografia. Embora as duas áreas estejam relacionadas, lexicologia não é sinônimo de produção de dicionários. Lexicologia é o estudo do significado, da estrutura, da origem e dos processos de formação das palavras, enquanto lexicografia é o estudo e prática de compilação de dicionários (Zé Amvela, 2004 :354). Em relação à lexicografia, o livro também deixa a desejar, pois mesmo levantando várias questões sobre a ordenação de sentidos em verbetes, não apresenta uma proposta concreta. Certamente, o ônus de apresentar alguma saída para os problemas (corretamente) elencados por Yallop será dos professores que porventura adotem o livro.

Se ignorarmos o marketing de que o livro se destina a alunos de graduação e supormos que seu público é de pesquisadores e alunos de pós-graduação, teremos uma outra apreciação da obra. Nesse contexto, os capítulos desempenham bem a função de textos independentes que levantam problemas mais do que apresentam soluções e que propõem conceitos novos, experimentais. Nessa perspectiva, pouco importa a falta de uniformidade dos capítulos.

Não se trata, contudo, de um livro incoerente. A unidade é garantida menos pela proximidade da Lexicologia com a Lingüística de Corpus do

que pela recorrência de dois temas: a tentativa de caracterização de uma lingüística social e o contraponto dessa visão com a lingüística chomskyana.

O primeiro tema funciona como um tipo de premissa central do livro, e me parece essencial para a unidade interna do livro. Além disso, constitui um avanço teórico, mesmo que pequeno, na medida em que assume posição, assim resumida por Yallop:

'in this book we take the view that language is social behavior and meaning a social phenomenon. By this we mean that language is more than an individual possession or ability, that language 'exists' because of its life in social interaction, that meaning is shaped and negotiated in social interaction and that meaning must be studied with due recognition of its social setting.' (:41-42)

Já o segundo tema, o contraponto com a lingüística de Chomsky, me parece muitas vezes desnecessário. À exceção de Halliday, os outros artigos exageram nas críticas a Chomsky. Algumas vezes o leitor sente que são ataques gratuitos, pois pouco acrescentam à argumentação corrente. As referências e ataques à Chomsky e a seus seguidores desdobram-se em muitos pontos dos artigos, especialmente nos de Teubert. Já que não há Lingüística de Corpus ou Lexicologia chomskyana, tal insistência não se justifica. Creio que esse contraponto deveria se resumir a aspectos históricos e a um pano de fundo geral que permitissem ao leitor apreciar as diferenças entre escolas lingüística.

A bem da verdade, uma boa dose de 'Chomsky bashing' tem sido norma na literatura de Lingüística de Corpus. São raras as exceções (e.g. Biber et al., 1998; Sinclair, 1991) ao longo dos 22 anos que nos separam do primeiro livro intitulado 'Corpus Linguistics', que foi Aarts e Meijs (1984). Essas críticas cumpriram seu papel, que é de demarcar território, confrontando o paradigma dominante. Entretanto, hoje a Lingüística de Corpus não precisa mais disso; seu território já está assegurado e só tende a crescer. Ela pode se sustentar sozinha, pelos seus próprios méritos, sem precisar contrapor-se a qualquer outra escola ou área da lingüística.

4. Conclusão

Lexicology and Corpus Linguistics é uma obra que não serve ao propósito a que foi anunciada, qual seja, o de ser uma introdução à Lexicologia e à

Linguística de Corpus, para alunos de graduação. Os textos exigem muito conhecimento prévio de linguística para atender a esse objetivo. Qualquer leitor que possua o conhecimento de linguística presumido já teria ouvido falar de Linguística de Corpus e de Lexicologia e portanto não precisaria de uma introdução a respeito. Além disso, os textos não são didáticos, na medida em que não apresentam os conceitos básicos das áreas de modo sistemático e progressivo. Finalmente, há uma falta de diferenciação entre lexicologia e lexicografia, o que pode confundir o leitor.

Por outro lado, a obra pode ser bem aproveitada em cursos de pós-graduação e também por pesquisadores de nível avançado. Ela firma posição para a Linguística de Corpus, ressaltando seu caráter social e afastando-a das ciências cognitivas.

Recebido em janeiro de 2007

Aprovado em fevereiro de 2008

E-mail: tony@corpuslg.org

REFERÊNCIAS BIBLIOGRÁFICAS

- AARTS, J. M. G., & W. MEIJS. 1984. *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi.
- BIBER, D., S. CONRAD, & R. REPPEN. 1998. *Corpus Linguistics - Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- FIRTH, J. R. 1957. *Papers in Linguistics - 1934-1951*. Oxford: Oxford University Press.
- KENNEDY, G. 1998. *An Introduction to Corpus Linguistics*. New York: Longman.
- MCENERY, T., & A. WILSON. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- PALMER, H., & A. S. HORNBY. 1933. *The Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, J. M., S. JONES, R. DALEY, & R. KRISHNAMURTHY. 2004. *English Collocation Studies: The OSTI Report*. London ; New York: Continuum.
- ZÉ AMVELA, E. 2004. Lexicography and lexicology. In M. Byram. Ed. *outledge Encyclopedia of Language Teaching and Learning* Oxford: Routledge.