

Co-ocorrência léxica no dicionário de Espanhol-Português

ÁLVARO IRIARTE SANROMÁN
(Universidade do Minho)

Para seleccionar ou intensionalizar¹ um referente, isto é, para nos referirmos às coisas² do mundo³ através de uma língua natural utilizamos as palavras dessa língua:

- (1) (a) Flor;
(b) Carro;
(c) Gato;

ou conjuntos de palavras combinadas segundo mecanismos ou regras dessa mesma língua:

- (2) (a) Uma flor vermelha;
(b) O carro do meu vizinho;
(c) O gato que está em cima da mesa.

Costuma-se afirmar que, no primeiro caso, estamos perante um processo de lexicalização, e, no segundo, um processo gramatical. Mas, como indica Crystal, em *The Cambridge Encyclopedia of Language*, a fronteira entre léxico e gramática é incerta (Crystal, 1987: 107), pois, dentro deste segundo processo, o que chamá-mos gramatical, podemos encontrar, para além de combinações livres de palavras, combinações restringidas, que chegam a comportar-se como unidades léxicas. Isto obrigar-nos-á, como veremos, a ultrapassar os limites da palavra como unidade léxica (lexema) e como unidade de tratamento lexicográfico (lema).

A semântica estrutural ensinava-nos que as palavras fragmentam o *continuum* que chamamos realidade de maneiras diferentes dentro das diversas comunidades linguísticas. É verdade que, ao situarmo-nos dentro dos limites da palavra, parece que a fragmentação da realidade difere de uma língua para outra, como no exemplo seguinte:

PORTUGUÊS	ESPAÑHOL
(3) hálito	
fôlego	aliento
alento	

O que acontece, neste caso, é que estamos perante três acepções diferenciadas da palavra espanhola *aliento*, como podemos constatar nos exemplos seguintes:

- (4) (a) *Toma pastillas contra el mal aliento;*
 (b) *Llegó sin ningún aliento;*
 (c) *Aún le queda aliento para seguir viviendo.*

Cada uma dessas acepções será uma unidade léxica, que chamaremos **lexema**, (*aliento*₁, *aliento*₂ e *aliento*₃), definida e delimitada pelo contexto de uso, a que correspondem três **semantemas**, e às quais correspondem em português três palavras distintas (*hálito*, *fôlego* e *alento*), e obviamente também três lexemas. Portanto, a estruturação da realidade nos exemplos concretos é a mesma, só que o “conceito” será verbalizado, intensionalizado, nas diferentes línguas de forma diferente: através de palavras, frases, sintagmas, etc. Não se pode afirmar que as denominações de uma língua não têm equivalência exacta noutras línguas pelo facto de não se utilizarem as mesmas unidades para fazer tal denominação (*vd.* Haensch *et al.*, 1982: 29-30) ou tal acto de selecção do referente.

Temos, portanto, três lexemas diferenciados que correspondem a uma única palavra. Mas um lexema poderá ser também uma acepção claramente delimitada de um grupo de palavras.

É generalizado o sentimento de que não falamos por palavras. Mas também é verdade que não falamos por meio de orações ou de textos. Falamos antes por conjuntos de palavras⁴.

Um caso evidente é o das expressões idiomáticas, que, segundo Alonso Ramos (1993: 178), devido a algumas das suas propriedades morfo-sintácticas e semânticas (a sua não composicionalidade semântica, a memorização, a não produtividade, a aceitação de algumas regras de derivação, a equivalência a uma palavra) devem ser consideradas elas próprias como unidades léxicas (lexemas) e não como combinação de várias unidades. Assim, em expressões como *esticar a canela* ou *esticar o pernil*, não são os lexemas *esticar* e *canela* (ou *pernil*) que se combinam para produzir um determinado sentido (como em *comprar um livro*). Ambas as palavras formam um bloco que significa, na sua totalidade, ‘morrer’ (exemplo adaptado de Alonso Ramos, *ibidem*).

A unidade léxica que consideraremos aqui será, portanto, o **lexema**, entendido como a acepção claramente delimitada de uma palavra ou grupo de palavras, e o seu sentido, o **semantema**.

Ultrapassar os limites da palavra como unidade léxica traz consigo novos problemas, como por exemplo: Qual é a unidade de tratamento lexicográfico com que se deve trabalhar? Que combinações léxicas devem ser incluídas como unidades de tratamento lexicográfico? O que é uma locução⁵? O que é uma colocação?, etc. E também, ultrapassar os limites da palavra como unidade léxica,

problematiza ainda mais a questão da co-ocorrência léxica, pois, como acabamos de ver, as expressões idiomáticas talvez devam ser consideradas como unidades léxicas e, por isso, não deveriam ser incluídas no estudo da co-ocorrência léxica: se uma expressão idiomática é uma unidade léxica, não há, então, co-ocorrência de unidades léxicas (Alonso Ramos, 1993: 178). O mesmo não acontece com combinações de palavras como

<i>mudança radical</i>	(cf. * <i>mudança febril</i>);
<i>ódio mortal</i>	(cf. * <i>ódio cego</i>);
<i>correr um risco</i>	(cf. * <i>sofrer um risco</i>);
<i>dar um passeio</i>	(cf. * <i>fazer um passeio</i>);
<i>falar inglês</i>	(cf. ≠ <i>falar bem, falar devagar</i>);
<i>dar ouvidos</i>	(cf. * <i>fazer ouvidos</i>); etc.

Aqui já não estamos perante unidades léxicas, como no caso das expressões idiomáticas. Trata-se agora da ocorrência de duas ou mais unidades léxicas, cuja combinação não é totalmente livre. Estamos perante casos de combinatória léxica restringida (as chamadas colocações), como fica evidenciado com os casos agramaticais que acompanham cada exemplo (**fazer um passeio*, etc.).

Tanto as expressões idiomáticas como as colocações devem ser tratadas nos dicionários. É por isso que, nestes casos, falaremos em “unidades de tratamento lexicográfico”⁶ e não em “unidades léxicas”, incluindo assim os casos de combinatória léxica restringida.

Já nas primeiras manifestações lexicográficas peninsulares, como tive oportunidade de verificar num trabalho realizado para o professor Azevedo Ferreira no âmbito do mestrado em Ensino da Língua e Literatura Portuguesas, detectámos como se ultrapassam os limites da palavra como unidade de tratamento lexicográfico.

Com efeito, tanto do *Dictionarium ex lusitanico latinum sermonem* de Jerónimo Cardoso (de 1562), como no *Vocabulario de romance en latin* de Antonio de Nebrija (de 1516), as entradas nem sempre coincidem com a unidade “palavra”, registando-se como lemas diferentes acepções das mesmas, assim como acepções provenientes da combinação de palavras. Este fenómeno era devido, nestas primeiras mostras lexicográficas bilingues, à necessidade de abranger as diferentes acepções ou correspondências latinas. Ex.:

Exemplos de entradas de Hieronymi Cardosi (1565): *Dictionarium ex lusitanico latinum sermonem*. Joannus Alvari typographi [].

cada dia	...
cada dia mais	costa arriba
cada ora	costa abaixo
cadanno	costa
cada somana	
cada mes	
cada hum	(<i>et passim</i>)

Exemplos de entradas de Antonio de Nebrija (1516): *Vocabulario de romance en latin*. Sevilla.

cada dos años	cada uno dos
cada tres años	...
...	cada qual
cada dia	...
cada dia adverbio	caer como quiera
cada dos dias	caer a menudo
...	caer juntamente
cada mes	caer otra vez
...	caer de arriba
cada uno sendos	caer abaxo

(et passim)

Muitas vezes deparamos com o facto de que os dicionários não só não registam significados provenientes de combinações de palavras, mas também com o facto de que consideram como sendo uma acepção de uma palavra o que, em rigor, é o significado dessa palavra juntamente com outros elementos com os quais co-ocorre.

É o caso, por exemplo, da expressão *ter bom ouvido*. Como indica Calderón Campos (1994: 58), embora para o espanhol, não é a palavra *ouvido* que significa 'aptidão para captar sons musicais'. Esse é o significado da expressão *ter bom ouvido*. Vejamos o diferente tratamento desta expressão em três dicionários portugueses (o *Porto Editora*, o *Aurélio* e o *Caldas Aulete*):

ouvido, s. m. audição; aparelho de audição; acto ou efeito de ouvir; facilidade em fixar de memória peças musicais, ou em distinguir faltas de afinação; ... (*Porto Editora*)⁷

ouvido. [Part. de *ouvir*] S. m. 1. [...] 2. [...]. 3. Aptidão para captar com relativa precisão sons musicais ou não, e de reproduzir aqueles sem o auxílio de partitura: *ter bom ouvido*. ... 4. [...] • [...] **Ter bom ouvido**. Ter fácil percepção de sons, especialmente musicais. (*Aurélio*)⁸

ouvido, s. m. um dos cinco sentidos [...]. || [...] *Ter bom ouvido*, ter boa disposição do órgão do ouvido para perceber os sons, e especialmente os musicais. [...]. (*Caldas Aulete*)⁹

No *Porto Editora*, o que num contexto real é expresso pela locução *ter bom ouvido* é considerado como sendo uma acepção da palavra *ouvido*.

No caso do *Aurélio* regista-se uma contradição, pois o mesmo significado ('aptidão para captar com relativa precisão sons musicais') é atribuído tanto à palavra *ouvido* como à expressão *ter bom ouvido*, que aparece como locução no fim do artigo.

Não é o caso do *Caldas Aulete*, que, como já indicava Rodrigues Lapa¹⁰, continua a ser o melhor dicionário do Português quanto ao tratamento da combinação léxica. Neste caso, regista-se este valor unicamente como locução.

Como vimos, podem dar-se dois tipos de combinações de unidades léxicas: combinações livres e combinações restringidas.

Segundo Ygor Mel'čuk (da Universidade de Montréal), uma combinação livre AB é um sintagma composto de dois ou mais lexemas A e B, cujo significado é a soma regular (ou união linguística)¹¹ dos significados dos lexemas constituintes 'A ⊕ B' e cujo significante é a soma regular dos seus significantes /A ⊕ B/ (Mel'čuk, no prelo; Alonso Ramos, 1993: 156). Exemplos de combinações livres são os apresentados no início desta comunicação (2).

Um sintagma não livre, ou frasema, AB é uma combinação de dois ou mais lexemas A e B, cujo significante é a soma regular dos significantes dos lexemas constituintes /A ⊕ B/, mas cujo significado, ou cujo uso pragmático, é diferente da soma dos significados dos lexemas constituintes.

Estruturas do tipo *perder a cabeça*, *baixar a cabeça*, *andar à nora*, *ser o braço direito*, *dar um passeio*, *ódio mortal*, *mudança radical*, *leite gordo*, etc. são exemplos de frasemas.

Ygor Mel'čuk (*idem*) distingue, dentro deste tipo de combinatória restringida, entre frasemas pragmáticos ("pragmatemas") e frasemas semânticos. Aqui só me referirei a estes últimos, que poderão ser de três tipos: expressões idiomáticas (ou frasemas completos), colocações (ou semi-frasemas) e quase-frasemas.

1) Expressões idiomáticas (ou frasemas completos)

Uma expressão idiomática, ou frasema completo AB, é uma combinação de dois ou mais lexemas A e B, cujo significante é a soma regular dos significantes dos lexemas constituintes /A ⊕ B/, mas cujo significado não é a união regular de A e B ('A ⊕ B') mas um significado diferente 'C', que não inclui nem 'A' nem 'B'. (Mel'čuk, *idem*)¹².

São exemplos de expressões idiomáticas estruturas do tipo:

<i>perder a cabeça</i>	(descontrolar-se);
<i>baixar a cabeça</i>	(obedecer);
[andar] <i>à nora</i>	([andar] desorientado) ¹³ ;
[ser] <i>o braço direito</i>	(o principal auxiliar); etc.

Como indicámos, para Alonso Ramos (1993), neste caso não deveria falar-se em combinação de unidades léxicas, mas numa única unidade léxica.

2) Colocações (ou semi-frasemas)

Uma colocação, ou semi-frasema, AB é uma combinação de dois ou mais lexemas A e B, cujo significante é a soma regular dos significantes dos lexemas constituintes /A ⊕ B/, e cujo significado 'X' inclui o significado do lexema A mais um significado 'C' ('X' = 'A ⊕ C'), de tal maneira que o lexema B que exprime 'C' não é seleccionado livremente¹⁴.

Consoante a natureza de 'C', Mel'čuk (*idem*) distingue quatro tipos de colocações:

- «1) either 'C' ≠ 'B' i.e. does not have (in the dictionary) the corresponding signified;
- and [a. 'C' is empty, that is, the lexeme B is, so to speak, a semi-auxiliary used to support a syntactic configuration;

or b. 'C' is not empty but the lexeme **B** expresses 'C' only in combination with A (or with a few other similar lexemes);

- 2) or 'C' = 'B', i.e. **B** has (in the dictionary) the corresponding signified; and [a. 'B' cannot be expressed by any otherwise possible synonym; or b. 'B' includes (an important part of) the signified 'A', that is, it is utterly specific].»

São exemplos de colocações (sublinhada a base da colocação, isto é, o lexema que conserva o sentido intacto):

Tipo 1a: dar um *passeio*; infligir uma *derrota*;

Tipo 1b: *ódio mortal*;

Tipo 2a: *leite gordo*¹⁵;

Tipo 2b: *cabelo loiro*.

3) Quase-frasemas

Para Mel'čuk (*idem*), nos quase-frasemas conservam-se os sentidos dos lexemas que o constituem mas acrescenta-se um novo sentido que não é dedutível da simples soma dos sentidos dos lexemas constituintes¹⁶. São exemplos de quase-frasemas, *cinturão negro*, onde encontramos o sentido de 'cinto' e de 'negro' mais um sentido aproximado de 'grau de conhecimento ou habilidade em artes marciais'. O mesmo acontece com expressões como *centro comercial* ou *dar o peito* (dar + peito + alimentar) (Alonso Ramos, 1993: 189).

O dicionário bilingue de espanhol-português não poderá limitar-se a fornecer apenas uma simples listagem das palavras existentes numa língua e o seu equivalente na outra (pois alguém que não conhecesse suficientemente bem as duas línguas poderia concluir que as diferenças entre uma e outra seriam apenas de tipo gráfico ou pouco mais):

Leche	Leite
Entero	Inteiro
Conclusión	Conclusão
Hacer	Fazer
Tirar	Tirar
Fotocopia	Fotocópia
Estar	Estar
Prohibido	Proibido
Calor	Calor
Cuestión	Questão
Provecho	Proveito
Apetito	Apetite

As possibilidades colocacionais (usos sintáctico-semânticos) e os usos pragmáticos variam de língua para língua. É por isso que o dicionário bilingue deverá fornecer não só os diferentes equivalentes, mas também informar sobre os colocadores típicos das palavras e os contextos de uso das mesmas:

Leche entera	Leite gordo
Leche desnatada	Leite magro
Sacar una conclusión	Tirar uma conclusão
Hacer una fotocopia	Tirar uma fotocópia
Estar prohibido	Ser proibido
Hoy hace calor	Hoje está calor
Plantear una cuestión	Colocar uma questão
Buen provecho	Bom apetite ¹⁷ .

Como **conclusão**: a informação gramatical no dicionário não só é pertinente e necessária, mas deve ser valorizada e aumentada se concebermos um dicionário cuja finalidade será tanto a codificadora como a decodificadora. Evidentemente, isto será ainda mais importante no caso dos dicionários electrónicos e dos geradores de texto.

É urgente trabalhar num dicionário que permita ou facilite o trabalho de codificação, fornecendo informações precisas relativas à eleição, combinação e uso correctos das palavras e locuções num contexto dado (contexto linguístico e contexto situacional).

NOTAS

- ¹ A *intensão* da expressão será entendida aqui como a maneira como são seleccionadas as entidades, os traços ou os critérios considerados na hora de determinar a extensão da expressão (Allwood *et alii*, 1981: 4-5), quer dizer, os *bits* de informação ou componentes de significado utilizados (Lyons, 1986: 466-467).
- ² Embora seja mais utilizado – em filosofia, linguística ou matemática – o termo *objecto*, neste trabalho preferimos utilizar a palavra *coisa* devido ao valor genérico que esta possui no seu uso corrente em português. Assim, quando a palavra *coisa* aparecer em *itálico*, será usada neste sentido lato: «... (coisa, na acepção mais extensa desta palavra “coisa”): acções, processos, qualidades, estados, relações, abstrações, possibilidades, irrealidades.» (Lopes, 1972: 49).
- ³ As *coisas* para que remete o acto de referência não devem ser concebidas como existentes no mundo real, mas no *universo do discurso*, criado e partilhado pelos interlocutores, e entendido como *modelo* de um universo exterior ao sistema.
- ⁴ «PEOPLE SPEAK IN SET PHRASES – rather than in separate words; hence the crucial importance of set phrases. At the same time, set phrases, or phrasemes, represent one of the major difficulties in theoretical linguistics as well as in dictionary making.» (Mel'čuk, no prelo).
- ⁵ E o que é um modismo, uma frase feita, uma expressão idiomática, uma expressão fixa, uma lexia complexa, uma unidade fraseológica, um sintagma, etc.
- ⁶ Seguindo Hausmann, F. e H. E. Wiegand (1989) [*apud* WERNER, R., C. CHUCHUY (1992: 101), in WOTJAK, G. (ed.) (1992)].
- ⁷ ALMEIDA COSTA J. e A. SAMPAIO E MELO (1994): *Dicionário da Língua Portuguesa da Porto Editora*. 7ª edição, revista e ampliada. Porto: Porto Editora.
- ⁸ FERREIRA, A. Buarque de Holanda (1986): *Novo Dicionário da Língua Portuguesa*. 2ª edição, revista e aumentada, 20ª impressão. Rio de Janeiro: Nova Fronteira.
- ⁹ AULETE, F. J. Caldas (1987): *Dicionário da Língua Portuguesa Caldas Aulete*. 5ª edição brasileira, revista, actualizada e aumentada por Hamílcar de Garcia e Antenor Nascentes. Rio de Janeiro: Editora Delta.
- ¹⁰ «É precisamente neste capítulo da fraseologia, muito importante, que os dicionários correntes deixam mais a desejar. O mais celebrado de entre eles e o mais moderno dos grandes dicionários, o de Cândido de Figueiredo, é muito pobre em grupos fraseológicos, o que constitui um grave defeito, porque é nessas locuções que se imprime o chamado génio da língua. Como repositório

de fraseologia, nada há que possa substituir entre nós o *Dicionário Contemporâneo* de Caldas Aulete.» (Lapa, 1984: 83).

- 11 Para o autor, a soma regular ou operação de união linguística (\oplus):
 «putting together linguistic items according to general rules of L while constructing expressions of higher order. The symbol \oplus is reminiscent of arithmetical summation, but linguistic union is much more complex than simple addition: it presupposes observing ALL general rules of L, and this, in conformity with the nature of items being united (signified are united in a different way from signifiers and syntactics, etc.) Thus, $X \oplus Y$ denotes the regular union of signs X and Y; 'X' \oplus 'Y' is the regular union of signifieds 'X' and 'Y'; etc.
 Informally, a phraseme is a phrase that cannot be constructed both unrestrictedly and regularly using exclusively the operation \oplus .» (Mel'čuk, no prelo).
- 12 $AB = \langle 'C'; /A \oplus B/ \rangle \mid 'C' \neq 'A' \ \& \ 'C' \neq 'B'$ (Mel'čuk, no prelo).
- 13 Repare-se que a expressão idiomática é só "à nora"; "andar à nora" (*andar* desorientado) é uma colocação, ou semi-frasema, formada por um verbo colocativo (*andar*) e uma expressão idiomática "à nora".
- 14 « $AB = \langle 'A \oplus C'; /A \oplus B/ \rangle \mid 'C'$ is expressed by B which is not selected unrestrictedly» (Mel'čuk, no prelo).
- 15 Em espanhol, *leche entera* seria uma colocação do tipo 1b.
- 16 $AB = \langle 'A \oplus B \oplus C'; /A \oplus B/ \rangle \mid 'C' \neq 'A' \ \& \ 'C' \neq 'B'$ (Mel'čuk, no prelo).
- 17 Cf. 'Bom proveito', 'Que te faça bom proveito' (espanhol: *Que te aproveche!*), com conotações negativas ou maior subjectividade (não participação do locutor?).

BIBLIOGRAFIA

- ALONSO RAMOS, M. (1993): *Las Funciones Léxicas en el modelo lexicográfico de I. Mel'čuk* (tese de doutoramento). Madrid: UNED.
- ALLWOOD, J., L.-G. ANDERSSON, Ö. DAHL (1981): *Logic in Linguistics*. Cambridge: Cambridge University Press. [Lund, 1971].
- CALDERÓN CAMPOS, M. (1994): *Sobre la elaboración de diccionarios monolingües de producción. Las definiciones, los ejemplos y las colocaciones léxicas*. Granada: Universidad de Granada.
- CRYSTAL, D. (1987): *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- HAENSCH, G., L. WOLF, S. ETTINGER e R. WERNER (1982): *La lexicografía. De la lingüística teórica a la lexicografía práctica*. Madrid: Gredos.
- LAPA, M. Rodrigues (1984): *Estilística da língua portuguesa*. Coimbra: Coimbra Editora.
- LOPES, Ó. (1972): *Gramática simbólica do português*. 2ª Ed (corrigida). Lisboa: Fundação Calouste Gulbenkian.
- LYONS, J. (1986): *Introducción en la lingüística teórica*. Barcelona: Teide [Cambridge University Press, 1968].
- MEL'ČUK, Y. (no prelo): "Collocations and Lexical Functions"
- WERNER, R., CHUCHUY, C. (1992): "¿Qué son los equivalentes en el diccionario bilingüe?", WOTJAK, G. (ed.) (1992), 99-107
- WOTJAK, G. (ed.) (1992): *Estudios de lexicología y metalexigrafía del español actual*. Tübingen: Max Niemeyer Verlag.