

POLISSEMIA: LUMPING E SPLITTING NOS LEARNER'S DICTIONARIES E A SEMÂNTICA COGNITIVA

Ana Flávia Souto de Oliveira¹

anaflavia10@gmail.com

RESUMO: Na Lexicografia, a polissemia impõe diversas questões ao trabalho do lexicógrafo. Uma delas diz respeito à delimitação dos significados que um item lexical possui. De tal questão, primariamente lexicológica, surgem dois fenômenos lexicográficos conhecidos como '*lumping*' e '*splitting*'. *Grosso modo*, o primeiro se relaciona à escolha por manter significados distintos agrupados em uma só acepção e o segundo consiste em separar os significados em diferentes acepções. Neste trabalho, avaliamos dois verbetes de *learner's dictionaries* a fim de verificar como tais fenômenos estão presentes nessas obras. A partir da concepção semântico-cognitiva de polissemia, propomos que essa teoria semântico-lexical permite compreender tais fenômenos de forma mais apropriada para seu tratamento no âmbito lexicográfico e auxilia na organização das acepções nos verbetes.

PALAVRAS-CHAVE: Lexicografia; polissemia; *lumping*; *splitting*.

INTRODUÇÃO

A polissemia, fenômeno no qual um item lexical apresenta dois ou mais significados relacionados, mesmo que geralmente não imponha problemas à linguagem cotidiana, traz questões de fundamental relevância à Lexicografia. Por exemplo, para o trabalho lexicográfico, é necessário (i) delimitar quantos e quais são os significados do item lexical a serem incluídos no verbete, problema intimamente relacionado a um método de análise lexicológica; (ii) descrever o conteúdo de cada um desses significados do item lexical, questão referente à problemática da definição; e, por fim, (iii) optar por uma forma de organizar esses significados, ou seja, definir de que modo agrupá-los e ordená-los no verbete (RAVIN; LEACOCK, 2000).

A primeira dessas tarefas, a delimitação dos significados do item lexical, é um assunto que recebe bastante atenção nas discussões semânticas, não apenas no âmbito da

¹ Doutoranda do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul – UFRGS. Bolsista CAPES.

Lexicografia, mas ocupando uma posição importante também em disciplinas como a Linguística de Corpus, o Processamento da Linguagem Natural (PLN) e a própria Semântica Lexical. Nesse viés, destacam-se, por exemplo, em PLN, os estudos relacionados à identificação desses significados, como os que buscam aprimorar os métodos para seu reconhecimento automático (cf. Tugwell e Kilgarriff, 2000).

A relação multidisciplinar entre essas áreas pode ser exemplificada pela seguinte situação: o lexicógrafo se utiliza de métodos lexicológicos para delimitar quais são os significados de um item lexical. Esses significados são utilizados para abastecer um programa de PLN, que buscará classificar ocorrências de uso em um *corpus*, a fim de medir, por exemplo, a frequência de uso desses significados em tal *corpus*. Porém, na Lexicografia, tal tarefa não se dá de forma simples, como pode ser facilmente observado em comparações entre verbetes de diferentes dicionários para o mesmo item lexical. Abaixo, apresentamos os verbetes do item lexical *accident* trazido em quatro *learner's dictionaries*² (as cores marcam significados considerados equivalentes):

CALD (2008) ³	LDCE (2009)	COBUILD (2006)	OALD (2005)
1 [C] something bad which happens that is not expected or intended, and which often damages something or injures someone	2 [C] an event in which a car, train, plane etc is damaged and often someone is hurt	1 N-COUNT An accident happens when a vehicle hits a person, an object, or another vehicle, causing injury or damage.	1 [C] an unpleasant event, especially in a vehicle, that happens unexpectedly and causes injury or damage
	3 [C] a situation in which someone is injured or something is damaged without anyone intending them to be	2 N-COUNT If someone has an accident, something unpleasant happens to them that was not intended, sometimes causing injury or death.	2 [C,U] something that happens unexpectedly and is not planned in advance
	4 [C, U] something that happens without anyone planning or intending it		

Quadro 1: Verbetes de *accident* em CALD (2008), LDCE (2009), COBUILD (2006) e OALD (2005)

² Os *learner's dictionaries* são dicionários monolíngues de inglês para falantes não nativos. Optamos por manter o termo, em vez de utilizar a tradução 'dicionários de aprendizes', pelo fato dessa denominação já ser consagrado na Metalexigrafia para designar especificamente esse genótipo lexicográfico. Além disso, no português, o termo 'dicionário pedagógico', outro possível equivalente, abarca diversos tipos de obras, desde os dicionários para aprendizes de língua materna até obras bilíngues voltadas ao público aprendiz, bastante diversas das que tratamos aqui.

³ Seguindo a tendência da metalexigrafia europeia, os dicionários serão representados pelo uso de siglas, seguidos do ano de publicação.

Os verbetes foram classificados a partir das seguintes glosas: “acontecimento inesperado”, “acontecimento com danos” e “acontecimento com danos envolvendo veículo”. As marcações referentes às cores são apresentadas abaixo:

Significados	Cor
“acontecimento inesperado”	XXXXXX
“acontecimento com danos”	XXXXXX
“acontecimento com danos envolvendo veículo”	XXXXXX

Quadro 2: Correspondência entre significados de *accident* e acepções apresentadas pelos dicionários

A partir desses dados, percebemos que (i) os dicionários trazem diferentes significados para *accident*, como o explicitado pelo fato de CALD (2008) não trazer acepção para “acontecimento com danos envolvendo veículo”, e (ii) nenhuma das obras separa e relaciona as acepções de forma idêntica: CALD (2008) traz uma única acepção para duas glosas; LDCE (2009) apresenta três acepções separadas, uma para cada glosa; e COBUILD (2006) e OALD (2005) relacionam duas glosas distintas na mesma acepção. Tal quadro materializa a afirmação de que “mesmo para lexicógrafos, não existem critérios objetivos para a análise de uma palavra em diferentes significados”⁴ (FILLMORE; ATKINS, 2000, p.101).

No caso de CALD (2008), no qual dois significados são apresentados numa única definição, temos uma *catch-all definition*, consequência do fenômeno lexicográfico chamado de *lumping*⁵. Já em LDCE (2009), no qual todos os significados recebem uma acepção distinta, temos um caso de *splitting*. A partir dessas considerações, neste trabalho, abordaremos os fenômenos de *lumping* e *splitting* e suas implicações para a organização das acepções nos verbetes, avaliando a forma como eles ocorrem nos *learner’s dictionaries* e buscando compreender esses fenômenos a partir da concepção de polissemia desenvolvida pela Semântica Cognitiva.

Na primeira parte do trabalho, definimos os conceitos de *lumping* e *splitting*, fornecendo exemplos extraídos de diferentes tipos de dicionários e considerando o reflexo desses fenômenos a partir das variáveis ‘tipo de obra’ e ‘perfil de usuário’. Na segunda parte, avaliamos verbetes de dois itens lexicais em *learner’s dictionaries*. Na seção três, introduzimos a concepção de polissemia da Semântica Cognitiva. Na parte quatro, buscamos avaliar os dados obtidos nas análises à luz dos postulados teóricos apresentados anteriormente e sua consequência para a organização dos verbetes dos dicionários.

⁴ [even for lexicographers there are no objective criteria for the analysis of a word into senses], tradução minha.

⁵ Neste trabalho, mantemos as designações do inglês para *lumping* e *splitting*, por julgarmos que não há uma tradução adequada para o português que capture a especificidade desses termos.

1. LUMPING E SPLITTING

Dados de uma pesquisa conduzida por Kilgarriff (1998) com lexicógrafos apontaram que a identificação dos significados de um item lexical foi considerada a segunda tarefa mais difícil pelos entrevistados quando da elaboração de um dicionário⁶. Além disso, essa dificuldade é agravada pelo fato de ela ser

uma questão para a qual os livros-texto não têm nada para nos dizer, e para a qual o treinamento dos lexicógrafos é inteiramente baseado em exemplos, em grande parte pelas bases lógicas para realizar *lumping* e *splitting* serem tão pouco compreendidas, mesmo por lexicógrafos que realizam um bom trabalho⁷ (KILGARRIFF, 1998, p.52).

Os fenômenos chamados de *lumping* e *splitting* constituem, dessa forma, uma questão central a ser resolvida ao identificar os significados dos itens lexicais. *Lumping* é definido por Fontenelle (2011, p.57) como o processo de “considerar dois padrões de uso levemente distintos como um único significado”⁸, enquanto que *splitting* ocorre “quando o lexicógrafo separa padrões de uso levemente diferentes em significados distintos”⁹. Porém, tal definição não torna a compreensão desses fenômenos menos complexa, visto que não especifica o que seria um uso “levemente distinto”.

Hanks (2008, p.127) reafirma o papel do lexicógrafo nessa decisão, ao dizer que os “lexicógrafos são às vezes classificados em “*lumpers*” e “*splitters*”: aqueles que preferem – ou melhor, que são forçados por questões de mercado – a unir usos em uma única acepção e aqueles que isolam distinções esmiuçadas”¹⁰. Dessa afirmação, fica evidente o caráter subjetivo que pode adquirir o tratamento desses fenômenos num projeto lexicográfico, ainda mais, tendo em vista, como Hanks destaca, as restrições editoriais impostas ao lexicógrafo.

Para desenvolver tais conceitos, podemos retomar os dois quadros apresentados anteriormente: aparentemente, as três glosas que sugerimos no Quadro 2 são suficientes para classificar os significados trazidos pelos dicionários do Quadro 1. Porém, abaixo

⁶ De acordo com a mesma pesquisa, redigir a definição [*wording*] foi considerada a tarefa mais difícil pelos lexicógrafos entrevistados.

⁷ [a matter on which textbooks have nothing to tell us, and lexicographers’ training is entirely example-based, largely because rationales for lumping and splitting are so little understood, even by lexicographers who do the job well], tradução minha.

⁸ [considering two slightly different patterns of usage as a single meaning], tradução minha.

⁹ [when the lexicographer separates slightly different patterns of usage into distinct meanings], tradução minha.

¹⁰ [lexicographers are sometimes classified into “*lumpers*” and “*splitters*”: those who prefer – or rather, who are constrained by marketing considerations – to lump uses together into a single sense, and those who isolate fine distinctions], tradução minha.

apresentamos verbetes de outros quatro tipos de dicionários para o item lexical *accident*, que dificultam essa classificação:

OBED (2006)	OALD (2005)	OCDCE (2005)	CEDCa (2005)
something bad that happens by chance [...]	1 an unpleasant event, especially in a vehicle, that happens unexpectedly and causes injury or damage 2 something that happens unexpectedly and is not planned in advance	1 an event that is without apparent cause, or is unexpected. 2 an unfortunate event brought about unintentionally. 3 an automobile collision or crash.	1 an unforeseen event or one without an apparent cause 2 anything that occurs unintentionally or by chance 3 a misfortune or mishap, esp one causing injury or death

Quadro 3: Definições do item *accident* presentes nos dicionários OBED (2006), OALD (2005), CEDCa (2005) e OCDCE (2005)

Desconsiderando a diferença no modo de agrupar e apresentar os significados, percebemos que, se tentarmos utilizar as mesmas glosas apresentadas no Quadro 2 para classificar tais acepções, isso não seria possível. Haveria a necessidade de incorporar a glosa “acontecimento não intencional” para caracterizar a acepção de número 2 do OCDCE (2005), pois o dicionário a separa de “acontecimento inesperado”, que é trazido na acepção 1. Além disso, tal glosa também se encaixaria na acepção trazida por OBED (2006) e na acepção 2 de CEDCa (2005).

Da mesma forma, teríamos que incluir a glosa “acontecimento sem causa aparente” para abarcar a especificidade das acepções 1 do OCDCE (2005) e do CEDCa (2005). Já a nuance “acontecimento ruim”, que parece perpassar diversos dos significados, pode ser encarada como um traço semântico, não como uma glosa, pois sozinha não é suficiente para distinguir *accident* de outros itens lexicais com tal traço, como *mishap* ou *misfortune*, por exemplo. Contudo, ela impõe a pergunta: seriam “acontecimento ruim inesperado” e “acontecimento inesperado” dois significados distintos? Por exemplo, ao classificar a ocorrência abaixo:

*Her pregnancy was an accident*¹¹

Claramente, a ocorrência carrega consigo a ideia de “acontecimento inesperado”, porém, dependendo do contexto, poderia também ser classificada como um “acontecimento ruim inesperado”, como na ocorrência *Her pregnancy was an accident that will not happen again*. Assim, parece clara a dificuldade não apenas em definir quais significados o item lexical possui, mas também em delimitar o conteúdo específico de cada um deles. Dessa

¹¹ Essa sentença foi atestada em 42.900 ocorrências no Google.

forma, pode ser justificada a separação das acepções “1. an event that [...] is unexpected” e “2. an unfortunate event brought about unintentionally” em OCDCE (2005).

Na tentativa oposta, de integrar diferentes significados em uma só acepção, é notável o papel do uso das conjunções, como *or* e *and*, como recurso para unir significados aparentemente, por que não “levemente”, distintos em uma única acepção, fato já explicitado por Geeraerts (2001). Por exemplo, a acepção de OCDCE (2005) “an event that is without apparent cause, *or* is unexpected” (grifo nosso), que une as glosas “acontecimento sem causa aparente” e “acontecimento inesperado”. Já na definição de OALD (2005) “an unpleasant event, especially in a vehicle, that happens unexpectedly *and* causes injury or damage” (grifo nosso), três glosas são apresentadas juntas, duas delas unidas pela conjunção *and*.

Assim, ampliando as glosas e classificando os significados apresentados pelos dicionários, teríamos dois novos quadros:

OBED (2006)	OALD (2005)	OCDCE (2005)	CEDCa (2005)
something bad that happens by chance [...]	1 an unpleasant event, especially in a vehicle, that happens unexpectedly and causes injury or damage 2 something that happens unexpectedly and is not planned in advance	1 an event that is without apparent cause, or is unexpected. 2 an unfortunate event brought about unintentionally. 3 an automobile collision or crash.	1 an unforeseen event or one without an apparent cause 2 anything that occurs unintentionally or by chance 3 a misfortune or mishap, esp one causing injury or death

Quadro 4: Definições do item *accident* presentes nos dicionários OBED (2006), OALD (2005), CEDCa (2005) e OCDCE (2005) com a classificação dos significados

Significados	Cor
“acontecimento inesperado”	XXXXXX
“acontecimento com danos”	XXXXXX
“acontecimento com danos envolvendo veículo”	XXXXXX
“acontecimento não-intencional”	XXXXXX
“acontecimento sem causa aparente”	XXXXXX

Quadro 5: Correspondência entre significados de *accident* e acepções apresentadas pelos dicionários

Nesse viés, perguntar qual o limite de distinções a serem feitas a respeito de quantos e quais significados um item lexical apresenta parece uma questão de caráter subjetivo, para a qual podem existir diferentes respostas, todas igualmente aceitáveis. Ademais, entra aqui também avaliar os limites de uma distinção clara entre casos de polissemia e vagueza.

Essa discussão se relaciona ainda ao tipo de dicionário em questão e ao perfil do usuário da obra, como o fato de OBED (2006), dicionário para aprendizes estrangeiros de nível básico, trazer apenas um significado para o item em questão e OALD (2005), dicionário para aprendizes estrangeiros de nível avançado, trazer duas acepções, que abarcam diferentes

glosas. Entendemos assim que a análise dessas questões deve passar também pela consideração dessas duas variáveis, como Kilgarriff aponta:

Splitting é em grande parte uma tarefa analítica. O lexicógrafo avalia como o comportamento da palavra deve ser analisado em significados distintos em princípio antes de começar a escrever, embora o número de distinções a serem feitas variem de acordo com o tamanho e estilo do dicionário (KILGARRIFF, 1998, p.51)¹².

Além do caráter analítico dessas considerações a respeito do *lumping* e *splitting*, há ainda uma parte prática importante a ser ressaltada para o trabalho lexicográfico: a delimitação dos significados de um item lexical tem influência direta na organização interna do verbete, na forma como as acepções serão dispostas e relacionadas. Isso porque, dependendo do critério que o dicionário utiliza para ordenar as acepções (como o critério etimológico e o de frequência), a quantidade de acepções e seus conteúdos semânticos – ou seja, como resolver casos de *lumping* e *splitting* – determinarão a organização microestrutural.

Para tornar essa asserção mais concreta, podemos utilizar como exemplo a avaliação de uma ordenação das acepções feita pela frequência. Dependendo do modo como os significados são percebidos pelo lexicógrafo, a avaliação das ocorrências num *corpus* pode mudar (OLIVEIRA, 2010), pois, por exemplo, considerar definições como a 1 de OALD (2005), “an unpleasant event, especially in a vehicle, that happens unexpectedly and causes injury or damage”, e a 3 de OCDCE (2005) de forma autônoma, “an automobile collision or crash”, para classificar ocorrências faz toda a diferença para a frequência absoluta de cada agrupamento de significados e levaria a resultados diferentes.

É importante, ainda, levar em consideração a classe gramatical e o funcionamento sintático do item lexical, pois eles excluem a possibilidade de alguns significados serem considerados compatíveis para a apresentação em uma só acepção. Nesse caso, adotar um viés de *splitting* traria benefício para as especificações sintáticas dos itens lexicais analisados.

¹² [Splitting is largely an analysis task. The lexicographer works out how the word’s behaviour is to be analysed into distinct senses in principle before starting to write, though the number of distinctions to be made will vary with dictionary size and style], tradução minha.

2. ANÁLISES NOS DICIONÁRIOS

2.1 MÉTODO

Para o presente estudo, avaliamos verbetes de quatro *learner's dictionaries*, o CALD (2008), o COBUILD (2006), o LDCE (2009) e o OALD (2005). Essas obras foram escolhidas por serem os principais dicionários monolíngues de inglês para aprendizes não nativos disponíveis no mercado. Dessas obras, foram retirados dois verbetes escolhidos de modo aleatório¹³, sendo um adjetivo (*fresh*, cuja análise repertoria apenas as acepções relativas a alimentos e flores) e um substantivo (*reason*, que recebeu uma análise integral).

Para determinar os significados apresentados por cada item lexical, realizamos uma avaliação do mesmo verbo em todos os dicionários e buscamos delimitar quais significados eram apresentados em cada uma das acepções. De forma a facilitar a comparação das acepções trazidas pelos dicionários, construímos tabelas com os verbetes e, com a utilização de cores, fizemos a marcação dos significados que julgamos ser equivalentes.

Após realizar tal delimitação, comparamos a distribuição dos significados nas acepções presentes nos verbetes e buscamos avaliar os casos onde havia *lumping* e *splitting*.

2.2 ANÁLISES

2.2.1 FRESH

Para o item *fresh*, avaliamos apenas as acepções referentes a ‘alimentos’ e a ‘flores’, visto que, nos verbetes dos dicionários, esses significados eram os únicos que apresentavam casos compatíveis com uma avaliação dos fenômenos de *lumping* e *splitting*. Isso porque todas as outras acepções desse item lexical foram trazidas pelas obras em definições de forma independente, como as compreendidas pelas glosas “novo”, “interessante por ser novo” e “recente” (para uma análise detalhada dos verbetes trazidos pelos *learner's dictionaries* para esse item lexical, cf. Oliveira, 2010).

Assim, as acepções trazidas pelos dicionários são apresentadas no Quadro 6. Para a análise das ocorrências relativas a ‘alimentos’ e ‘flores’, o item em questão apresenta os

¹³ Na verdade, retiramos esses verbetes de uma análise realizada anteriormente em Oliveira (2010), na qual o propósito da análise era avaliar a organização geral das acepções nos verbetes. Nesse trabalho anterior, tais itens foram escolhidos de forma aleatória, apenas tendo em vista se o item lexical era polissêmico.

significados repertoriados no Quadro 7, que abarcam todas as nuances de significados trazidas pelos dicionários (de forma conjunta) e que parecem ser supostamente diferenciáveis. Por esse motivo, no mesmo quadro, são apresentados exemplos de ocorrências¹⁴ que ilustram o uso isolado de cada uma das glosas.

CALD (2008)	COBUILD (2006)	LDCE (2009)	OALD (2005)
5 NATURAL (of food or flowers) in a natural condition rather than artificially preserved by a process such as freezing	3 Fresh food has been picked or produced recently, and has not been preserved, for example by being frozen or put in a tin.	5 FOOD/FLOWERS a) fresh food has recently been picked or prepared , and is not frozen or preserved b) fresh flowers have recently been picked	FOOD 1 (usually of food) recently produced or picked and not frozen, dried or preserved in tins or cans

Quadro 6: Acepções de *fresh* referentes a ‘alimentos’ e a ‘flores’

Significado	Exemplo
“recém colhido”	* Baptisia australis has indigo-blue flowers and superbly soft, bluish-grey foliage which is wonderful to use fresh .
“recém preparado”	* Fish and chips are a Northern favourite and always fresh here. * Serve with fresh bread
“não preservado artificialmente”	* [...] 104 varieties of fresh and frozen pasta [...]. * [...] meringue basket filled with fresh or frozen raspberries [...].

Quadro 7: Significados distinguíveis de *fresh* que se relacionam a alimentos e suas ocorrências

Esses exemplos poderiam dar a impressão de que é possível, a partir das ocorrências de uso, diferenciar claramente entre os significados que os dicionários apresentam de forma integrada na mesma definição, o que levaria a uma contestação dos motivos que fizeram os lexicógrafos adotarem, na maior parte das definições, o procedimento de *lumping*. Porém, em diversas outras ocorrências¹⁵, um ou mais significados são acionados ou são difíceis de serem explicitamente reconhecidos ou diferenciados. Por exemplo, *fresh basil* [manjeriço fresco] em:

Pesto [...] makes a useful substitute for fresh basil in a vinaigrette dressing [o pesto [...] torna-se um ótimo substituto para o manjeriço fresco em um molho vinagrete] (British National Corpus)

possivelmente recebe a leitura manjeriço “não conservado artificialmente”, não desidratado, mas pode também se referir a manjeriço “recém colhido”.

Retornando aos verbetes, no LDCE (2009), aceção 5, aponta que *fresh* pode se referir, em **a**, a “alimentos recém colhidos”, “alimentos recém preparados” e a “alimentos não preservados artificialmente”; e, em **b**, a flores “recém colhidas”. Assim, a obra opta por

¹⁴ Os exemplos foram retirados do *British National Corpus* (BNC) e do *Collins WordbanksOnline English corpus* (CWB).

¹⁵ Para uma análise estatística dessas ocorrências, cf. Oliveira, 2010, p.124-133.

apresentar três significados em uma só acepção, utilizando as conjunções *and* e *or*. O OALD (2005) apresenta as significações “alimentos recém preparados”, “alimentos recém colhidos” e “alimentos não preservados artificialmente”, todas na mesma acepção. Já o CALD (2008) apresenta apenas a acepção “não preservados artificialmente”, tanto em relação a alimentos quanto em relação a flores.

Nesse sentido, cabe destacar que os significados referentes a flores são apresentados, nos dicionários analisados, com diferenças significativas. Por um lado, o LDCE (2009) traz a nuance “recém colhida”, por outro, o CALD (2008) apresenta a nuance “natural” em oposição à “preservada artificialmente”. OALD (2005) e COBUILD (2006), contudo, não apresentam nenhuma informação explícita quanto a flores. Assim, a partir dessas considerações, constatamos que o único ponto em comum entre os quatro dicionários, neste viés, é a acepção “alimentos não preservados artificialmente”.

Essas questões estão em sintonia com os problemas discutidos anteriormente: de que forma delimitar quantos e quais são os significados de um item lexical. Assim, notamos que os significados, na maioria dos verbetes, são apresentados integrados na mesma definição – como o caso de *fresh* com os significados “alimento recém colhido”, “recém produzido” e “não conservado artificialmente”, em OALD (2005) e COBUILD (2006) – ou seja, um caso de *lumping*.

2.2.2 REASON

Para avaliar as acepções apresentadas pelos dicionários para *reason*, utilizamos todas as que eram apresentadas nos verbetes (Quadro 8).

CALD (2008)	COBUILD (2006)	LDCE (2009)	OALD (2005)
EXPLANATION 1 [C or U] the cause of an event or situation or something which provides an excuse or explanation	IN-COUNT The reason for something is a fact or situation which explains why it happens or what causes it to happen.	1 CAUSE [C] why something happens, or why someone does something	1 [C] a cause or an explanation for sth that has happened or that sb has done
JUDGEMENT 3 [U] the ability of a healthy mind to think and make judgments, especially based on practical facts	2 N-UNCOUNT If you say that you have reason to believe something or to have a particular emotion, you mean that you have evidence for your belief or there is a definite cause of your feeling.	2 GOOD OR FAIR [U] a fact that makes it right or fair for someone to do something	2 [U] a fact that makes it right or fair to do sth
	3 N-UNCOUNT The ability that people have to think and to make sensible judgments can be referred to as reason .	4 GOOD JUDGMENT [U] sensible judgment and understanding	3 [U] the power of the mind to think in a logical way, to understand and have opinions, etc.
		7 ABILITY TO THINK [U] the ability to think, understand, and form judgments that are based on facts	4 [U] what is possible, practical or right

Quadro 8: Verbetes de *reason*

Identificamos, nos dicionários, quatro significados, que são apresentados no Quadro 9, com as respectivas marcações de cores.

Significados	Cor
“causa contável”	
“causa não contável”	
“razoabilidade”	
“raciocínio”	

Quadro 9: Correspondência entre significados de *reason* e acepções apresentadas pelos dicionários

Notamos que o CALD (2008) separa os significados em duas acepções, uma incluindo as glosas “causa contável” e “causa incontável” e outra unindo “razoabilidade” e “raciocínio”. Nesse caso, então, o dicionário demonstra uma tendência ao procedimento de *lumping*, inclusive desconsiderando o caráter contável ou incontável dos dois primeiros significados, que poderia ter sido usado para justificar uma divisão em acepções distintas, ou seja, *splitting*. COBUILD, por sua vez, separa “causa contável” e “causa incontável”, mas apresenta uma única acepção para as glosas “razoabilidade” e “raciocínio”. As outras duas obras, LDCE (2009) e OALD (2005), trazem uma acepção para cada um dos significados, caracterizando, assim, o processo de *splitting*.

Dessa forma, mesmo que a ordem na qual os significados aparecem seja idêntica em três obras – CALD (2008), COBUILD (2006) e OALD (2005) –, percebemos que em apenas duas delas a apresentação das acepções segue o mesmo padrão quanto à delimitação dos significados do item lexical. Isso pode ser melhor visualizado no quadro abaixo:

CALD (2008)	COBUILD (2006)	LDCE (2009)	OALD (2005)
“causa contável” “causa não-contável”	“causa contável”	“causa contável”	“causa contável”
“raciocínio” “razoabilidade”	“causa não contável”	“causa não contável”	“causa não contável”
	“raciocínio” “razoabilidade”	“razoabilidade”	“raciocínio”
		“raciocínio”	“razoabilidade”

Quadro 10. Divisão das acepções de *reason* nos quatro dicionários

Mais uma vez, a forma como cada obra explicita o entendimento da polissemia do item é distinta, o que, claramente, demonstra a dificuldade em decidir pela adoção de um dos dois procedimentos, *lumping* ou *splitting*.

3. SEMÂNTICA COGNITIVA LEXICAL E POLISSEMIA

A polissemia pode ser definida como o fenômeno no qual dois ou mais significados relacionados são associados a uma mesma forma linguística (TAYLOR, 2003). Nesses casos, a relação mantida entre os significados pode ser descrita com base na noção de motivação, na qual um significado dá origem a outro. Para a Semântica Cognitiva, essa motivação se dá por princípios cognitivos gerais, de modo que uma definição mais precisa de polissemia para esse paradigma seria a de perceber os itens lexicais individualmente “como categorias de significados relacionados que são motivados por princípios cognitivos como metáfora, metonímia, generalização, especialização e transformações por esquemas de imagem”¹⁶ (CUYCKENS; ZAWADA, 2001, p.xiv).

Na abordagem da semântica cognitiva lexical, os modelos de descrição da polissemia são construídos a partir da extensão da Teoria dos Protótipos desenvolvida na psicologia cognitiva por Eleanor Rosch (ROSCH, 1999), na qual, a partir de uma série de estudos experimentais, Rosch e sua equipe avaliaram de que forma objetos físicos eram percebidos e categorizados por seres humanos. Segundo Geeraerts (2006b, 2010), as pesquisas

¹⁶ [categories of related senses which are motivated by cognitive principles such as metaphor, metonymy, generalization, specialization, and image-schema transformations], tradução minha.

desenvolvidas por Rosch resultaram em duas direções de estudo que atribuem importância fundamental aos postulados da Teoria dos Protótipos: a psicolexicologia formal (com a busca de modelos formais para a representação da memória e operações conceituais) e as aplicações na linguística. Nessa última área, diversos pesquisadores buscaram estender os achados de Rosch, que eram inicialmente para objetos físicos, para a descrição de fenômenos linguísticos, principalmente no que diz respeito ao estudo da estrutura de categorias linguísticas.

Segundo Geeraerts (2006a; 2010), quatro tipos de fenômenos são tradicionalmente associados à prototipicidade: (1) a existência de graus de tipicidade, ou seja, nem todos os membros são igualmente representativos de uma categoria; (2) a presença de uma estrutura de semelhança de família [*family resemblance*]¹⁷; (3) a falta de definição [*blurred*] nos limites da categoria e (4) a falta de uma definição em termos de um conjunto único de atributos (necessários e suficientes). O autor propõe a avaliação dos tipos de efeitos prototípicos a partir de duas dimensões, de um lado a distinção entre uma caracterização extensional e intensional e, de outro, a não igualdade (ou seja, a diferença de status entre os membros da categoria) e a não discricção (a dificuldade e a flexibilidade de demarcação). Essas características são sintetizadas no quadro abaixo:

	EXTENSIONALMENTE (a nível referencial)	INTENSIONALMENTE (a nível dos significados)
NÃO IGUALDADE (efeitos de saliência, estrutura interna centro+periferia)	(1) graus de representatividade entre os membros de uma categoria;	(2) agrupamentos em semelhanças de família
NÃO DISCRICÇÃO (problemas de demarcação, flexibilidade)	(3) flutuações nas margens de uma categoria;	(4) impossibilidade de definições em termos de “condições necessárias e suficientes”.

Quadro 11. Efeitos prototípicos (adaptado de Geeraerts, 2010, p.189)

Tais características, como salienta Geeraerts (2010), não precisam ocorrer conjuntamente para que uma categoria seja caracterizada como prototípica. Ao invés disso, a presença de um ou mais desses fatores já indicaria a existência de efeitos prototípicos em sua estrutura. Assim, Geeraerts aponta para uma prototipicidade da própria concepção de protótipo (GEERAERTS, 2006a), com casos mais representativos que outros.

Como exemplo, apresentamos a análise do item lexical *fruit* desenvolvida por Geeraets, 2001; 2010). Quanto à dimensão extensional, a característica (1) pode ser entendida a partir de estudos experimentais que apontaram que, quando o significado cotidiano “parte

¹⁷ Como Geeraerts (2010, p.187) explica, “a estrutura semântica [dessas categorias] adquire a forma de um grupo radial de leituras agrupadas que se entrecruzam” [their semantic structure takes the form of a radial set of clustered and overlapping readings], tradução minha.

comestível doce e suculenta de uma planta, que contém sementes” estava em jogo, sujeitos americanos entendiam que a maçã era uma fruta mais típica do que, por exemplo, o romã, ou seja, eles julgavam que os elementos incluídos nessa leitura de *fruit* apresentavam diferentes graus de representatividade. A partir desse exemplo, fica claro também que a prototypicalidade e o pertencimento à categoria são duas propriedades distintas: mesmo que romã não seja uma fruta típica, é assim mesmo uma fruta. Já a segunda característica extensional, de número (3), pode ser exemplificada pelo caso da azeitona. Seria azeitona uma fruta? A partir do significado em questão, seria difícil delimitar se azeitona é ou não um membro, mesmo que um significado de caráter técnico, “a parte de uma planta ou árvore que contém semente”, autorize prontamente tal caracterização.

Quanto aos efeitos prototípicos intensionais, se considerarmos novamente a definição cotidiana “parte comestível doce e suculenta de uma planta, que contém sementes” para *fruit* e os elementos banana, morango, maçã e limão, podemos perceber que mesmo elementos que intuitivamente entram nessa leitura não podem ser descritos com tal definição. Tradicionalmente, as plantas que contém frutas são árvores, o que não se aplica para morango. A banana, por sua vez, não é uma fruta que seja “suculenta”. O limão não apresenta a característica “doce”. Desse modo, apenas maçã poderia ser descrita por essa definição de um modo tradicional, com características necessárias e suficientes. Assim, percebemos que a característica prototípica (4) se aplica também a esse significado de *fruit*, pois o único traço comum que sobra da definição, “parte comestível que contém semente”, não é suficiente para distinguir *fruit*, por exemplo, de vagem. Desse modo, segundo Geeraerts (2010, p.190), “os atributos que entram na descrição semântica de *fruit* demarcam vários subgrupos da gama total de aplicação de *fruit*”, correspondendo, assim, à característica (2), a existência de agrupamentos com traços que se sobrepõem parcialmente. A figura abaixo demonstra mais claramente tal configuração:

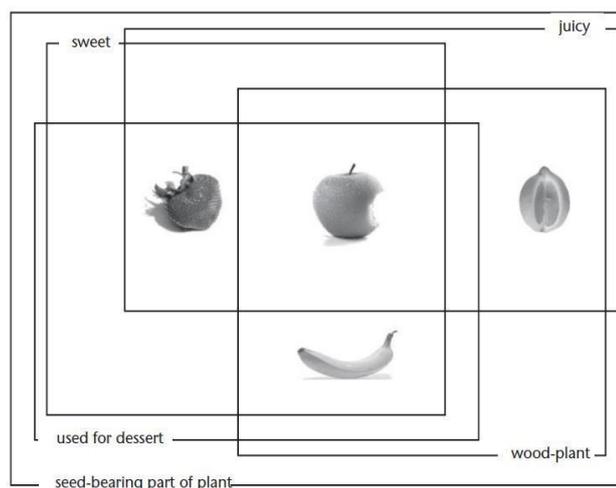


Figura 1: Sobreposição de traços exemplificando a semelhança de família (Fonte: Geeraerts, 2010, p. 191)

Esse modelo de entendimento da estrutura das categorias foi, então, aplicado à compreensão da polissemia. Para sua descrição, um item lexical é, assim, entendido como uma categoria, que pode apresentar os mesmos efeitos postulados pela Teoria Prototípica. Nessa perspectiva, o interesse fundamental é descrever a estrutura semasiológica do item, ou seja, descrever os diversos significados de um item lexical e a forma como eles estão relacionados.

Os efeitos prototípicos apontados para significados individuais podem ser vistos como aplicáveis à estrutura geral do item lexical, ou seja, à forma como os significados individuais se relacionam uns com os outros na estrutura interna do item. Há, assim, um duplo nível dessa análise, pois, da mesma forma que a estrutura geral do item é entendida como uma categoria, os significados considerados de forma individual também são percebidos desse modo, pois cada significado tem seu centro prototípico, aplicações mais representativas, e o item como um todo tem seu centro prototípico, significados mais centrais. Lewandowska-Tomaszczyk (2007, p.148) resume tais postulados da seguinte forma:

Palavras polissêmicas consistem em várias categorias radiais relacionadas, mesmo que cada um dos significados polissêmicos possa ele mesmo apresentar uma estrutura prototípica complexa. O membro central da categoria radial fornece um modelo cognitivo que motiva os significados não centrais. As extensões de significado agrupadas em torno da categoria central são relacionadas por uma variedade de relações possíveis, tais como transformações por esquemas de imagem, metáfora, metonímia [...]¹⁸.

¹⁸ [Polysemic words consist of a number of radially related categories even though each of the polysemic senses can itself display a complex prototype structure. The central radial category member provides a cognitive model that motivates the noncentral senses. The extended senses clustered around the central category are related by a variety of possible links such as image schema transformations, metaphor, metonymy (...)], tradução minha.

Assim, fica clara a importância central do conceito de *semelhança de família*¹⁹ para o entendimento da polissemia nesses termos. Para o modelo semântico-cognitivo de descrição da polissemia, os significados de um item lexical estariam dispostos em agrupamentos de significados inter-relacionados, centrados em torno de um significado (mais básico), no qual efeitos de semelhança de família caracterizariam a relação entre os vários significados, com o significado central geralmente combinando o máximo de características salientes (GEERAERTS, 2010, p.193). Ou seja, não há necessariamente atributos compartilhados por todos os significados da categoria, mas a categoria seria unida a partir de uma estrutura na qual os atributos se entrecruzam.

Dessa concepção, segundo Geeraerts (1995), surge a necessidade de lidar com diversos dados, sendo que o principal para a presente discussão é a dificuldade no estabelecimento dos limites da polissemia. Isso porque, pela dificuldade em delimitar os limites de cada significado, pela natureza multidimensional que a estrutura semasiológica do item lexical adquire, “não há necessariamente uma solução que seja única e ideal para traçar em um item lexical linhas divisórias ao redor de um significado nem entre significados” (GEERAERTS, 2001, p.10).

Como exemplo de polissemia, trazemos novamente a análise do item lexical *fruit* apresentada por Geeraerts (2010, p.193). O autor apresenta as seguintes glosas para os significados do item lexical:

- a. parte doce, suculenta, normalmente utilizada como sobremesa;
- b. parte de uma planta, que contém a semente;
- c. resultado comestível de um processo vegetal;
- d. resultado natural de um processo orgânico;
- e. resultado positivo de um processo ou de uma atividade;
- f. resultado de um processo ou de uma atividade.

Segundo o autor, uma possível representação da forma como esses significados se relacionam é o diagrama apresentado na Figura 2, que mostra graficamente uma possível leitura para essa estrutura de semelhança de família.

Por exemplo, o significado central **a**, relaciona-se com o significado **c** por generalização e com o significado **f** por metáfora, mesmo que, em parte, os significados **c** e **f** também sejam semanticamente relacionados. Há, dessa forma, uma inter-relação entre os significados do item.

¹⁹ Esse conceito foi cunhado por Wittgenstein em sua discussão da definição da palavra do alemão *Spiel*, “jogo”. O autor argumenta que não há um grupo de características comuns aos membros dessa categoria que sejam suficientes para distingui-los de não jogos. Ao contrário, a categoria é estruturada por uma rede de entrecruzamento e sobreposição de similaridades, a qual ele deu o nome de *family resemblance* (TAYLOR, 2003).

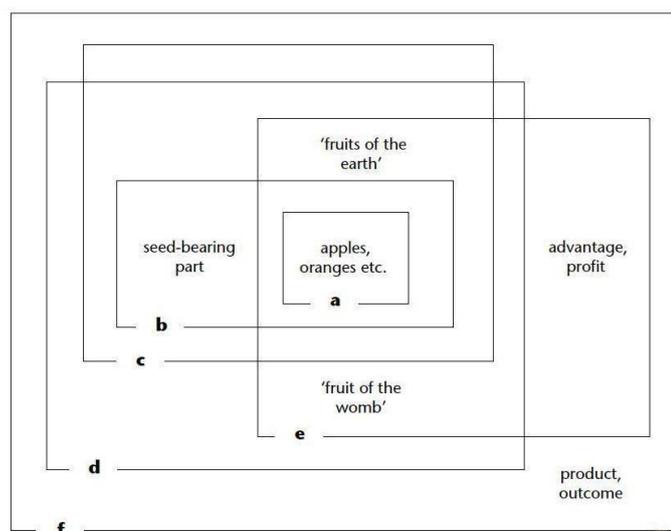


Figura 2. Efeitos prototípicos na categoria *fruit* (Fonte: Geeraerts, 2010, p.194)

Geeraerts (2006d) ressalta ainda a dificuldade em se diferenciar a polissemia de casos de vagueza – nos quais dois significados não distintos estão unidos como subcasos de um significado mais geral. Geeraerts aponta para a falta de um conjunto coerente de critérios para tal distinção, pois uma comparação entre os diversos testes (lógico, linguístico e definicional) demonstrou que eles geram resultados distintos em contextos diferentes (GEERAERTS, 2006d). Nessa linha, Tuggy (2006) afirma que tais considerações levam a uma concepção na qual as próprias noções de polissemia e vagueza não podem ser tidas como absolutas e que esses fenômenos fazem parte de um *continuum* homonímia–polissemia–vagueza.

4. ANÁLISES A PARTIR DA SEMÂNTICA COGNITIVA

Geeraerts (2001; 2006c) destaca que, pela própria linearidade da escrita e consequente linearidade da estrutura de um dicionário, há a impossibilidade de que se represente, em uma estrutura linear de arranjo de acepções, a estrutura cognitiva multidimensional que a polissemia apresenta. Assim, a polissemia dos itens lexicais é afetada pelo que o autor chama de “obrigatoriedade de linearização” dos significados imposta pelos dicionários. Ele resume o problema da seguinte forma:

enquanto que as microestruturas lexicográficas consistem basicamente em uma lista de significados lexicais bem delimitados e numerados em sequência, a estrutura de conceitos prototipicamente organizados é caracterizada pelo agrupamento e pela máxima sobreposição. Como pode, então, a estrutura multidimensional de conceitos

prototípicos ser mapeada na ordem linear do dicionário?²⁰ (GEERAERTS, 2006c, p.330-331)

Nesse viés, o autor considera que tal estrutura não consegue ser propriamente mapeada por formas de ordenamento tradicionais utilizadas pelos dicionários, de modo que a utilização de uma teoria que perceba a estruturação radialmente agrupada da polissemia é o ponto de partida fundamental para a consideração desse fenômeno no âmbito lexicográfico (GEERAERTS, 2001, p.18; 2007, p.1168).

Como apresentamos acima, a concepção semântico-cognitiva de significado pressupõe a impossibilidade de definir muitas categorias prototípicas a partir de traços necessários e suficientes. Isso leva à consideração do significado como um fenômeno altamente flexível, para o qual diferentes tipos de representação podem ser julgados adequados. Esse fato está em consonância com as diferenças encontradas entre as divisões apresentadas pelos dicionários nas acepções dos itens lexicais analisados no que se refere aos processos de *lumping* e *splitting*, principalmente no caso de *fresh*.

Assim, os problemas trazidos para os dicionários com relação à demarcação e à flexibilidade dos significados são entendidos como uma consequência da própria estrutura semasiológica dos itens lexicais. Desse modo, parece pertinente dizer que “de forma alguma está obvio o que são os significados lexicais, ou como eles devem ser estabelecidos, ou, no caso de palavras supostamente polissêmicas, quantos significados distintos devem ser postulados²¹” (TAYLOR; CUYCKENS; DIRVEN, 2003, p.17).

Além disso, retomando a noção de estrutura multidimensional dos itens lexicais, parece-nos que a grande contribuição de uma teoria como a Semântica Cognitiva Lexical para a Lexicografia está em utilizar a ideia de protótipo e sobreposição – a organização em semelhança de família – para a organização microestrutural do verbete. Buscar manter os agrupamentos de significado na organização das acepções de um dicionário parece questão fundamental para a Semântica Cognitiva.

Desta forma, uma organização que buscasse manter e explicitar as relações semânticas entre os significados de um item lexical auxiliaria a solucionar possíveis problemas trazidos pelo *lumping* e *splitting* para a lexicografia. Isso porque, diferentemente da frequência – que separa significados semanticamente relacionados, pois o critério para ordenar as acepções é a

²⁰ [whereas lexicographical microstructures basically consist of a list of neatly separated consecutively numbered lexical meanings, the structure of prototypically organized concepts is characterized by clustering and maximal overlapping. How then can the multidimensional structure of prototypical concepts be mapped onto the linear order of the dictionary?], tradução minha.

²¹ [it is by no means obvious what word meanings are, nor how they are to be stated, nor, in the case of words presumed to be polysemous, just how many different meanings need to be postulated], tradução minha.

quantidade de ocorrências de cada um deles –, uma organização com preceitos semântico-cognitivos tentaria manter unidos os grandes blocos de acepções relacionadas semanticamente. De tal modo, independentemente de se optar pelo *lumping* ou *splitting*, os significados relacionados estariam juntos, mesmo que definidos por uma *catch-all definition* ou discriminados individualmente.

5. CONCLUSÃO

Neste trabalho, avaliamos os fenômenos de *lumping* e *splitting* nos *learner's dictionaries*. A partir das análises, demonstramos que esses fenômenos estão presentes e fazem parte do trabalho lexicográfico. Mesmo que diversas questões práticas restrinjam a forma como o lexicógrafo realizará a divisão dos significados nos verbetes – como as restrições editoriais, o tipo de dicionário em questão e o usuário para qual a obra é feita –, o caráter subjetivo presente em tal tarefa é explicitado pelo tratamento distinto dado pelos dicionários avaliados, todos de mesmo tipo e com foco no mesmo público-usuário.

Com a aproximação da concepção semântico-cognitiva de polissemia, buscamos demonstrar que essa teoria semântico-lexical permite compreender tais fenômenos de forma mais natural e, possivelmente, mais apropriada para seu tratamento no âmbito lexicográfico. Isso porque a compreensão da estrutura semasiológica dos itens lexicais como sendo multidimensional, com uma organização a partir de semelhanças de família, se aplicada à organização lexicográfica, parece aproximar a descrição linguística realizada pelo lexicógrafo de um aporte que permita lidar com as questões de demarcação e flexibilidade do significado linguístico, fundamentais na discussão dos fenômenos de *lumping* e *splitting*.

REFERÊNCIAS

1. CUYCKENS, H.; ZAWADA, B. Introduction. In: _____; _____. *Polysemy in cognitive linguistics: selected papers from the International Cognitive Linguistics Conference*. Amsterdam: John Benjamins, 2001.
2. FILLMORE, C. J.; ATKINS, S. Describing polysemy: the case of 'crawl'. In: RAVIN, Y.; LEACOCK, C. (org.). *Polysemy: theoretical and computational approaches*. Oxford: OUP, 2000. p.91-109
3. FONTENELLE, T. Lexicography. In: SIMPSON, J. *The Routledge Handbook of Applied Linguistics*. London and New York: Routledge, 2011.

4. GEERAERTS, D. Representational formats in cognitive semantics. *Folia linguística*, 29:1-21-2, Mouton de Gruyter, p.21-41, 1995.
5. _____. The definitional practice of dictionaries and the Cognitive Semantic conception of polysemy. *Lexicographica*, p.6-21, v.17, 2001.
6. _____. A rough guide to Cognitive Linguistics *In: GEERAERTS, D. (org.). Cognitive Linguistics: Basic Readings*. Berlin/New York: Mouton de Gruyter, 2006a.
7. _____. Prospects and problems of prototype theory. *In: GEERAERTS, D. (org.). Cognitive Linguistics: Basic Readings*. Berlin/New York: Mouton de Gruyter, 2006b.
8. _____. The lexicographical treatment of prototypical polysemy. *In: GEERAERTS, D. Words and other wonders*. Berlin: Mouton de Gruyter, 2006c.
9. _____. Vagueness' puzzles, polysemy vagaries. *In: GEERAERTS, Dirk. Words and other wonders*. Berlin: Mouton de Gruyter, 2006d.
10. _____. Lexicography. *In: _____; CUYCKENS, H. The Oxford Handbook of Cognitive Linguistics*. New York: OUP, 2007.
11. _____. *Theories of Lexical Semantics*. Oxford/New York: OUP, 2010.
12. HANKS, P. Do word meanings exist? *In: FONTENELLE, T. (org.). Practical lexicography: a reader*. Oxford: OUP, 2008.
13. KILGARRIFF, A. The hard parts of Lexicography. *International Journal of Lexicography*, v.11, n.1, 1998.
14. LEWANDOWSKA-TOMASZCZYK, B. Polysemy, Prototypes, and Radial Categories. *In: GEERAERTS, D.; CUYCKENS, H. (org.). The Oxford Handbook of Cognitive Linguistics*. New York: OUP, 2007. p.139-169
15. OLIVEIRA, A. F. S. *Subsídios da Semântica Cognitiva para a disposição das acepções nos learner's dictionaries*. 2010. 231f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.
16. RAVIN, Y.; LEACOCK, C. Polysemy: an overview. *In: _____; _____. Polysemy: theoretical and computational approaches*. Oxford: OUP, 2000. p.1-29
17. ROSCH, E. Principles of categorization. *In: MARGOLIS, E.; LAURENCE, S. (org.). Concepts: core readings*. Cambridge: MIT Press, 1999.
18. TAYLOR, J. *Linguistic Categorization*. Oxford/New York: Oxford University Press, 2003.
19. TAYLOR, J.; CUYCKENS, H.; DIRVEN, R. Introduction: new directions in cognitive lexical semantic research. *In: _____; _____; _____. Cognitive Approaches to Lexical Semantics*. Berlin/New York: Mouton de Gruyter, 2003. p.1-28

20. TUGGY, D. Ambiguity, polysemy and vagueness. *In: GEERAERTS, D. (org.). Cognitive Linguistics: Basic Readings*. Berlin/New York: Mouton de Gruyter, 2006.
21. TUGWELL, D.; KILGARRIFF, A. Harnessing the Lexicographer in the Quest for Accurate Word Sense Disambiguation. *In: SOJKA, P.; KOPEČEK, I.; PALA, K. (org.). TSD 2000, LNAI 1902*, pp. 9–14, 2000.

OBRAS LEXICOGRÁFICAS:

22. CALD. *Cambridge Advanced Learner's Dictionary*. Cambridge: Cambridge University Press, 2008.
23. LDCE. *Longman Dictionary of Contemporary English*. Essex: Pearson Education Limited, 2009.
24. CEDCa. *Collins English Dictionary Canadian Edition*. Glasgow: Harper Collins Publishers, 2005.
25. COBUILD. *Collins COBUILD Advanced Learner's Dictionary*. Glasgow: Harper Collins Publishers, 2006.
26. OALD. *Oxford Advanced Learner's Dictionary*. Oxford: OUP, 2005.
27. OBED. *Oxford Basic English Dictionary*. Oxford: OUP, 2006.
28. OCDCE. BARBER, K. *et al. Oxford Canadian Dictionary of Current English*. Oxford: OUP, 2005.

ABSTRACT: Polysemy poses a wide range of issues to the lexicographer's work. One of those issues concerns how to establish how many senses a lexical item has. Starting from this question, two lexicographic phenomena known as lumping and splitting arise. The former is related to the choice of keeping distinct meanings of a lexical item under a single sense, while the latter consists of separating its meanings into different senses. This paper evaluates two entries extracted from learner's dictionaries in a way to verify in which forms both these phenomena are presented in those works. Based on the cognitive semantics conception of polysemy, we argue that this lexical semantic theory enables us to understand these phenomena on a more appropriate way for their lexicographic treatment and is useful for sense arrangement.

KEYWORDS: Lexicography; polysemy; lumping; splitting.

Recebido no dia 29 de novembro de 2012.

Aceito para publicação no dia 05 de março de 2013.