

# OS PROBLEMAS DA CLASSIFICAÇÃO TRADICIONAL DAS UNIDADES LÉXICAS E UMA PROPOSTA DE SOLUÇÃO: O CRITÉRIO SÊMIO-TÁXICO

Aldo Luiz BIZZOCCHI<sup>1</sup>

- **RESUMO:** Este trabalho discute os problemas que envolvem a delimitação das unidades léxicas da língua – que costumam ser identificadas ao conceito intuitivo e nebuloso de “palavra” –, tanto no sistema (lexias) quanto no discurso (vocábulos), por força das peculiaridades morfossintáticas e gráficas de cada língua, que conduzem muitas vezes a uma dificuldade no reconhecimento da unidade simples, composta ou complexa (na verdade, seqüência de unidades simples ou compostas), com conseqüentes repercussões nos estudos lexicológicos e lexicográficos, em que é fundamental reconhecer com clareza a fronteira entre tais unidades léxicas. A seguir, propõe-se um critério metodológico de delimitação das unidades léxicas, com vistas à solução do problema em questão.
- **PALAVRAS-CHAVE:** Unidades léxicas; lexia; vocábulo; lexicologia; lexicografia; critério sêmio-táxico.

## Introdução

Um dos grandes problemas que sempre se colocam quando se estuda o léxico de uma língua do ponto de vista morfossintático é a classificação das lexias ou dos vocábulos nas categorias simples, composto, complexo e textual. Embora tal distinção não seja ociosa, porquanto diga respeito aos próprios processos de formação dessas unidades léxicas,

---

<sup>1</sup> Programa de Mestrado em Comunicação e Ciências da Linguagem – Instituto de Ciências Sociais e Comunicação – Universidade Paulista – 04026-002 – São Paulo.

ela acarreta uma série de embaraços, já que a fronteira entre as categorias supramencionadas é quase sempre fluida, além de os critérios tradicionais de classificação se basearem freqüentemente no aspecto gráfico da palavra, o qual é em geral bastante enganador. Assim, procuramos aqui propor um critério rigoroso de classificação morfossintática das unidades léxicas, que escape aos percalços das classificações tradicionais e que seja válido para qualquer língua que admita as categorias léxicas em questão.

## **A classificação morfossintática das unidades léxicas**

Pottier (1974, p.266-7) classifica as lexias em simples, compostas, complexas e textuais, conforme segue:

a) *La lexie simple corresponde au "mot" traditionnel dans de nombreux cas: chaise, pour, mangeait, la.*

b) *La lexie composée est le résultat d'une intégration sémantique, qui se manifeste formellement: tire-bouchons, vert-bouteille, rez-de-chaussée. Le lien peut être très étroit entre un lexème et un grammème, pour former un lexème secondaire: fourchette (qui ne s'oppose plus à fourche), remanier († manier).*

*Toute séquence peut s'intégrer et former un nouveau lexème. Esp.: ensimismarse (sur "en si mismo"), pordiosear (sur "¡ por Dios!").*

c) *La lexie complexe est une séquence en voie de lexicalisation, à des degrés divers:*

*La guerre froide, un complexe industriel, prendre des mesures, feu rouge, bel e bien, hot dogs.*

...

d) *La lexie textuelle est une lexie complexe qui atteint le niveau d'un énoncé ou d'un texte: hymne national, prière, tirade, devinette, proverbe...*

Note-se que a tipologia da lexia proposta por Pottier se atém sobretudo ao grau de liberdade combinatória dos lexemas que a formam (Pais, 1977, p.62), o qual é atestado principalmente pela grafia (as lexias simples consistem uma só palavra, as lexias compostas por vezes têm seus lexemas constituintes ligados por hífen, as lexias complexas e textuais são formadas de palavras livres, separadas por brancos gráficos).<sup>2</sup> Na

<sup>2</sup> Neste trabalho, estamos empregando o termo *palavra* como seqüência de letras não separadas entre si por espaços em branco ou qualquer outro sinal gráfico.

pesquisa lexicológica, a unidade léxica detectada no *corpus* é sempre uma palavra, manifestação efetiva de um vocábulo, unidade de discurso. Como o vocábulo não é senão uma lexia que se atualiza num discurso determinado, sofrendo, portanto, uma redução sêmio-táxica de seu significado, ocorre que também os vocábulos se dividem em simples, compostos e complexos (entretanto, devemos nos referir à atualização discursiva de uma lexia textual como enunciado atualizado). Embora os vocábulos possam ser lexicais ou gramaticais, segundo contenham ou não lexemas em sua estrutura morfossintática, analisaremos aqui apenas os vocábulos lexicais. As definições de vocábulo simples, composto e complexo que daremos a seguir são definições tradicionais e baseadas sobretudo no aspecto gráfico dos vocábulos. Tais definições aplicam-se bastante bem a línguas como o português ou o francês. Entretanto, não funcionam para línguas como o alemão e o holandês, por exemplo. Assim sendo, partiremos delas para em seguida demonstrar a insuficiência, problematizando a questão com alguns contra-exemplos. A partir daí, proporemos um novo critério para a identificação dos tipos de vocábulo. O estabelecimento de uma definição consistente de vocábulo simples, composto e complexo é de fundamental importância em lexicologia e lexicografia, uma vez que a pesquisa lexicológica e lexicográfica quase sempre envolve a coleta de unidades a partir de um *corpus*, sendo então necessário delimitar tais unidades, isto é, definir que tipos de vocábulos constituirão o objeto da análise; para tanto, é imprescindível saber se uma determinada unidade léxica é ou não decomponível em unidades menores.

O vocábulo lexical simples é aquele formado de um único lexema, e um número qualquer de gramemas:

$$V_s = \{1, g_1, g_2, \dots, g_n\}$$

Gráficamente, o vocábulo simples apresenta-se como uma seqüência ininterrupta de letras, percebida e seguida de espaços em branco, seqüência esta que, como já fizemos anteriormente, denominaremos, na falta de termo melhor, como palavra. Exemplos: cadeira, elefante, bonito, amar.

O vocábulo lexical composto resulta da combinação de dois ou mais vocábulos simples. Gráficamente, os vocábulos simples que constituem o vocábulo composto apresentam-se aglutinados ou ligados por hífen, de modo a formarem uma seqüência ininterrupta de sinais gráficos, pre-

cedida e seguida de espaços em branco, ou seja, uma palavra. Exemplos: cata-vento, morfossintaxe, guarda-roupa, pé-de-moleque.

O vocábulo lexical complexo é aquele formado a partir da combinação de dois ou mais vocábulos lexicais simples ou compostos e, eventualmente, também vocábulos gramaticais, que servem de relatores. Graficamente, os vocábulos lexicais e gramaticais constituintes do vocábulo complexo apresentam-se destacados entre si, isto é, separados por espaços em branco. Podemos dizer, então, que o vocábulo complexo é formado de duas ou mais palavras. Exemplos: aula magna, greve geral, copa do mundo, processamento de dados.

O principal problema que essas definições tradicionais acarretam diz respeito exatamente ao aspecto gráfico dos vocábulos, pois não é verdade que o vocábulo simples sempre se constitua graficamente de uma única palavra, ou que o vocábulo composto sempre apresente seus formantes aglutinados ou hifenados, assim como nem sempre as palavras constituintes do vocábulo complexo aparecem separadas por espaços em branco. Senão vejamos:

a) Sejam formas verbais compostas como *está chovendo* ou *havia feito*. Em cada um desses exemplos, temos duas palavras distintas, entretanto, *está chovendo* pode perfeitamente comutar com uma forma simples como *chove*, assim como *havia feito* pode comutar com *fizera*. Não se pode dizer que o verbo auxiliar seja um vocábulo lexical independente, já que não remete a nenhum *designatum*, mas desempenha apenas uma função gramatical, que é a de indicar um tempo e um aspecto verbais, no caso, o presente continuativo e o pretérito mais-que-perfeito. Nesse sentido, o verbo auxiliar possui em línguas predominantemente sintagmáticas como o português, o francês e o inglês, a mesma função de uma desinência modo-temporal numa língua paradigmática como o latim. Cada uma das formas verbais tomadas como exemplo constitui, portanto, um vocábulo simples, contendo um único lexema.

b) Sejam os *phrasal verbs* do inglês, como, por exemplo, *go out*, *get in*, *look up*, *take away* etc. Graficamente, constituem-se de duas ou mais palavras, e, por vezes, podemos inserir outros elementos entre elas (por exemplo, *take it away*). Não obstante, todo *phrasal verb* constitui um único vocábulo, por sinal, um vocábulo simples, já que apresenta um único lexema. Assim, o inglês *go out*, "sair", resulta da combinação de *go*, "ir", e *out*, "para fora", da mesma forma que o latim *exire*, "sair", resulta de *ex*, "para fora" e *ire*, "ir". Temos, assim, no inglês, um caso de derivação semelhante ao que ocorre na prefixação em latim, com a única dife-

rença de que no inglês o afixo é posposto ao verbo e possui certa autonomia sintática. O mesmo vale para os verbos destacáveis do alemão (por exemplo, *ausgehen*, "sair", mas *geht aus*, "sai").

c) Por outro lado, temos contrações gráficas de vocábulos distintos, por exemplo, o inglês *cannot*, contração de *can* e *not*, e a forma não contratada *can not* não é permitida em inglês. Trata-se aí de mero capricho ortográfico, uma vez que *cannot* comuta com *is not*, *has not*, *may not*, *must not*, formas não contratadas. Assim, *cannot* constitui-se de dois vocábulos, um vocábulo lexical simples e um vocábulo gramatical.

d) Sejam os vocábulos alemães *Generalstreik* e *Produktionsprogramm*, por exemplo. A despeito de se apresentarem graficamente como uma única palavra, cada um desses exemplos constitui um vocábulo complexo, pois *Generalstreik*, "greve geral", se opõe a *siegreicher Streik*, "greve vitoriosa", assim como *Produktionsprogramm*, "programa de produção" se opõe a *produktives Programm*, "programa produtivo". Ocorre que, em alemão, sempre que o adjunto adnominal ou o complemento nominal de um substantivo é outro substantivo, deve este vir graficamente justaposto ao substantivo que lhe serve de base, o que acontece também com certos adjetivos (caso de *General-*).

e) Compare-se agora o vocábulo mesa redonda nas duas frases a seguir:

(1) *Minha sala de jantar possui uma mesa redonda e quatro cadeiras.*

(2) *Os líderes dos países industrializados realizaram uma mesa-redonda em Paris.*

Na frase (1), mesa redonda é perfeitamente decomponível em mesa e redonda, pois nesse caso mesa redonda se opõe a mesa quadrada, mesa marrom, e a almofada redonda, bandeja redonda etc. Pode-se inclusive intercalar outros elementos entre mesa e redonda: "Esta mesa é redonda". Já na frase (2), mesa-redonda constitui um todo indecomponível. Por isso mesmo, é grafada com hífen. Assim, em (1) temos um vocábulo complexo e em (2), um vocábulo composto. Podemos dizer que em (1) trata-se de fato de uma mesa que possui por atributo ser redonda. Em (2), o significado recoberto pelo significante mesa-redonda na verdade nada tem a ver com mesa ou com a forma redonda. Temos aí uma simples reminiscência da motivação metonímica inicial do vocábulo, hoje já totalmente desaparecida.

f) Em muitos casos, vocábulos resultantes de processos morfossintáticos semelhantes são grafados de forma diferente, atestando, em alguns casos, diferentes estágios de lexicalização, em outros, simples incoerência

ortográfica. Comparem-se os vocábulos ingleses *milkman*, *songwriter*, *fig-tree* e *data processing*. No primeiro caso, temos um caso de composição em que o substantivo de base, *man*, praticamente perdeu o estatuto de vocábulo, tendo-se transformado em mero sufixo (compare-se com o português *leiteiro*). O segundo caso é semelhante ao primeiro, porém aí a motivação de *writer* é muito mais forte, de modo que grafias *song-writer* e mesmo *song writer* seriam igualmente aceitáveis, embora desusadas, ao contrário de *\*milk-man* e, menos ainda, de *\*milk man*. Em *fig-tree* (também grafado como *fig tree*), temos algo semelhante a *songwriter*, porém com uma grafia diferente. Já em *data processing*, qualquer contratação seria impossível. *Data processing* possui o mesmo estatuto do português *processamento de dados*, que pode sofrer intercalação (processamento automático de dados) ou comutação de seus constituintes (processamento de informações, análise de dados). Comparem-se, ainda, os vocábulos portugueses *desordem* e *contra-ordem*, ou os vocábulos ingleses *subordinate* e *co-ordinate*. O mesmo processo de prefixação resulta ora em aglutinação, ora em ligação por hífen, conforme o prefixo utilizado.

Os exemplos citados revelam que a grafia é freqüentemente um critério enganoso para se tipificar vocábulos, ainda mais quando é preciso delimitar com precisão a extensão das unidades que se deseja analisar. Assim, substituiremos tal critério por outro, que, a nosso ver, permite dar conta de todos os exemplos apresentados. Chamá-lo-emos de critério sêmio-táxico.

## **O critério sêmio-táxico de classificação de vocábulos**

Quando duas lexias são colocadas em combinatória sintagmática num enunciado, a semia resultante dessa combinatória é um subconjunto do produto cartesiano dos sememas das lexias envolvidas. No caso de um sintagma nominal formado, por exemplo, de um substantivo qualificado por um adjetivo, temos uma relação base x adjunto, de tal sorte que o semema do substantivo pelo semema do adjetivo, entendidos ambos como conjuntos sêmicos. Assim, essa relação se reduz a um produto cartesiano entre dois conjuntos. Antes de passarmos adiante, convém tornarmos mais claro como se dá esse produto cartesiano. Para tanto, nos basearemos no modelo proposto por Pais (1977, p.74-82) acerca da combinatória semêmica no enunciado simples, modelo este que também se

aplica no caso da combinatória semêmica intra-sintagmática, adaptando, entretanto, tal modelo às nossas necessidades específicas.

Cumpra lembrar que numa situação normal de comunicação, todo enunciado simples contém uma carga de informação pressuposta comum aos dois sujeitos da enunciação, o emissor e o receptor, bem como uma carga de informação supostamente pertencente apenas ao emissor, o que justifica o ato da comunicação. Essas cargas de informação, segundo Pottier (1974), são chamadas, respectivamente, suporte e aporte. Diz Pais (1977) em seu trabalho:

O suporte contém uma carga de informação própria que é o resultado da combinatória dos sememas das lexias que o integram. Essa combinatória intra-suporte é uma atribuição que se realiza numa relação base x adjunto entre o termos do grupo substantivo e do grupo adjetivo. Resulta desse produto o semema do suporte, que se caracteriza por conter apenas semas descritivos, numa visão nominal, estática; contém necessariamente semas da semântica lexical, ligados à descrição dos objetos do universo antropocultural – manifestos ou representados por substitutos – e semas gramaticais, pertencentes à estrutura interna da língua, as taxes, como gênero, número, grau e outros, necessários à definição da combinatória.

O aporte leva o suporte, atribui-lhe uma carga semântica suplementar de informação. Esta carga semêmica resulta por sua vez de uma combinatória intra-aporte, uma atribuição que se realiza, também, numa relação base x adjunto entre os termos do grupo verbal, de um lado, e do grupo adjetivo e do grupo dos complementos do outro lado. Da mesma forma, resulta desse produto o semema do aporte, que vai entrar em combinatória com o do suporte, e que contém, como no caso do suporte, semas lexicais e gramaticais.

O enunciado é uma seqüência de lexias de que o suporte e o aporte são subsequências distintas. Quanto à carga semântica de informação, estrutura-se a semia do enunciado como o subconjunto antropocultural – biofatos, sociofatos, psicofatos e manufatos – ou de semas descritivos de processos, que intervêm entre aqueles objetos, ou seja, semas estáticos e dinâmicos, respectivamente.

No primeiro caso, teremos em estrutura profunda, um esquema lógico-conceptual atributivo, através do qual o semema do suporte recebe semas que o tornam mais específico. No segundo caso, teremos, em estrutura profunda, um esquema lógico-conceptual ativo, através do qual o semema do suporte recebe semas que não o tornam necessariamente mais específico mas o colocam numa relação conjuntural com o aporte, segundo o qual o suporte desencadeia um processo que pode eventualmente atingir outros actantes.

Do que foi dito, resulta que, em qualquer enunciado, ocorrem, tanto no nível intra-suporte, quanto no nível intra-aporte, como ainda no nível

suporte-aporte, relações de combinatória sêmica que produzem a semia final do enunciado, ou esquematicamente:

- NÍVEL DO SUPORTE: base x adjunto = semia do suporte
- NÍVEL DO APORTE: base x adjunto = semia do aporte
- NÍVEL DO ENUNCIADO: suporte x aporte = semia do enunciado

Restringindo-nos aos dois primeiros níveis, encontraremos sintagmas nominais ou verbais, ambos caracterizados por possuir um elemento de base e eventualmente um ou mais adjuntos. Nosso problema é saber se esse conjunto base-adjunto constitui um vocábulo composto ou complexo. Segundo o critério sêmio-táxico, o que vai indicar de que tipo de vocábulo se trata é o seu semema. Vejamos como isso acontece.

Seja o sintagma nominal *mesa redonda*, tal qual depreendido da frase:

*Minha sala de jantar possui uma mesa redonda e quatro cadeiras.*

Temos o semema de mesa:

$$\langle \text{mesa} \rangle = \{+M, +D, -A, +P, +T, +C, +E, \pm R, \pm Q, \pm O, \dots\}$$

e o semema de redonda:

$$\langle \text{redonda} \rangle = \{+M, +D, \pm A, \dots +R, \dots\}$$

onde: M = material, D = descontínuo, A = animado, P = com pés, T = com tampo, C = para comer, E = para escrever, R = redonda, Q = quadrada, O = oval.

A semia do sintagma será um subconjunto do produto cartesiano do semema de mesa pelo semema de redonda, ou seja:

$$\langle \text{mesa redonda} \rangle \subset \langle \text{mesa} \rangle \times \langle \text{redonda} \rangle$$

O conjunto resultante cartesiano  $\langle \text{mesa} \rangle$  e o segundo é um sema do semema  $\langle \text{redonda} \rangle$ , de modo a combinar, um a um, cada sema do primeiro semema a cada um dos semas do segundo semema:

$$\langle \text{mesa} \rangle \times \langle \text{redonda} \rangle = \{(+M, +M), (+M, +D), \dots (+M, \pm R), \dots (+D, +M), (+D, +D), (+D, \pm A), \dots (-A, \pm A), \dots (\pm R, +R), \dots\}$$

Alguns dos pares ordenados desse produto são incompatíveis, porque combinam semas de níveis paradigmáticos diferentes, tais como  $(+M, +D)$  e  $(+D, \pm A)$ , por exemplo. Por essa razão, devem ser eliminados, restando apenas os pares de mesmo nível paradigmático. Para estes, vale a seguinte regra de sinais:

$+ x + = +$   
 $+ x - = -$   
 $- x + = -$   
 $- x - = -$   
 $\pm x + = +$   
 $\pm x - = -$   
 $+ x \pm = +$   
 $- x \pm = -$

Assim, para que um sema pertença à semia do produto, é preciso que ele esteja presente em ambos os sememas fatores; se este pertencer a apenas um dos sememas, ou a nenhum, conseqüentemente, não pertencerá ao produto. O sinal  $\pm$  indica a neutralidade do sema que afeta. Assim, no semema <mesa>, os semas  $\pm R$ ,  $\pm O$ ,  $\pm Q$  podem ou não ocorrer, porém a ocorrência de um deles exclui a dos demais. Da mesma forma, no semema <redonda>, o traço distintivo "material" é positivamente marcado (somente objetos do mundo material, isto é, biofatos ou manufatos, podem ter forma redonda), porém o traço "animado" é neutro, pois redondo pode ser atributo tanto de seres animados quanto de seres inanimados. Partindo da perspectiva semântica estrutural de Greimas (1966), podemos dizer que os semas  $+M$ ,  $+D$  etc. pertencem ao núcleo sêmico do semema <mesa>, ao passo que os semas  $\pm R$ ,  $\pm Q$  etc. constituem semas contextuais, mutuamente exclusivos, cuja inflexão em termos de articulação sêmica pode ser subsumida por um sema único como "forma", por exemplo. No semema <redonda>, o sema  $+R$  pertence ao núcleo sêmico, de modo que a combinação semântica <mesa> x <redonda> vai provocar a atualização do sema complexo "forma", do semema <mesa>, como "forma redonda", à exclusão dos demais semas, "forma quadrada", "forma oval" etc. Estabelece-se, assim, entre <mesa> e <redonda> uma isotopia.

Aplicada a regra dos sinais a cada um dos pares do produto cartesiano, resulta que o semema de mesa redonda tem a seguinte configuração:

$$\langle \text{mesa redonda} \rangle = \{+M, +D, -A, \dots +R, \dots\}$$

Observe-se que mesa redonda é um objeto material, descontínuo, não animado, e, além disso, de forma redonda.

Tomemos, agora, o sintagma nominal mesa-redonda depreendido da frase:

Os líderes dos países industrializados realizaram uma mesa-redonda em Paris.

Nesse caso, o semema de mesa-redonda não resulta de nenhum produto cartesiano entre os sememas de mesa e de redonda. Temos aí semas descritivos de um sociofato, e não de um manufato.

Poderíamos aqui ensaiar uma primeira definição de vocábulo simples, composto e complexo com base no grau de integração semântica e de liberdade combinatória dos lexemas constituintes desse vocábulo. Poderíamos dizer então que:

a) o vocábulo lexical simples é aquele que possui um único lexema e um número qualquer de gramemas;

b) o vocábulo lexical composto resulta da integração semântica de dois ou mais vocábulos simples, de tal modo que o semema desse vocábulo não seja produto de uma combinação dos sememas dos vocábulos simples que o constituem;

c) o vocábulo lexical complexo resulta da combinação sintática de dois ou mais vocábulos simples ou compostos, de sorte que o semema desse vocábulo é o resultado da combinação dos sememas dos vocábulos simples ou complexos que o constituem.

Desse ponto de vista, podemos dizer que mesa redonda constitui um vocábulo complexo, ao passo que mesa-redonda constitui um vocábulo composto.

Nos termos de Pottier (1974) a *lexia complexa* é uma seqüência em vias de lexicalização, portanto, é natural que todo vocábulo complexo constantemente repetido tenda a tornar-se uma unidade cristalizada, e dessa maneira indivisível, e como tal seja dicionarizada, atingindo, assim, o estatuto de vocábulo composto. Tal processo é particularmente facilitado em línguas que, a exemplo do inglês e, principalmente, do alemão, apresentam grande flexibilidade combinatória de seus lexemas, por mecanismos de combinação sintagmática como a justaposição, por exemplo. Desse modo, tanto os vocábulos complexos quanto alguns vocábulos compostos apresentam semema resultante da combinação sêmica dos sememas dos vocábulos integrantes. Vê-se, assim, que a definição de vocábulo composto que demos anteriormente ainda é insuficiente, uma vez que o critério sêmio-táxico de identificação de vocábulos tem valor basicamente negativo: se o semema do vocábulo em questão não resulta da combinação semântica dos sememas dos lexemas que o integram, tal vocábulo é indubitavelmente composto. Caso contrário, é preciso recorrer a um critério adicional que permita distinguir os dois tipos de

vocábulo é o resultado do produto cartesiano do semema do lexema e dos sememas dos gramemas. Entretanto, o semema lexical não deve resultar de nenhum produto cartesiano anterior, vale dizer, não deve ser decomponível em sememas menores. Graficamente, um vocábulo simples como cadeira teria a seguinte representação:

PC	{S <sub>1</sub> , S <sub>2</sub> , S <sub>3</sub> ,... S <sub>n</sub> }
PE	cadeira

onde PC = plano do conteúdo e PE = plano de expressão.

Vocábulo composto é aquele cuja estrutura morfossintática contém mais de um lexema, e um número qualquer de gramemas, e que possua pelo menos uma das seguintes características:

- a) sua semia não resulta do produto cartesiano dos sememas de seus lexemas;
- b) ele não é decomponível em unidades menores sintaticamente autônomas.

Convém lembrar que, diacronicamente, um vocábulo como *fidalgo* é primeiramente complexo (*fidalgo* = filho d'algo), e a seguir composto. Numa perspectiva sincrônica, tal vocábulo deve ser considerado como simples, uma vez que, atualmente, *fidalgo* possui apenas um lexema, não sendo mais sentida pelos falantes a composição *filho d'algo*.

Graficamente, um vocábulo composto como *guarda-chuva* seria representado como:

{S <sub>1</sub> , S <sub>2</sub> , S <sub>3</sub> ,... S <sub>n</sub> }
guarda-chuva

Já um vocábulo composto como *saca-rolhas* poderia ser representado como:

{S <sub>1</sub> , S <sub>2</sub> , S <sub>3</sub> ,... S <sub>m</sub> }	x	{S' <sub>1</sub> , S' <sub>2</sub> , S' <sub>3</sub> ,... S' <sub>n</sub> }	=
sacar		rolha	
{S <sub>1</sub> , S <sub>2</sub> , S <sub>3</sub> ,... S <sub>n</sub> } x {S' <sub>1</sub> , S' <sub>2</sub> , S' <sub>3</sub> ,... S' <sub>n</sub> }			
saca-rolhas			

No primeiro caso (*guarda-chuva*), temos a indivisibilidade sintática e semântica do vocábulo, no segundo caso (*saca-rolhas*), temos apenas a indivisibilidade sintática.

Vocábulo complexo é aquele que contenha em sua estrutura morfossintática mais de um lexema, e um número qualquer de gramemas, e que possua as seguintes características:

- a) sua semia resulta do produto cartesiano dos sememas de seus lexemas;
- b) ele é decomponível em unidades menores sintaticamente autônomas.

Graficamente, um vocábulo complexo como *greve geral* seria representado como:

$$\begin{array}{|c|} \hline \{S_1, S_2, S_3, \dots S_m\} \\ \hline greve \\ \hline \end{array} \times \begin{array}{|c|} \hline \{S'_1, S'_2, S'_3, \dots S'_n\} \\ \hline geral \\ \hline \end{array} = \begin{array}{|c|} \hline \{S_1, S_2, S_3, \dots S_n\} \times \{S'_1, S'_2, S'_3, \dots S'_n\} \\ \hline greve geral \\ \hline \end{array}$$

Assim, poderíamos esquematizar nossas definições da seguinte maneira:

Tipo de vocábulo lexical	Número de lexemas	Número de sememas lexicais da semia resultante	Ocorrências em dicionários
simples	1	1	sim
composto	mais de 1	1 ou mais	sim
complexo	mais de 1	mais de 1	não

Retomando agora nossos exemplos anteriores, com base no critério sêmico-táxico de classificação, teremos que:

- a) as formas verbais *está chovendo* e *havia feito* constituem vocábulos simples, possuindo cada uma um único lexema e um único semema lexical na semia resultante,<sup>3</sup>

3 Comparem-se as formas verbais *vai fazer* e *quer fazer*. No primeiro caso, o verbo auxiliar *vai* exerce a função de uma simples desinência. No segundo caso, o verbo modal *quer* remete não a uma função gramatical, mas sim a um conceito lingüístico, o conceito de desejo. Note-se que, do ponto de vista formal, *vai fazer* comuta com uma forma sintética como *fará*, ao passo que *quer fazer* só pode comutar com uma forma igualmente analítica como *deseja fazer*.

b) os *phrasal verbs* do inglês, bem como os verbos destacáveis do alemão, são igualmente vocábulos simples;

c) os vocábulos alemães *Generalstreik* e *Produktionsprogramm* são ambos complexos, pois possuem dois lexemas e dois sememas lexicais na semia resultante, e, além disso, seus vocábulos simples constituintes possuem autonomia sintática;

d) *mesa-redonda* é um vocábulo composto e *mesa redonda* é um vocábulo complexo;

e) os vocábulos ingleses *milkman*, *songwriter* e *fig-tree* são compostos, pois, nesse caso, o processo de integração dos lexemas e, conseqüentemente, dos sememas, conduziu tanto a uma gramaticalização progressiva do segundo lexema quanto a uma cristalização sintática. Isso se reflete na lexicalização desses vocábulos, já perfeitamente dicionarizados. Nesse sentido, tais vocábulos tornaram-se unidades léxicas indecomponíveis em unidades menores.

## Conclusão

Como vimos, o princípio tradicional de classificação de lexemas ou vocábulos nas categorias simples, composto, complexo e textual é insuficiente e impreciso, por basear-se em critérios pouco rigorosos e pouco confiáveis, como por exemplo a grafia, a qual tem estado sempre sujeita a vicissitudes históricas e idiosincrasias de toda sorte, o que torna muitas vezes a escolha da representação gráfica das palavras arbitrária e subjetiva. Além disso, tal princípio de classificação não é universal nem mesmo no âmbito das línguas européias. Assim, parece-nos que a análise sêmio-táxica das unidades léxicas fornece bases verdadeiramente científicas para a classificação morfossintática dessas unidades. Sua combinação com o critério da dicionarização conduz, portanto, a uma classificação rigorosa e livre das incoerências dos métodos tradicionais.

BIZZOCCHI, A. L. The problems of the traditional classification of lexical units and a solution proposal: the semio-taxical criterion. *Alfa (São Paulo)*, v.43, p.89-103, 1999.

- **ABSTRACT:** *This paper discusses the problems involving the delimitation of the lexical units of the language – which are usually identified to the intuitive and nebulous concept of “word” –, both at the level of the system (lexias) and the discourse (vocables), because of the morphosyntactic and spelling peculiarities of each language, which many times lead to a difficulty in recognising the simple, compound or complex unit (actually, a sequence of simple or compound units), with consequent repercussions in the lexicological and lexicographical studies, in which it is fundamental to recognise neatly the boundary between such lexical units. Furthermore, a methodological criterion for the delimitation of lexical units, aiming at solving the problem at issue, is proposed.*
- **KEYWORDS:** *Lexical units; lexia; vocable; lexicology; lexicography; semio-taxical criterion.*

## **Referências bibliográficas**

- GREIMAS, A. J. *Sémantique structurale: recherche de méthode*. Paris: Larousse, 1966.
- PAIS, C. T. *Ensaio semiótico-lingüístico*. Petrópolis: Vozes, 1977.
- POTTIER, B. *Linguistique générale: théorie et description*. Paris: Klincksieck, 1974.