

GRAMÁTICA DO PORTUGUÊS

VOLUME I

Comissão Organizadora

Eduardo Buzaglo Paiva Raposo

Maria Fernanda Bacelar do Nascimento

Maria Antónia Coelho da Mota

Luísa Segura

Amália Mendes

Colaboração de

Graça Vicente e Rita Veloso



FUNDAÇÃO
CALOUSTE
GULBENKIAN

2

LÉXICO

8 Organização do léxico

Rui P. Chaves

9 Processos de lexicalização

Maria Fernanda Bacelar do Nascimento

10 Processos de gramaticalização

Amália Mendes



8 ORGANIZAÇÃO DO LÉXICO	185
8.1 Aspetos do significado lexical	187
8.2 Vagueza	189
8.3 Homonímia e polissemia	190
8.3.1 Homonímia	192
8.3.2 Polissemia	192
8.4 Sinonímia	196
8.5 Antonímia	198
8.6 Hiponímia-hiperonímia	200
8.7 Meronímia-holonímia	205
8.8 Relações proposicionais	208
8.8.1 Relações e predicadores simétricos	209
8.8.2 Relações e predicadores reflexos	210
8.8.3 Relações e predicadores transitivos	211



Nem sempre isto acontece com duas palavras diferentes. A intensão diferente de duas palavras ou expressões distintas não implica necessariamente que a sua extensão também o seja. Assim, p.e., as expressões nominais *animal com rins* e *animal com coração* têm sentidos diferentes mas denotam extensionalmente a mesma classe de entidades, já que todos os animais que possuem coração também possuem rins e vice-versa; neste caso, os dois conjuntos são idênticos. Uma possibilidade mais frequente é que a extensão de duas expressões intensionalmente diferentes seja distinta mas com membros comuns; um exemplo é o par de palavras *estudante* e *jovem*: muitos estudantes são jovens (e vice-versa), mas nem todos os estudantes são jovens e nem todos os jovens são estudantes.

Distinguindo-se da denotação (quer intensional quer extensional), a **referência** é a propriedade que têm algumas expressões linguísticas, chamadas **expressões referenciais**, de designarem ou representarem uma entidade particular ou um grupo particular de entidades do **universo do discurso**. As expressões referenciais têm tipicamente como núcleo um nome comum (também chamado substantivo), um nome próprio ou um pronome, i.e., são sintagmas nominais (cf. Cap. 20). Os seguintes sintagmas nominais (sublinhados) constituem expressões referenciais, pois são usados para designar determinadas entidades num determinado contexto enunciativo: a minha vizinha foi às compras, a Maria cozinhou um ótimo jantar e ela zangou-se com o marido.

À entidade, individual ou coletiva, designada por uma expressão referencial chama-se o **referente** da expressão. O referente corresponde, assim, essencialmente, àquilo de que se está a falar quando usamos um determinado sintagma nominal num enunciado concreto (cf. Cap. 20). Por isso, a referência depende não só do sentido das palavras que formam o sintagma nominal (em particular, do substantivo e outros elementos que o modifiquem, como os adjetivos), mas também da situação em que o enunciado é produzido. Por exemplo, na frase *o gato adormeceu*, o sintagma nominal *o gato* refere normalmente um animal diferente de cada vez que a frase é produzida num enunciado concreto, consoante o contexto situacional, o lugar e o tempo em que esse enunciado é produzido. Isto acontece ainda que as palavras *o* e *gato* tenham a mesma contribuição semântica em todos esses enunciados (sobre a interpretação do artigo definido, cf. Caps. 20 e 22).

A dependência da referência relativamente ao contexto situacional ou discursivo em que um enunciado é produzido também é visível em frases como *todos os cães perseguiram o gato*. Nesta frase, o sujeito não é normalmente interpretado como designando todos os cães do mundo (perseguido um determinado gato), mas sim como referindo a totalidade dos cães mencionados anteriormente ou que se encontram na situação concreta em que o enunciado é produzido (e que podem ser apenas três ou quatro). Em resumo, a referência depende do contexto situacional e discursivo em que a frase é dita, da intenção comunicativa de quem enuncia a frase e do conhecimento da situação partilhado pelos interlocutores; em contrapartida, a denotação, incluindo o sentido, não depende do contexto.

A distinção entre denotação e referência é particularmente evidente em casos em que as palavras, num determinado contexto situacional ou discursivo, não possuem um referente. Por exemplo, no enunciado *nenhum gato foi ferido na rodagem do filme, até porque não havia gatos no estúdio*, os sintagmas em que se insere a palavra *gato* (i.e., *nenhum gato* e *gatos*), embora tenham uma denotação,

não referem nenhum animal em particular, já que se declara que não havia gatos na situação em questão. Quando um SN não refere qualquer entidade, diz-se que é **não referencial** (cf. Caps. 20 e 22).

Estabelecidas estas diferenças, é importante assinalar que os aspetos do significado léxico abordados nos pontos seguintes deste capítulo têm a ver não com a referência, mas sim com a denotação das palavras, em particular a denotação intensional ou sentido.

8.2 Vagueza

O sentido da maioria das palavras inclui propriedades conceptuais de natureza muito diferente. Tomemos como exemplo um dos sentidos comuns da palavra *janela*. Esta palavra denota uma abertura na parte exterior de uma edificação ou de um veículo que permite a passagem do ar e da luz para o interior dessa edificação ou veículo, bem como o seu revestimento com vidro ou outro material transparente.

Estas propriedades, ainda que necessárias (ou, pelo menos, típicas) na caracterização daquilo que constitui uma janela, são, no entanto, muito gerais, permitindo alguma variação na sua caracterização precisa. Em particular, deixa-se em aberto toda uma série de possibilidades quanto à forma ou ao tamanho da abertura da superfície, a natureza da edificação ou do veículo, o material de que é feito o revestimento, a sua espessura, a possibilidade de se abrir ou não de modo a deixar passar o ar (por exemplo, as janelas de um avião e de muitos edifícios modernos não se podem abrir), etc. Ou seja, *janela*, e na realidade a maioria das palavras, tem um **sentido vago** (a propriedade semântica correspondente chama-se **vagueza**).

Os sentidos vagos permitem, assim, que exista variação e flexibilidade na especificação interna das propriedades que compõem o sentido das palavras, podendo essa especificação ser diferente de falante para falante; e quanto mais vago for o sentido de uma palavra, mais abrangente será a sua extensão. Uma janela, por exemplo, pode ser redonda, quadrada, retangular, pode ser pequena ou grande, o seu revestimento pode ser feito de um material mais ou menos translúcido, de maior ou menor espessura, pode aplicar-se sobre edificações ou veículos de natureza variada, desde que tenham um volume interno e uma superfície externa adequada, incluindo submarinos, aviões, carros, casas, cabanas, tendas, etc.

Para além das propriedades necessárias, o sentido das palavras inclui outras propriedades que estão frequentemente (e tipicamente) presentes nos objetos por elas designados, mas que não constituem características necessárias da definição. Por exemplo, as janelas típicas são feitas de vidro (mas não necessariamente), podem abrir-se através de um fecho (mas não necessariamente) e têm um caixilho (mas não necessariamente).

Para dar outro exemplo, consideremos a palavra *carro*, de cujo sentido são atributos típicos ter uma carroçaria, um volante, um motor, quatro rodas revestidas de pneus, um chassis, bancos, janelas com um revestimento de vidro, etc. No caso extremo, se removermos todas as peças de um carro à exceção do volante, então certamente já não temos um carro, mas unicamente um volante. No entanto, a determinação dos elementos necessários para que se possa dizer que (ainda) temos um carro varia provavelmente de falante para falante. Ou seja, se removermos as

peças de um carro uma a uma, os falantes poderão divergir sobre se o resultado de cada remoção é ainda um carro ou não. Isto significa que a definição precisa do sentido da palavra *carro* não é clara nem para os falantes nem para os linguistas que estudam o funcionamento da linguagem. Nesse sentido, a maioria das palavras de uma língua tem uma definição mental ligeiramente diferente para cada falante, relacionada com o seu conhecimento do mundo e a sua experiência pessoal. No entanto, a comunicação decorre, em geral, sem problemas, pois muito raramente os falantes se encontram numa situação em que a maioria dos atributos que associam à definição de uma palavra não seja satisfeita pelo seu referente.

Um exemplo extremo de vagueza é a palavra *algo* (cf. Cap. 23). Esta palavra é usada por muitos falantes para designar entidades abstratas ou concretas, e, embora seja tipicamente intuída como o pronome indefinido aplicável a entidades não humanas (e não animadas), é muitas vezes admissível o seu uso mesmo quando o referente inclui ou é constituído por seres animados, incluindo mesmo seres humanos: cf. a aceitabilidade de *algo matou os meus ratinhos da índia, possivelmente o frio, o stress, a gripe, a idade, o gato, o vizinho ou o bebedouro elétrico*, ou de *algo fez o alarme soar, e só pode ter sido um ladrão*. Nesse sentido, *algo* contrasta com *alguém* (outra palavra vaga), corretamente intuído como sendo um pronome indefinido unicamente aplicável a entidades humanas, mas não a animais ou a entidades não animadas, como se pode ver pela estranheza de *#alguém fez o alarme soar, e só pode ter sido ou o gato ou o vento*.

8.3 Homonímia e polissemia

Na sua maioria, as palavras de qualquer língua são ambíguas, isto é, a uma mesma forma lexical correspondem sentidos diversos, e tantos mais quanto maior for a frequência dessas palavras no uso; pelo contrário, são em muito menor número as **palavras monossémicas**, ou seja, aquelas em que há uma relação biunívoca entre a forma e o sentido. São exemplos deste último caso certos nomes científicos, como *camussela*, nome de um pássaro, *urzela*, nome de um líquen ou *batíscafo*, nome de um pequeno submarino para observação científica dos oceanos. São também palavras monossémicas algumas formas derivadas que se relacionam apenas com um dos sentidos da palavra base, como *colunável*, que significa, apenas, 'digno de aparecer nas colunas da imprensa', e que deriva de um dos vários sentidos da palavra *coluna*, o de 'secção regular da imprensa escrita'.

No âmbito da ambiguidade lexical, distinguem-se dois casos: o da **homonímia** e o da **polissemia**. Em ambos, a uma mesma forma fonética e gráfica correspondem sentidos diferentes, pondo-se como problema decidir em que casos se considera estarmos perante palavras distintas ou perante uma só palavra. No primeiro caso, trata-se de homonímia; no segundo, de polissemia.

Para dar um exemplo clássico, será que *verde*, como nome de cor, e *verde*, como designação do estado de maturação de um fruto, são duas palavras homónimas, ou uma só palavra polissémica? Alguns falantes não detetam qualquer relação entre a cor e o estado de maturação, considerando-as, por isso, palavras diferentes. Outros, no entanto, relacionam os dois sentidos, dado que a fruta que tem cor verde é, frequentemente, fruta que ainda não amadureceu, pelo que consideram tratar-se

de uma única palavra. Compreende-se, assim, que nem sempre seja fácil distinguir entre homonímia e polissemia⁴.

Tendo em conta unicamente o conhecimento lexical do falante médio numa determinada época, e seguindo, portanto, uma perspetiva puramente sincrónica, diz-se que duas ou mais palavras são homónimas quando têm a mesma forma fonética e gráfica e sentidos completamente distintos, isto é, em que não se distinguem, presentemente, traços semânticos comuns⁵, como, p.e., *serra* 'instrumento para cortar' e *serra* 'elevação de terreno'.

Em contrapartida, considera-se que uma palavra é polissémica quando a uma mesma forma lexical correspondem sentidos diferentes para os quais é possível estabelecer uma relação através de um ou mais traços semânticos comuns, como, p.e., *vão* no sentido de 'vazio, oco', e *vão* no sentido de 'sem valor, desnecessário'. No caso das palavras polissémicas, podem conceber-se pontes semânticas entre os vários sentidos, atribuídas a processos diversos, sendo particularmente sistemáticos na língua os processos de extensão ou restrição dos sentidos, nomeadamente por metáfora e metonímia (vulgarmente chamados "sentidos figurados"). Neste capítulo, segue-se esta perspetiva sincrónica e não uma perspetiva histórica.

Algumas formas lexicais como, por exemplo, a palavra *banco* podem ser tanto homónimas como polissémicas. Esta forma constitui uma palavra polissémica quando se compara a interpretação de 'instituição' (cf. *o banco vai falir*) com a de 'edifício' onde funciona a instituição (cf. *o banco vai ruir*). Em contrapartida, a forma instancia um caso de homonímia (ou seja, representa duas palavras distintas) quando se compara a interpretação de 'instituição' com a interpretação de 'assento' (cf. *ele sentou-se no banco do jardim*)⁶.

Um dos critérios usados em lexicografia para distinguir entre homonímia e polissemia é observar a história da palavra. Ou seja, seguindo uma perspetiva diacrónica, consideram-se palavras homónimas aquelas que, tendo a mesma forma, têm origens etimológicas diferentes, convergindo numa mesma forma devido à sua evolução fonética. Assim, por exemplo, nos dicionários que adotam este critério, formas lexicais idênticas têm entradas diferentes, equivalendo a palavras distintas, ou seja, homónimas, sempre que as formas associadas aos diversos sentidos têm étimos também diferentes, como, p.e., *manga* 'parte de uma peça de vestuário', cujo étimo é *manica*, do latim, e *manga* 'fruto da mangueira', cujo étimo é *manga*, do malaiala, ou *canto* 'ângulo, esquina, aresta', do latim *canthus*, e *canto* 'ato de cantar, som musical', do latim *cantus*. Em contrapartida, a uma forma lexical cujos sentidos não correspondem a étimos diferentes é dada uma única entrada; neste caso, temos uma única palavra polissémica.

⁴ Por este motivo, a distinção é considerada como praticamente insolúvel por vários semanticistas (cf. Lyons 1986:43ss).

⁵ Frequentemente, palavras com sentidos muito distintos têm formas morfológicas derivadas também distintas (*banco* 'assento' > *bancada*, *banqueta* vs. *banco* 'instituição bancária' > *bancário*, *banqueiro*), o que corrobora a adoção de uma metodologia sincrónica para se distinguir entre polissemia e homonímia.

⁶ O item *banco* pode também ser usado em expressões como *banco de areia* ou ainda *banco de sangue*, *banco de dados*, em que significa um lugar onde se guarda uma quantidade elevada da matéria ou das entidades em causa. É plausível que estes dois últimos usos instanciem um sentido relacionado com o de instituição financeira, na medida em que todos remetem para um lugar onde se guarda um volume elevado de alguma coisa.

Este critério histórico não é relevante para a maioria dos falantes porque estes não têm, de um modo geral, conhecimento da etimologia das palavras, nem do seu percurso histórico. Assim, existem palavras formalmente idênticas, com o mesmo étimo e, conseqüentemente, com sentidos historicamente relacionados, mas que não são intuídas como relacionadas pelos falantes de uma determinada geração (ou seja, para esses falantes, essas palavras são homónimas e não polissémicas), como, p.e., a palavra *curso*, do latim *cursus*, que pode significar, por um lado, 'ato de correr, movimento' e, daí, 'caminho percorrido por um rio', e, por outro lado, 'programa de estudos, lição, disciplina', sentidos que o falante não relaciona.

8.3.1 Homonímia

Como vimos, na perspetiva sincrónica seguida neste capítulo, palavras que têm a mesma forma gráfica e fonética mas sentidos diferentes são consideradas palavras homónimas. Por exemplo, o nome *canto* é ambíguo entre duas palavras diferentes, cada uma com sentidos distintos: uma ocorre em expressões como *o canto da sala* e a outra em expressões como *o canto do rouxinol*. Outros casos de homonímia são, por exemplo, *banco* 'assento' e *banco* 'instituição bancária', *cavalo* 'animal equídeo' e *cavalo* 'heroína (droga)'⁷.

Em todos estes exemplos, a mesma forma tem, atualmente, sentidos suficientemente diferentes para ser tratada no âmbito da homonímia, independentemente de considerações sobre as suas origens etimológicas, que são diferentes em *canto* e em *banco*, ao contrário do que acontece em *cavalo*.

Muitas das palavras homónimas pertencem a classes diferentes: *foca* (verbo, em *a máquina foca bem*) vs. *foca* (nome, em *eu vi a foca*); *são* (adjetivo, em *ele está são*) vs. *são* (verbo, em *elas são altas*); *mato* (nome, em *estamos no mato*) vs. *mato* (verbo, em *eu mato as moscas*), entre muitas outras. Este tipo de homonímia, que diz respeito a palavras homónimas não pertencentes à mesma classe, é designado por **homonímia parcial**.

Outro tipo de homonímia, a **homonímia absoluta**, diz respeito a palavras que pertencem à mesma classe, como, por exemplo, *vela* 'objeto de cera' vs. *vela* 'tecido para propulsão eólica'; *banco* 'instituição financeira' vs. *banco* 'assento'; *manga* 'fruto' vs. *manga* 'parte de uma peça de vestuário'; *pena* 'castigo judiciário' vs. *pena* 'órgão cutâneo que reveste o corpo das aves'; *letra* 'sinal gráfico' vs. *letra* 'título de crédito'. Todas estas palavras pertencem à classe dos nomes e possuem sentidos independentes.

8.3.2 Polissemia

Como foi dito, designa-se como **polissemia** a associação de uma forma lexical única a sentidos diferentes que mantêm entre si alguma relação. Dois processos analógicos muito produtivos e com um papel importante neste domínio da significação são a **metáfora** e a **metonímia**. No primeiro caso, há transferência de sentido resultante de uma associação por semelhança. Assim, por exemplo, a palavra *joia* tem o sentido básico de ornamento valioso em *usava uma joia na lapela* e, por semelhança,

⁷ O nome de "cavalo" para heroína vem do efeito que produz, semelhante ao de um coice de cavalo. Há, portanto, uma relação de analogia, mas nem todos os falantes têm conhecimento disto.

o de pessoa com qualidades valiosas em *a tua irmã é uma joia*. No segundo caso, há expansão ou restrição de sentido por contiguidade. Essa contiguidade pode ser (i) entre o continente e o conteúdo (*copo* significa recipiente em *enche o copo com água* ou o líquido que o recipiente contém – tipicamente uma bebida alcoólica – em *vamos beber um copo*); (ii) entre o todo e a parte (*teto* significa uma parte da casa em *pintei o teto de branco* ou toda a casa em *não tem um teto onde se abrigar*); (iii) entre a matéria-prima e o produto (*prata* tem o sentido de metal precioso em *a prata vale menos do que o ouro* e de objeto feito de prata em *roubaram-lhe todas as pratas*).

Outro exemplo de polissemia é a palavra *livro*, que pode significar o conteúdo intelectual de uma obra escrita (em frases como *o livro é extremamente polémico*) ou o suporte físico no qual o texto está escrito (em frases como *o rato roeu o livro*). Embora os dois sentidos (conteúdo vs. suporte) sejam diferentes, apresentam relações metonímicas que o falante reconhece.

Podemos distinguir dois tipos de polissemia, a **polissemia regular** e a **polissemia irregular**. Na polissemia regular, todas as palavras pertencentes a uma determinada classe semântica exibem o mesmo padrão polissêmico. Um exemplo é o das palavras que remetem para uma obra escrita (*carta, diário, dicionário, jornal, livro de código, livrete, manual, panfleto, relatório, revista, etc.*), que exibem exatamente a mesma dualidade de sentidos da palavra *livro*. Se for dito a um falante do português que existe uma palavra nova, por exemplo *lídito* (palavra inventada por nós), que remete para um tipo de obra escrita mas cujo significado exato não é bem conhecido, os falantes concluirão que podem usar essa palavra tanto com o sentido de conteúdo como com o sentido de suporte material.

No caso da polissemia irregular, não existe qualquer padrão polissêmico partilhado pelas palavras da mesma classe semântica. Por este motivo, as palavras e os seus sentidos têm de ser aprendidos um a um, não podendo ser inferidos a partir da classe semântica. Por exemplo, a palavra *coração* tem um sentido biológico, quando designa um órgão muscular, e um sentido emocional, quando designa metaforicamente a sede de algumas emoções. Contudo, esta dualidade de sentidos não é observada em outras palavras da mesma classe semântica, tais como *bexiga, pâncreas, pulmão, rim, etc.*⁸

Na polissemia regular, pode acontecer que os vários sentidos da palavra estejam simultaneamente disponíveis na mesma frase. Por exemplo, é possível usar os dois sentidos da palavra *livro* em *o livro que está em cima da mesa contém erros sobre a história de Portugal*. Aqui, o objeto físico livro está a ser localizado no espaço, ao mesmo tempo que se qualifica o seu conteúdo intelectual. Este tipo de polissemia regular é chamado **polissemia compatível**.

Contudo, nem todas as palavras polissêmicas podem ser usadas desta forma; há um segundo tipo de polissemia regular a que chamamos **polissemia incompatível**. Como exemplo, consideremos a palavra *colher*, que pode ser usada no sentido de utensílio recipiente ou também, por metonímia, como medida de quantidade que o recipiente pode conter (cf. Cap. 21). Os dois sentidos são ilustrados, respetivamente, em *a colher está torta* e em *ponha uma colher de manteiga no refogado*. Esses

⁸ Também a palavra *estômago* possui uma dualidade de sentido ('órgão' e 'ânimo') semelhante à da palavra *coração*, mas o sentido de 'ânimo' apenas se verifica na expressão idiomática *não ter estômago para*.

sentidos são observáveis em palavras da mesma classe: *balde, copo, pá, panela, pipa, pipeta, saco* ou *tacho*. Contudo, a polissemia incompatível não permite que estejam ativos na mesma frase, em simultâneo: cf. a estranheza de *#a colher de manteiga que está torta deve ser posta no refogado*, uma frase que não pode ser interpretada como uma instrução para colocar uma certa porção de manteiga num refogado usando uma colher torta. Abaixo são listados exemplos de palavras pertencentes a estas duas classes.

❖ Polissemia compatível

- (1) a. O *relatório* era falso. [conteúdo]
 b. O *relatório* foi rasgado. [suporte]
 c. O *relatório* que foi rasgado era falso. [conteúdo + suporte]

(outros exemplos: *livro, manual, carta, etc.*)

- (2) a. O *hospital* despediu o segurança. [instituição]
 b. O *hospital* foi assaltado. [edifício]
 c. O *hospital* que foi assaltado despediu o segurança. [instituição + edifício]

(outros exemplos: *banco, escola, empresa, etc.*)

- (3) a. A *cidade* foi inundada. [local, região]
 b. A *cidade* está revoltada contra a Câmara. [população]
 c. A *cidade* que foi inundada está revoltada contra a Câmara. [local, região + população]

(outros exemplos: *cidade, vila, aldeia, país, etc.*)

❖ Polissemia incompatível

- (4) a. Ele partiu um *copo*. [recipiente]
 b. Ele pôs um *copo* de vinagre no estufado. [medida de quantidade]
 c. #Ele pôs um *copo* de vinagre estalado no molho. [recipiente + medida de quantidade]

(outros exemplos: *colher, caneca, jarro, balde, etc.*)

- (5) a. O *panda* está a dormir. [espécime (cf. Cap. 21)]
 b. O *panda* está atualmente ameaçado. [espécie]
 c. #O *panda* que está atualmente ameaçado está a dormir. [espécime + espécie]

(outros exemplos: *tigre, canário, sardinha, etc.*)

- (6) a. Ele usa um fio de prata. [substância]
 b. A prata tem o número atômico 47. [elemento]
 c. #Ele usa um fio de prata que tem o número atômico 47. [substância + elemento]

(outros exemplos: *ouro, ferro, platina, etc.*)

A polissemia não está restringida a formas nominais. Há também verbos, adjetivos e preposições que exibem um comportamento polissêmico.

Assim, os verbos *fechar* e *abrir* tanto podem remeter para ações físicas (*fechar a gaveta, abrir o portão*) como para atividades abstratas (*fechar o balanço, fechar a sessão, abrir a sessão, abrir a aula com uma anedota*, etc.). Também se podem considerar polissêmicos alguns verbos com realizações sintáticas distintas associadas a leituras semânticas parcialmente diferentes, ainda que relacionadas. É este o caso dos verbos que entram na “alternância causativa-incoativa”, como *afundar* (cf. Caps. 13 e 28). Este verbo ocorre numa frase transitiva com dois sintagmas nominais argumentais, desempenhando as funções de sujeito e de complemento direto (cf. (7a)), e numa frase intransitiva com apenas um sintagma nominal argumental, desempenhando a função de sujeito (cf. (7b)):

- (7) a. O submarino *afundou* o barco. (afundar = fazer com que vá ao fundo)
 b. O barco *afundou-se*. (afundar-se = ir ao fundo)

Nas duas frases, o sentido de *afundar* é essencialmente o mesmo. Nomeadamente, o mesmo sintagma nominal é complemento direto da versão transitiva (7a) e sujeito da versão intransitiva (7b), sendo-lhe atribuído pelo verbo o mesmo papel temático de paciente afetado em ambas as frases (cf. Cap. 11). No entanto, há uma diferença: em (7b) apenas se exprime a situação resultante do evento causal representado em (7a), sem qualquer menção, explícita ou implícita, desse evento causal e do agente causador (representados em (7a)). Tal como é característico da polissemia regular, o tipo de diferença semântica relacionada com a alternância causativa-incoativa é observado em muitos outros verbos, tais como *abrir, danificar, fechar, partir, submergir* ou *sujar* (cf. Cap. 28).

Existem também adjetivos polissêmicos como *alto, baixo, belo, bom, grande, mau, pequeno* ou *verdadeiro*. Por exemplo, em (8a), o adjetivo atributivo *grande*, em posição pré-nominal, denota a elevada capacidade de jogo do indivíduo, ao passo que, em (8b), o mesmo adjetivo atributivo, em posição pós-nominal, denota a sua elevada estatura física. Já em (8c), o mesmo adjetivo, agora com função predicativa, é ambíguo entre um sentido em que denota dotes físicos e outro em que denota dotes morais:

- (8) a. O Vasco é um *grande* jogador.
 b. O Vasco é um jogador *grande*.
 c. O Vasco é *grande*.

Frequentemente, o sentido dos adjetivos depende da entidade que qualificam, o que torna impossível listar todos os sentidos que podem ter. Este fenómeno é ilustrado nas frases em (9), onde o adjetivo *bom/boa* é sempre usado com uma conotação positiva, embora ligeiramente diferente em cada caso:

- (9) a. A carne não está *boa*. ('fresca')
 b. Foi *bom* passear pela cidade. ('agradável')
 c. A caldeirada estava *boa*. ('saborosa')
 d. Isto é *bom* para as enxaquecas. ('curativo')
 e. Este nó é *bom* de se desfazer. ('fácil')
 f. Ela é uma pessoa *boa*. ('bondosa')
 g. Ela é uma *boa* médica. ('competente')

O mesmo tipo de comportamento polissêmico pode ser observado em relação a adjetivos como *falso*, *lento*, *novo*, *rápido*, *velho*, *veloz*, *verdadeiro*: cf., por exemplo, *este é um jogo rápido*, *este carro é rápido* e *este teste é rápido*. No primeiro caso, *rápido* não qualifica a velocidade do jogo, mas sim a das jogadas. No segundo caso, o adjetivo qualifica a velocidade de um veículo, não a rapidez da condução. E, no terceiro caso, o adjetivo qualifica o tempo em que se pode fazer ou ler o teste.

Também existe polissemia nas preposições. Por exemplo, a preposição *em* significa inclusão, mas este sentido pode tomar diferentes matizes, alguns deles abstratos, tal como ilustrado em (10) (cf. Cap. 32):

- (10) a. O livro está *em Lisboa*. [inclusão geográfica]
- b. O futuro está *em ti*. [inclusão metafórica]
- c. O peixe está *em óleo quente*. [inclusão física]
- d. A avestruz está *em pânico*. [inclusão psicológica]
- e. O sistema está *em fase de testes*. [inclusão abstrata]
- f. Um dos condutores está *em contramão*. [inclusão numa situação]
- g. O relatório está *em disquete*. [inclusão num suporte]
- h. O meu quadro está *em pedaços*. [inclusão num estado]
- i. A minha vida está *em pedaços*. [inclusão metafórica num estado]

8.4 Sinonímia

Chamam-se **sinónimas** as palavras distintas, da mesma classe lexical, com sentidos semelhantes (cf. *japonês* e *nipónico*, *nitrogénio* e *azoto*, *narcótico* e *estupefaciente*, *amável* e *gentil*, *célebre* e *famoso*, *horrível* e *horroroso*, *abelha-mestra* e *abelha rainha*, etc.).

A **sinonímia absoluta** é um fenómeno extremamente raro nas línguas humanas. De facto, praticamente não existem palavras que tenham exatamente o(s) mesmo(s) sentido(s) e as mesmas condições de uso, i.e., que possam substituir-se uma à outra num enunciado sem provocar qualquer diferença de significado expressivo ou de adequação pragmática do enunciado ao contexto situacional ou discursivo. Por vezes, existem também diferenças “colocacionais” entre sinónimos (i.e., que têm a ver com a sua capacidade combinatória com outras palavras; cf. Firth 1951:190ss e Lyons 1986:50ss): assim, *célebre*, mas não *notável*, pode ocorrer com o advérbio *tristemente*: cf. *tristemente célebre* mas não *#tristemente notável* (cf. Cap. 9).

Por este motivo, é mais correto (e produtivo) restringir a definição de sinonímia aos aspetos do significado que têm a ver com o valor de verdade das frases nas quais se integram os termos em questão⁹ – a chamada **sinonímia proposicional** (cf. Cruse 2004). Com base neste critério, são sinónimas as palavras ou expressões linguísticas que se possam substituir numa frase sem alterar o seu valor de verdade (cf. *muitos japoneses/nipónicos emigraram para o Brasil*, *o azoto/nitrogénio faz parte do ar que respiramos*, *o tráfico de narcóticos/estupefacientes pode dar cabo de um país*, *a Maria é extremamente amável/gentil*, *esse filme é horrível/horroroso*, *a {abelha-mestra/abelha rainha} acabou de entrar na colmeia*, etc.).

⁹ Dizendo melhor, que têm a ver com o valor de verdade das proposições expressas por essas frases (cf. Cap. 11). Para simplificar a exposição, continuamos a usar o termo “frase” neste contexto.

Quando uma palavra é polissêmica, determinados sinónimos podem abranger apenas algum ou alguns dos seus sentidos; nesse caso, a sinonímia diz-se **parcial**. Por exemplo, o nome polissêmico *alto* tem *cume* como sinónimo em contextos como *o alto/cume da montanha*, mas não em contextos como *ter um alto na cabeça* (cf. a estranheza de *#ter um cume na cabeça*); neste contexto, *alto* tem como sinónimo *protuberância* (cf. *tem uma protuberância na cabeça*); a palavra *caro* tem como sinónimo *dispendioso* em *o aluguer é caro/dispensioso*, mas tem como sinónimo *querido* em fórmulas de cortesia como *meu caro/querido amigo*.

Há palavras com sentidos semelhantes mas também com algumas diferenças semânticas que podem aparecer realçadas em determinados contextos, podendo, nalguns deles, a sua substituição alterar subtilmente o valor de verdade da frase; é o caso de palavras como *vaidade* e *arrogância*, *orgulho* e *soberba*; *célebre* e *notável*; e *guiar* e *conduzir*. Estas palavras são chamadas **quase sinónimos** ou **parassinónimos**¹⁰. Por exemplo, *guiar* está usualmente associado ao ato simples de orientar um volante ou um guiador, ao passo que *conduzir* remete para uma atividade mais complexa que inclui a manipulação de outras peças (pedais, alavancas, manetes, etc.). Assim, dizemos que um piloto conduz um avião, um submarino, etc., mas não que o guia.

Em muitos casos, as palavras sinónimas são variantes dialetais da mesma língua (cf. *sertã* e *frigideira*; *quarto de banho* e *casa de banho*; *borrego*, *cordeiro* e *anho*, *café*, *bica* e *cimbalino* (no sentido de café de máquina servido em cafés e restaurantes), *fino* e *imperial* (no sentido de copo de cerveja tirada à pressão)¹¹.

Entre variedades nacionais do português, há também casos de sinonímia, como, por exemplo, *autocarro* (português europeu), *ônibus* (português do Brasil) e *machim-bombo* (português de Angola e de Moçambique). As palavras *ônibus* e *machimbombo*, para além de significarem transporte urbano de passageiros, tal como *autocarro*, têm outro sentido, correspondente ao da palavra do português europeu *camioneta* (transporte interurbano de passageiros).

A sinonímia também tem uma dimensão temporal, ou porque uma das palavras sinónimas é um arcaísmo (cf. *bragal* como sinónimo de *enxoval*, *ósculo* como sinónimo de *beijo*), ou porque é um neologismo (*rebaixas* como sinónimo de *saldos*, *flop* como sinónimo de *fracasso*).

Também entre registos de língua diferentes ocorrem frequentemente sinónimos: cf. *pai* (termo neutro) vs. *papá* (sinónimo familiar); *comer* (termo neutro) vs. *papar* (sinónimo familiar) e *morfar* (sinónimo popular). Outro exemplo é o das várias palavras que denotam intoxicação alcoólica, como *bebedeira* e os seus correspondentes populares: *borracheira*, *piela*, *pifo*, *tosga*, a par de outros, menos partilhados pela generalidade dos falantes, como *bezana*, *buba*, *cadela*, *cardina*, *manta*, *narda*, etc.

[1] Muitos dos termos da linguagem científica, incluindo nomes latinos e fórmulas químicas, são sinónimos de termos da linguagem corrente (cf., entre outros, *síndrome de Down* e *mongolismo*, *cefaleia* e *dor de cabeça*, OH_2 e *água*, *merluccius merluccius* e *pescada*). O termo científico, no entanto, tem um sentido muito restrito, que se

¹⁰ Os parassinónimos são, no entanto, frequentemente apresentados como sinónimos nos dicionários de uso.

¹¹ Com o tempo, a variação dialetal pode esbater-se ou mesmo desaparecer, e as palavras sinónimas passam a ser usadas por falantes do mesmo dialeto, quer como termos equivalentes quer com associações contextuais ou estilísticas diferentes.

reduz à nomeação científica da espécie ou categoria em questão, abstraindo de toda a riqueza expressiva do termo corrente e das suas várias associações de sentido com as outras palavras da língua. Por esse motivo, na maioria dos casos, o termo científico tem um contexto de uso limitado, restringido a textos científicos e legais, dicionários e enciclopédias, comunicação entre especialistas, etc. Por exemplo, *cefaleia* é termo usado por médicos, sendo *dor de cabeça* expressão corrente; *mongolismo* e *síndrome de Down* são ambos termos científicos, mas o primeiro é de uso mais corrente, sendo hoje em dia menos utilizado do que o segundo entre os especialistas; e ninguém diz que quer para o jantar *merluccius merluccius* com molho tártaro.

8.5 Antonímia

As palavras da mesma classe lexical com formas diferentes e sentidos opostos chamam-se **antónimos**, designando-se por **antonímia** a relação semântica entre elas (cf. *alto* e *baixo*, *despedir* e *contratar*, *macho* e *fêmea*, *noite* e *dia*, entre muitos outros).

A natureza semântica da oposição que é estabelecida entre palavras antónimas pode assumir várias formas, das quais destacamos quatro classes principais.

❖ Antonímia complementar

Os **antónimos complementares** são aqueles que instanciam uma relação que só possui dois pontos de oposição possíveis, isto é, em que os dois antónimos têm um sentido incompatível entre si e se excluem mutuamente. Neste caso, uma entidade que pode ser descrita por um deles não o pode ser pelo outro. Para além disso, não existem valores intermédios entre os conceitos opostos. É este o caso, p.e., de adjetivos como *par/ímpar*, *existente/inexistente* e *vivo/morto* (no sentido literal, não metafórico, destes dois últimos termos – cf. Nota 12). Ou seja, a negação de um implica a afirmação do outro: um número natural é par ou ímpar, uma pessoa está viva ou morta, etc.

Devido ao facto de não existirem estados intermédios entre dois antónimos complementares, estes não podem combinar-se com advérbios que especificam grau ou intensidade variável, tais como *muito*, *pouco*, *mais*, *menos* ou *bastante*: cf. **ele está muito/pouco morto*, **a água é mais inexistente em Marte do que na Terra* ou **este número é bastante ímpar* (mas cf. Nota 12).

❖ Antonímia de grau

Os **antónimos de grau** são aqueles que se definem numa escala contínua com pontos intermédios entre dois extremos opostos; são exemplos pares de adjetivos como *quente/frio*, *molhado/seco*, *fácil/difícil*, *novo/velho*, *limpo/sujo*, *sóbrio/bêbedo*, *curto/longo*, *alto/baixo*, *puro/impuro*, *amar/odiar*, *rir/chorar*, *longe/perto*. Estes antónimos podem combinar-se com advérbios que especificam grau ou intensidade: cf. *o chão está muito molhado*, *é mais difícil sobreviver em Marte do que na Terra*, *ele está bastante bêbedo*, etc.

Neste tipo de antonímia, existem com frequência palavras que remetem para estados intermédios, tais como *morno* (estado entre *quente* e *frio*), *húmido* (estado entre *molhado* e *seco*), ou *gostar* e *detestar* (par de antónimos situados entre os antónimos mais extremos *amar* e *odiar*). Nalgumas escalas existe mais do que um termo aplicável aos seus pontos intermédios. Assim, na escala de temperatura, tem-se *gelado*, *frio*, *fresco*, *morno*, *quente*, *tórrido*. Nem sempre é claro em que ponto

da escala se encontram tais palavras, mas estas escalas contrastam claramente com as dos antónimos complementares, onde não existem termos intermédios entre os dois antónimos¹².

❖ Antonímia reversível

Os **antónimos reversíveis** denotam extremos opostos de uma escala espacial, como o par *dentro/fora*, e podem envolver movimento, orientação ou localização, de tal modo que as entidades a que se aplicam podem “circular” de modo contínuo pelos vários pontos da escala espacial num sentido ou noutro, como ilustram os pares *expandir/encolher*, *abrir/fechar*, *tirar/pôr* e *subir/descer*. Assim, p.e., a ação de abrir uma porta pode ser invertida na ação de a fechar, o inverso de estar dentro é estar fora (e as duas situações são reversíveis), depois de enrolar uma corda é possível desenrolá-la, etc.

Os antónimos reversíveis são semelhantes aos antónimos de grau, na medida em que admitem termos de comparação ou intensidade: cf. *um elevador subiu mais do que o outro* ou *o camião está mais dentro da garagem do que fora*. Muitos verbos criados pelo prefixo *des-* são antónimos reversíveis da forma correspondente sem o prefixo, tais como *atar/desatar*, *apertar/desapertar*, *enrolar/desenrolar* e *ligar/desligar*.

❖ Antonímia relacional

Os **antónimos relacionais** denotam uma mesma situação, mas a partir de uma perspetiva oposta. Por exemplo, os verbos *dar* e *receber* remetem ambos para uma ação de transferência, mas ao passo que *dar* representa a transferência a partir da perspetiva de quem dá, *receber* representa-a a partir da perspetiva de quem recebe. Ou seja, para uma determinada situação particular, quer se use *dar* quer se use *receber*, estão envolvidos os mesmos intervenientes, com os mesmos papéis temáticos (fonte e destinatário), ainda que a função gramatical das expressões que representam esses papéis temáticos não sejam as mesmas: assim, a frase *o Pedro deu um beijo à Leonor* e a frase *a Leonor recebeu um beijo do Pedro* representam a mesma situação. De uma forma mais geral, ‘*x dar algo a y*’ implica ‘*y receber algo de x*’, e vice-versa (relativamente a uma determinada situação).

Outros exemplos de antónimos relacionais são *ensinar/aprender*, *empregado/patrão*, *professor/aluno*, como se pode ver em pares de frases que exprimem essencialmente a mesma situação, tais como *o Bruno é empregado do Carlos* e *o Carlos é patrão do Bruno*.

O verbo *alugar* é interessante a este respeito, pois pode expressar duas perspetivas diferentes sobre uma situação. De facto, este verbo tanto pode denotar o ato de ceder a outrem o direito ao uso de um bem móvel ou imóvel, com contrapartidas financeiras, como o ato de adquirir o direito a usar esse bem, mediante um pagamento: cf. *o João alugou a casa à Maria*, que tanto pode significar que o João, enquanto senhorio, cedeu à Maria o direito de usar uma determinada casa (provavelmente a sua) mediante um pagamento, como pode significar que o João, enquanto inquilino, adquiriu da Maria o direito de usar uma determinada casa (provavelmente propriedade da Maria). Assim, este verbo é polissémico entre duas interpretações que estão associadas entre si por uma relação de antonímia relacional.

¹² Num sentido não literal, figurativo ou metafórico, alguns antónimos complementares podem ser usados como antónimos de grau e combinar-se com advérbios de grau. É este o caso, p.e., do par *vivo/morto*. Assim, é possível dizer *essa pessoa está [mais morta do que viva/meio morta]*.

8.6 Hiponímia-hiperonímia

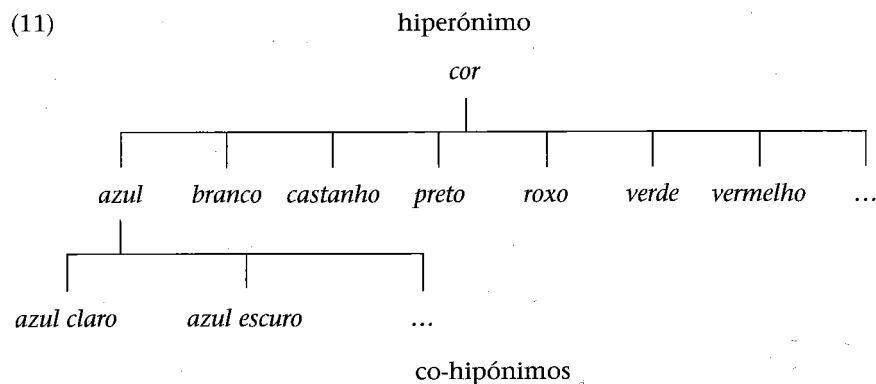
Freqüentemente, palavras que pertencem à mesma área semântica encontram-se relacionadas entre si em termos da maior ou menor especificidade do seu sentido. Por exemplo, o item *azul* é mais específico e, logo, mais informativo do que o item *cor*, por denotar um tipo particular de cor; *azul claro* é mais específico e informativo do que *azul* por denotar um tipo particular de azul, e assim sucessivamente. Estas relações são chamadas **relações de hiponímia-hiperonímia**. Assim, *azul claro* é **hipónimo** de *azul* e *azul* é **hipónimo** de *cor*. Inversamente, *azul* é **hiperónimo** de *azul claro* e *cor* é **hiperónimo** de *azul*. Para dar outro exemplo de uma área semântica diferente, *maçã* ou *pera* são hipónimos de *fruta*, e este termo é hiperónimo daqueles.

A relação lógica entre um hipónimo e o seu hiperónimo é a de inclusão. Assim, do ponto de vista extensional, a classe (ou conjunto) de entidades denotada por um hiperónimo contém (ou inclui) a classe (ou conjunto) de entidades denotada pelo hipónimo. Assim, p.e., a classe das entidades coloridas contém a classe das entidades azuis, a classe das frutas contém a classe das peras, etc.

Inversamente, do ponto de vista intensional, os conceitos que formam o sentido de um hipónimo formam uma classe mais abrangente, rica e informativa do que a classe de conceitos que formam o sentido do hiperónimo. Assim, p.e., o sentido de *azul* contém não só a noção de *cor*, como a noção adicional que tem a ver com uma determinada localização no espectro das ondas luminosas; o sentido de *maçã* inclui o sentido de *fruta* mais uma série de propriedades que individualizam as maçãs relativamente às peras, às uvas, etc.

As relações de hiponímia e hiperonímia são simétricas (mudando os predicados): sempre que x é hipónimo de y , y é hiperónimo de x , e vice-versa; p.e., *azul* é hipónimo de *cor* e *cor* é hiperónimo de *azul*.

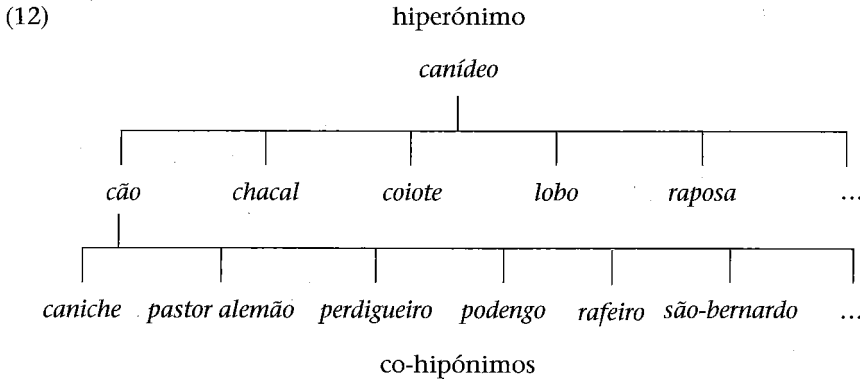
Geralmente, um hiperónimo possui mais do que um termo hipónimo. Assim, todas as palavras que designam cores particulares são hipónimas do termo *cor*: *amarelo*, *branco*, *castanho*, *cor-de-rosa*, *laranja*, *preto*, *verde*, etc. As palavras que partilham o mesmo hiperónimo são chamadas **co-hipónimas**. Assim, p.e., as relações de hiponímia/hiperonímia do termo *cor* instanciam uma hierarquia de palavras ilustrada (de modo incompleto) em (11):



A mesma palavra pode entrar numa relação de hiponímia com mais do que um termo da mesma área semântica, numa organização hierárquica articulada em níveis

distintos; assim, p.e., *azul claro* é um **hipónimo indireto** de *cor*, visto que é **hipónimo direto** de *azul*, e, por sua vez, *azul* é hipónimo direto de *cor*.

Em alguns casos, a hierarquia de hiponímia/hiperonímia que se estabelece é relativamente complexa, como, p.e., a de *canídeo*, que se apresenta a seguir:



O termo *canídeo*, por sua vez, faz parte de uma hierarquia muito mais vasta (*canídeo* é hipónimo de *mamífero*, e esta palavra, por sua vez, é hipónimo de *animal*).

Existem vários critérios que permitem a identificação de um par de palavras relacionadas desta forma. Assim, é possível dizer que existe uma relação de hiponímia entre um par de palavras (aqui simbolizadas como “*x*” e “*y*”) se estas não forem sinónimas e se a verdade de *a* é *x* implicar necessariamente a verdade de *a* é *y*. Este critério baseia-se no facto de um hipónimo remeter para um conceito mais específico do que o seu hiperónimo. Assim, p.e., se for verdade que o Rex é um caniche, é também necessariamente verdade que é um cão, visto que *caniche* é um hipónimo de *cão*. É, assim, possível comprovar em (12) que *caniche* é um hipónimo de *cão*¹³.

Um outro critério para a identificação de hipónimos/hiperónimos resulta da impossibilidade de coordenar um hipónimo com um hiperónimo (cf. (13a)), contrariamente à coordenação de dois co-hipónimos (cf. (13b)) ou de dois termos que se situem em ramificações distintas de uma hierarquia de hiponímia-hiperonímia (cf. (13c)).

- (13) a. #Eu tenho em casa um caniche e um cão.
 b. Eu tenho em casa um caniche e um pastor alemão
 c. Eu tenho em casa um cão/caniche e um {gato/gato siamês}.

A estranheza de (13a) deve-se ao facto de uma estrutura de coordenação de sintagmas nominais representar normalmente um conjunto de entidades distintas, cada uma correspondendo a um dos termos coordenados. Em (13b,c), a coordenação exprime um conjunto com dois animais distintos, mas em (13a) isso não

¹³ É fácil para o leitor verificar que as relações de implicação inversas, i.e., de hiperónimo para hipónimo, não se verificam. Se Rex é um mamífero, não é necessariamente um canídeo, se é um canídeo, não é necessariamente um cão, e assim por diante.

acontece, visto que todos os caniches são cães. Naturalmente, se mais informação for acrescentada ao termo coordenado *um cão*, então a frase tornar-se-á aceitável: cf. *eu tenho em casa um caniche e um cão rafeiro*. Ou seja, nesta frase a coordenação é possível porque *caniche* é co-hipónimo de *cão rafeiro*.

Embora a relação de hiponímia/hiperonímia seja usualmente bastante intuitiva, nalguns casos não o é. Por exemplo, há uma diferença subtil entre as palavras *pássaro* e *ave*, que se reflete no contraste entre as duas frases de (14):

- (14) a. Todos os pássaros são aves.
b. #Todas as aves são pássaros.

O exemplo (14b) é estranho porque a palavra *ave* é um termo genérico usado para qualquer ser bípede com o corpo coberto de penas, com asas e bico, ao passo que a palavra *pássaro* é um termo mais específico, que é usado para qualquer ave capaz de voar. Assim sendo, *pássaro* é hipónimo de *ave*, o que torna ilegítima a relação de inclusão expressa por (14b).

Nem todos os co-hipónimos de um hiperónimo estão em pé de igualdade. Alguns são melhores representantes do que outros da categoria representada pelo hiperónimo, sendo, por isso, considerados prototípicos (sobre esta noção, cf. a introdução a este capítulo). Por exemplo, quando um falante do português pensa na palavra *ave*, há vários co-hipónimos que se apresentam como representantes mais naturais desta classe (*águia*, *andorinha*, *canário*, *pardal*, *pombo*, etc.) e outros que se apresentam como representantes menos óbvios (*faisão*, *galinha* ou *pinguim*). Da mesma forma, se for pedido a alguém que enumere exemplos de cores, é mais provável que *azul*, *vermelho* e *amarelo* sejam enunciados antes de cores como *bege*, *magenta* ou *lilás*. Embora todos estes termos sejam exemplos de cor, alguns são hipónimos mais salientes na organização do léxico mental dos falantes do português.

A divisão dos co-hipónimos em prototípicos e não prototípicos pode ter consequências linguísticas a nível lexical. Para dar um exemplo, o nome coletivo *cardume* denota um grupo de peixes nadando em proximidade, e pode ser especificado pela espécie particular de peixes que formam o cardume (p.e., *cardume de sardinhas*). No entanto, é relativamente estranho falar de um cardume de tubarões, pelo facto de os tubarões não representarem exatamente o protótipo de um peixe. De facto, de modo geral, na nossa cultura, as classes prototípicas de peixes são aquelas cujos membros são animais comestíveis de porte relativamente pequeno (sardinhas, pescadas, linguados, carapaus, etc.). Ora, são precisamente os nomes destas classes que, também tipicamente, se aceitam como complementos do coletivo *cardume*. Repare-se que isso nada tem a ver com o facto de os tubarões não nadarem em grupo (os tubarões-martelo, na realidade, deslocam-se frequentemente em grupo), mas apenas com a aplicabilidade ou não aplicabilidade linguística do nome da classe ao nome coletivo. Naturalmente, também nesta área, falantes diferentes têm intuições diferentes, dependendo da cultura ou subcultura na qual estão integrados, do seu conhecimento e da sua experiência.

Existem certas relações de inferência lógica entre asserções contendo hipónimos e hiperónimos, que variam consoante o tipo de predicado. Por exemplo, verbos como *detestar* implicam que se for verdade que *x detesta y*, então é necessariamente

verdade que *x* detesta todos os hipónimos (diretos ou indiretos) de *y*. Este padrão de inferências é ilustrado abaixo, em que as frases de (16) são verdadeiras sempre que a frase (15) for verdadeira, isto é, em que a asserção *ele detesta cães* implica que ele detesta todos os tipos de cães, como se exemplifica em (16):

(15) Ele detesta cães.

- (16) a. Ele detesta caniches.
 b. Ele detesta galgos.
 c. Ele detesta cachorros.

Repare-se que *detestar* não permite inferências do hipónimo para o hiperónimo. Se a frase (17a) é verdadeira, não é necessário que a frase (17b) também seja verdadeira:

- (17) a. Ele detesta caniches.
 b. Ele detesta cães.

Outros verbos, como os verbos de percepção (*ouvir, ver, etc.*) ou os verbos que denotam uma ação em que o complemento direto é um paciente afetado (*afagar, alimentar, comprar, fotografar, vacinar, etc.*) permitem a inferência inversa. Tomando *ver* como verbo representativo, se for verdade que *x* viu *y*, então não é obrigatoriamente verdade que *x* viu todos os hipónimos de *y*. Assim, o facto de a frase (18) ser verdadeira não implica que as frases de (19) o sejam:

- (18) Ele viu cães no jardim.
 (19) a. Ele viu caniches.
 b. Ele viu galgos.
 c. Ele viu cachorros.

Pelo contrário, verbos como *ver* permitem a inferência do hipónimo para o hiperónimo, ao contrário de *detestar*. Assim, a verdade da frase (20) implica a verdade das frases de (21):

- (20) Eu vi um cão.
 (21) a. Eu vi um canídeo.
 b. Eu vi um mamífero.
 c. Eu vi um animal.

A diferença entre os verbos de percepção e de ação (como *ver, afagar, alimentar, etc.*) e os verbos de atitude afetiva (como *detestar, odiar, etc.*) relativamente às suas inferências hiponímicas e hiperonímicas reside na forma como se combinam semanticamente com o seu complemento direto. Em *ver cães*, p.e., o verbo aplica-se a um conjunto de entidades na extensão do substantivo *cão* (por outras palavras, *ver cães* significa 'avistar seres que são cães'). Como um cão é, por definição, um canídeo, um mamífero e um animal (ou seja, o nome *cão* é caracterizável pelas propriedades semânticas 'canídeo', 'mamífero' e 'animal'), é possível, a partir de 'ver cães', inferir 'ver canídeos', 'ver mamíferos' e 'ver animais'. Contudo, a inferência não "desce" na hierarquia dos hipónimos (como *caniche*, p.e.: se vejo um cão, não

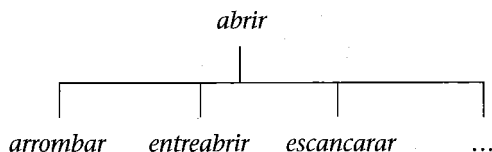
veja necessariamente um caniche), visto que estes contêm especificações semânticas adicionais que não estão contidas no hiperónimo.

Em contrapartida, em *detestar cães*, o verbo aplica-se à intensão do substantivo *cão*, ou seja, ao conjunto de propriedades que compõem o sentido desse substantivo. Assim, *detestar cães* significa 'ter uma opinião negativa em relação às propriedades que fazem uma entidade ser um cão'. Por inferência, essa atitude afetiva estende-se a todos os seres denotados por palavras que são hipónimos de *cão* (caniches, pastores alemães, cães rafeiros, etc.), visto que todas elas são caracterizáveis pela propriedade 'cão'. Contudo, a atitude não "sobe" na hierarquia dos hiperónimos (como, p.e., *canídeo* ou *mamífero*), visto que estes não se caracterizam pela propriedade 'cão' (odiar cães não implica que se odeiem lobos ou baleias).

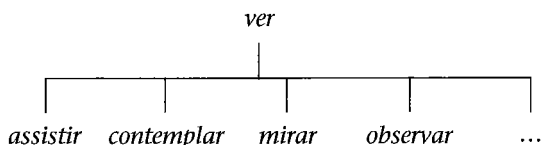
Verbos como *odiar*, *proibir*, *respeitar* e *temer* comportam-se da mesma forma que *detestar*. Este facto não é surpreendente, pois alguns destes verbos estão eles próprios relacionados com outros por meio de uma relação paralela à de hiponímia/hiperonímia nos nomes, relação usualmente referida como **troponímia**. Esta relação existe entre verbos que denotam eventos mais gerais (sobreordenados) e verbos que denotam eventos mais específicos (subordenados). Pode dizer-se, por exemplo, que essa relação existe entre *andar* e *coxear*, pois *coxear* pode ser definido como 'andar de um modo específico/particular' (cf. Barreto 2002).

Os verbos *desgostar*, *detestar* e *odiar* denotam um mesmo tipo de sentimento, embora com diferentes intensidades. Assim, sempre que a proposição 'x odeia y' é verdadeira, então a proposição 'x detesta y' também o é, mas não vice-versa: 'x detesta y' não implica 'x odeia y'. Em (22) e (23) ilustram-se duas hierarquias de hipónimos verbais (isto é, de relações de troponímia):

(22)



(23)



Como o hiperónimo e os seus hipónimos partilham parte do seu sentido, muitos processos lexicais aplicam-se tanto ao hiperónimo como aos seus hipónimos, diretos ou indiretos. Por exemplo, a relação de polissemia regular, discutida acima (cf. 8.3.2), observa-se de uma forma sistemática em todos os hipónimos de um hiperónimo. Assim, tanto *recipiente* como os seus hipónimos exibem dois sentidos polissémicos relacionados, o de 'utensílio' (cf. *amolguei o recipiente*) e o de 'capacidade' (cf. *entornei o recipiente*). Os hipónimos *colher*, *copo*, *panela* ou *tacho* possuem o mesmo tipo de polissemia observado no hiperónimo, como o leitor pode facilmente observar.

Também os hipónimos de um verbo tipicamente partilham parte do sentido do seu hiperónimo. Assim, p.e., todos os hipónimos de *abrir* em (22) selecionam o mesmo tipo de argumentos, combinam-se com o mesmo tipo de adjuntos adverbiais e podem ocorrer no mesmo tipo de estruturas, como exemplificado em (24) e (25):

(24) Ele {abriu/arrombou/entreabriu/escancarou} a porta com um tijolo.

(25) Ele {abriu/arrombou/entreabriu/escancarou} rapidamente a porta.

No entanto, um hipónimo verbal pode impor mais restrições do que o seu hiperónimo. Por exemplo, *abrir* permite exprimir a duração do estado resultante da ação através de um adjunto adverbial introduzido por *durante* (cf. a frase (26a), que significa que a janela esteve aberta durante cinco minutos depois de ter sido aberta); em contrapartida, *arrombar* apenas exprime a ação em si mesma, não permitindo exprimir a duração do estado resultante (cf. a estranheza de (26b)). Repare-se que *arrombar*, no entanto, permite exprimir a duração da própria ação, através de um adjunto adverbial introduzido por *em* (cf. (26c)):

- (26) a. Ele abriu a janela *durante cinco minutos*.
 b. #Ele arrombou a janela *durante cinco minutos*.
 c. Ele arrombou a janela *em cinco minutos*.

Caso semelhante ocorre com *matar* e *assassinar*. Tal como é ilustrado em (27), o hiperónimo *matar* não requer um agente volitivo ou humano, contrariamente ao hipónimo *assassinar*. Por essa razão, (27b) é semanticamente anómalo:

- (27) a. O maremoto matou Abel.
 b. #O maremoto assassinou Abel.

8.7 Meronímia-holonímia

Um **merónimo** é uma palavra que denota uma parte incluída num todo maior. O termo que denota esse todo, por sua vez, chama-se **holónimo**. Por exemplo, as palavras *capa*, *contracapa* e *página* são merónimos de *livro* e, inversamente, *livro* é holónimo dessas palavras. O verbo *ter* pode ser usado para exprimir a relação de meronímia-holonímia, sendo o sujeito o holónimo e o complemento o merónimo: cf. *o livro tem muitas páginas* (*página* é merónimo de *livro*), *a cidade tem várias avenidas* (*avenida* é merónimo de *cidade*), *a casa tem um telhado* (*telhado* é merónimo de *casa*), *o júri tem cinco membros* (*membro* é merónimo de *júri*), *um dia tem vinte e quatro horas* (*hora* é merónimo de *dia*), etc.

Um dos aspetos distintivos desta relação é o facto de a entidade designada por um merónimo não estar necessariamente presente como uma parte de todas as entidades designadas pelo holónimo. Por exemplo, *índice* é um merónimo de *livro*, embora não seja necessariamente verdade que todos os livros têm um índice.

Um segundo aspeto desta relação é o facto de se poderem estabelecer estruturas relacionais complexas, pois uma palavra pode ter mais do que um merónimo e mais do que um holónimo. Por exemplo, *livro* tem como merónimos *capa*, *contracapa* e *página*, e tem como holónimos *coletânea* (no sentido de 'coleção de livros

de um mesmo autor, ou com o mesmo tema') e *biblioteca* (conjunto ordenado de livros possuídos por uma entidade). Da mesma forma, cada um destes merónimos e holónimos pode ter os seus próprios merónimos e holónimos. Assim, a palavra *página* (no sentido do conjunto de texto escrito) tem como merónimo *parágrafo*, que, por sua vez, tem como merónimo *linha*, e assim por diante, até *caráter*.

Para além disso, um merónimo pode ter mais do que um holónimo em campos semânticos diferentes. Por exemplo, a palavra *janela* é um merónimo de variadas palavras, incluindo meios de transporte (*autocarro, camião, carro, metro, etc.*) e edificações (*cabana, casa, prédio, etc.*). Em todos estes casos, uma parte integral que consiste numa abertura com uma determinada forma e disposição, ou no revestimento dessa abertura, é chamada *janela*.

Existem várias classes de meronímia-holonímia, que passamos a descrever mais detalhadamente.

❖ Meronímia quantitativa

Na **meronímia quantitativa**, a relação parte-todo liga uma porção arbitrária a um todo sem partes distintas intrínsecas. Por exemplo, uma melancia não possui partes integrais distintas, como são, por exemplo, os gomos de um citrino. Contudo, um corte de parte de uma melancia é referido pelo nome *talhada*. Este termo requer uma forma específica (por exemplo, um cubo de melancia não é uma talhada) e requer um determinado tipo de holónimo (ver abaixo): um corte de queijo exatamente com a mesma forma e dimensão não é referido como *talhada*, mas sim como *fatia*. Em geral, classes de alimentos possuem termos específicos que referem as porções respetivas, como se ilustra em (28) (cf. também o Cap. 21):

- (28) a. talhada – melancia, melão, meloa, abóbora
 b. fatia – bolo, tarte, queijo, pão
 c. naco – pão, carne, presunto
 d. posta – peixe
 e. bife – vaca, porco, frango, peru

❖ Meronímia integral

Na **meronímia integral** ou **estrutural** (cf. Cap. 21), a parte é integral, individualizável, com funções próprias e claramente distinta do todo onde está inserida. É este o caso, por exemplo, de partes e peças, e o respetivo mecanismo onde estas se inserem. Alguns exemplos são dados em (29):

- (29) a. asa – chávena
 b. asa – pássaro
 c. gomo – laranja
 d. dedo – mão ou pé
 e. manga – camisa, vestido, blusa
 f. gargalo – garrafa
 g. tronco – árvore
 h. trinco – fechadura
 i. parede – casa

❖ Meronímia inclusiva

Na **meronímia inclusiva**, a relação parte-todo liga uma entidade autónoma, com a sua própria identidade, a uma coleção. Por exemplo, uma floresta é formada por árvores, uma alcateia é formada por lobos, etc. Por outras palavras, a meronímia inclusiva estabelece uma ligação entre nomes **coletivos** (ou **grupais**) e os nomes que designam as entidades que compõem esses coletivos. Apresentam-se alguns exemplos em (30):

- (30) a. lobo – alcateia
 b. árvore – floresta
 c. eucalipto – eucaliptal
 d. sobreiro – montado
 e. porco – vara
 f. osso – esqueleto
 g. tecla – teclado
 h. peixe – cardume

Em muitos casos, o sentido do nome que representa o holónimo não se reduz a uma versão plural do sentido do nome que representa o merónimo. Assim, p.e., uma alcateia não é simplesmente um conjunto de lobos, visto que inclui também a ideia de uma hierarquia bem definida no grupo; e uma floresta não é simplesmente qualquer conjunto de árvores (um conjunto de árvores arrancadas do chão e empilhadas num armazém não é uma floresta).

À semelhança do que se exemplificou em (28), alguns nomes coletivos não possuem só um merónimo: p.e., uma manada pode ser formada por vários tipos de animal de grande porte (bovinos, equinos, paquidermes, etc.).

❖ Meronímia material

Na **meronímia material**, a relação parte-todo liga uma substância ingrediente (tipicamente expressa por um nome não contável, ou massivo; cf. Caps. 21 e 25) a uma substância mais complexa. Por exemplo, *álcool* é um merónimo material de *vinho* porque é um dos seus ingredientes necessários, mas não uma parte distinta, discernível e funcional, do vinho (contrariamente, p.e., ao volante de um carro). Ou seja, o sentido da palavra *vinho* de alguma forma implica um componente – entre outros – denotado por *álcool*. Apresentam-se outros exemplos em (31):

- (31) a. açúcar – doce
 b. limão – limonada
 c. ginja – ginjinha
 d. gema – gemada
 e. nitroglicerina – dinamite
 f. carne – bife

Note-se que as relações de meronímia podem entrelaçar-se. Por exemplo, *carne* é um merónimo material de *bife*, o qual, por sua vez, é um merónimo quantitativo de *vaca*, *porco* ou *peru*.

Na maior parte destes exemplos, a substância designada pelo merónimo é um fator necessário para se poder utilizar o holónimo correspondente como nome da

substância complexa. Por exemplo, não existe ginjinha sem ginja, dinamite sem nitroglicerina, ou limonada sem limões. É interessante notar, no entanto, que isso não acontece em relação a todos estes pares. Devido a determinados desenvolvimentos tecnológicos, existe hoje em dia café sem cafeína (também designado por *déscafeinado*) e cerveja sem álcool; assim, em casos particulares, *cafeína* não é um merónimo de *café*, nem *álcool* um merónimo de *cerveja*.

❖ Meronímia de subatividade

Na **meronímia de subatividade**, a relação parte-todo liga uma ação a outra ação na qual a primeira está incluída. Por exemplo, *pagar* é uma atividade que faz parte de uma atividade mais complexa, referida como *comprar*; outros casos são exemplificados em (32):

- (32) a. falar – entrevistar
 b. mastigar – comer
 c. focar – fotografar

É importante notar que *pagar* e *comprar* não estão relacionados através de hiponímia-hiperonímia. As duas palavras pertencem a áreas semânticas distintas, embora estejam relacionadas pela meronímia de subatividade.

Repare-se que a subatividade designada pelo merónimo não se encontra necessariamente presente em todas as atividades ou situações designadas pelo holónimo. Por exemplo, não é necessário focar para fotografar, falar para entrevistar, ou mastigar para comer (cf. a expressão de censura *não comas sem mastigar!*).

Em certos casos, o merónimo e o holónimo de subatividade são palavras homófonas: cf., por exemplo, *pintar* ('aplicar tinta em alguma coisa' e 'criar uma pintura') e *escrever* ('inserir caracteres numa página' e 'criar um texto').

❖ Meronímia espacial

A **meronímia espacial** é um tipo de meronímia em que a relação parte-todo liga uma área espacial a uma outra, mais abrangente, e em que as fronteiras entre as duas áreas são algo subjetivas. Ou seja, informalmente, *x* é merónimo espacial de *y* se *x* é uma subárea de *y*. Por exemplo, *baixa* é uma subárea de uma *cidade* densamente urbanizada, ainda que as suas fronteiras possam não ser completamente definidas. Em (33) são dados alguns outros exemplos de meronímia espacial.

- (33) a. oásis – deserto
 b. palma – mão
 c. lombo – dorso
 d. testa – face
 e. nuca – cabeça
 f. cume – montanha

8.8 Relações proposicionais

Nem todas as propriedades semânticas lexicais têm a ver com relações entre palavras. Algumas destas propriedades manifestam-se em inferências que envolvem

unidades fráscas, e têm a ver, em particular, com a relação que se estabelece entre um predicador (cf. Cap. 11) e os seus argumentos. Nesta secção, descrevem-se três relações que têm sido particularmente estudadas na literatura da Lógica e da Semântica: a simetria, a reflexividade e a transitividade.

8.8.1 Relações e predicadores simétricos

Com a maioria dos predicadores que seleccionam dois argumentos (i.e., os predicadores de dois lugares; cf. Cap. 11), a ordem dos argumentos não é reversível. Assim, p.e., mantendo constante a referência dos nomes, a frase *a Joana gosta do Vasco* tem um significado diferente do da frase *o Vasco gosta da Joana*. Em particular, não se pode inferir a verdade de qualquer uma delas a partir da verdade da outra (ou seja, a Joana pode perfeitamente gostar do Vasco sem que o Vasco goste da Joana e vice-versa).

No entanto, existem predicadores de dois lugares que admitem uma tal reversibilidade, i.e., em que uma frase que corresponda ao esquema proposicional $a \ x \ b$ implica a frase que corresponde ao esquema $b \ x \ a$ e vice-versa (a e b representam argumentos, x representa o predicador)¹⁴. Assim, em (34a-c), a primeira frase implica a segunda e vice-versa:

- (34) a. O Vasco casou(-se) com a Joana. = A Joana casou(-se) com o Vasco.
 b. A Idalina é vizinha da Beatriz. = A Beatriz é vizinha da Idalina.
 c. O Manuel é irmão da Maria. = A Maria é irmã do Manuel.

Dizemos então que *casar(-se) (com)*, *(ser) vizinho (de)* e *(ser) irmão/ã (de)* são **predicadores simétricos**.

Como se observou acima, a maioria dos predicadores são **não simétricos**, como, p.e., os verbos *beijar*, *gostar (de)*, *ver* e o nome *amigo (de)*. Assim, a frase *o Vasco viu a Ana* não implica necessariamente a frase *a Ana viu o Vasco*; e o Pedro pode ser amigo do Paulo sem que o Paulo seja necessariamente amigo do Pedro.

Muitas línguas possuem uma construção gramatical que torna um predicador não simétrico num predicador simétrico: nomeadamente, a chamada **construção recíproca**. Nesta construção, os argumentos a e b constituem os dois termos de um sujeito coordenado, e no predicado ocorre o adjunto adverbial “de reciprocidade” *um + preposição + o outro* (*um do outro*, *uns com os outros*, etc.), como se ilustra em (35) (cf. Caps. 13 e 41)¹⁵:

- (35) a. O Vasco e a Joana beijaram-se (um ao outro).
 b. A Idalina e a Beatriz gostam uma da outra.
 c. O Manuel e a Maria viram-se (um ao outro).
 d. O Pedro e o Paulo são amigos (um do outro).

¹⁴ As relações de implicação descritas nesta subsecção e nas seguintes estabelecem-se não entre as frases enquanto entidades gramaticais, mas sim entre as proposições expressas por essas frases (cf. Cap. 11). Para não sobrecarregar a exposição, no entanto, usamos o termo “frase” em vez de “proposição”, pedindo ao leitor que faça o ajuste relevante.

¹⁵ Se o predicador é um verbo transitivo, este é necessariamente “conjugado” com o pronome *se* e o adjunto de reciprocidade é opcional, como em (35a, c); quando o predicador é um nome, o adjunto é igualmente opcional, como em (35d). Na coordenação, os termos também são reversíveis: comparar (35a) com *a Joana e o Vasco beijaram-se (um ao outro)*, etc.

Os predicadores simétricos também podem ocorrer nesta construção, como se ilustra em (36) (comparar com (34)):

- (36) a. O Vasco e a Joana casaram (um com o outro).
 b. A Idalina e a Beatriz são vizinhas (uma da outra).
 c. O Manuel e a Maria são irmãos (um do outro).

Existem predicadores, chamados **assimétricos**, que, por razões lógicas, nunca permitem uma inferência de simetria, como *pai (de)*, *mais alto (do que)*, entre outros. Com estes, obviamente, nem as frases com o esquema proposicional $a \times b$ são equivalentes às frases com o esquema $b \times a$ (cf. (37a)), nem é possível construir uma frase recíproca bem formada semanticamente (cf. (37b)):

- (37) a. O Eduardo é pai do Manuel. (\neq *o Manuel é pai do Eduardo*)
 b. #O Eduardo e o Manuel são pais um do outro.

8.8.2 Relações e predicadores reflexos

Quando os dois argumentos de um predicador de dois lugares referem a mesma entidade, obtém-se uma **frase reflexa**. Em português, estas frases contêm uma só ocorrência do argumento, com a função de sujeito, e no seu predicado ocorre um adjunto adverbial “de reflexividade”, preposição + *si próprio* ou preposição + *si mesmo* (e suas variantes morfológicas em género e número), como se ilustra em (38) (cf. Caps. 13 e 41)¹⁶:

- (38) a. O doente lavou-se (a *si próprio*).
 b. O Pedro não gosta de *si mesmo*.
 c. A Maria é amiga de *si própria*.

A verdade destas frases é, obviamente, contingente, ou seja, não é necessariamente verdadeiro que um doente se lave a *si próprio*, nem que o Pedro de que se fala goste de *si mesmo*, nem que a Maria que eu conheço seja amiga de *si própria*.

No entanto, existe um pequeno número de predicadores com os quais, por motivos lógicos, as frases reflexas são necessariamente verdadeiras, qualquer que seja o contexto. Ou seja, com estes predicadores, chamados **predicadores reflexos**, o esquema proposicional $a \times a$ (a designa os argumentos, x o predicador), linguisticamente expresso por uma frase reflexa, é sempre verdadeiro. A maioria, se não a totalidade, dos predicadores reflexos representa relações de identidade ou de igualdade: cf. *o João é igual a si próprio*, *a Maria é tão alta como ela própria*, *o Pedro tem a mesma idade que ele próprio*. Por serem tautologias, estas frases são geralmente sentidas como estranhas (e pouco informativas) pela maioria dos falantes.

A maioria dos predicadores de dois lugares (*lavar*, *gostar (de)*, *amigo (de)*, etc.) são **não reflexos**; ou seja, como se observou acima, com eles as proposições que correspondem ao esquema $a \times a$ são apenas contingentemente verdadeiras. Finalmente, existem vários predicadores que, por razões lógicas, são **irreflexos**; ou seja,

¹⁶ De novo, se o predicador é um verbo transitivo, este é necessariamente “conjugado” com o pronome “reflexo” *se* e o adjunto de reciprocidade é opcional, como em (38a). Em construções comparativas, em vez de uma preposição, ocorre uma conjunção-complementador (cf. os exemplos dois parágrafos a seguir, no texto).

com esses predicadores, as proposições representadas pelo esquema $a x a$ nunca podem ser verdadeiras: cf., por exemplo, *(ser) pai (de)*, *(ser) mais alto (do que)*, etc.

Tal como a construção recíproca torna um predicador não simétrico num predicador simétrico (cf. 8.8.1 e os exemplos em (35)), também a construção reflexa, acima descrita, torna um predicador não reflexo num predicador reflexo (cf. (38)).

Em português existem verbos inerentemente reflexos, como o verbo *suicidar-se* (cf. *ele suicidou-se* e a impossibilidade de **ele suicidou o Luís*)¹⁷. Certos prefixos, como *auto-*, também convertem um predicador não reflexo num predicador reflexo, exigindo sempre o pronome clítico *se*: cf. *ele autorrecreminou-se*. Outros verbos são convertidos em predicadores “reflexos” quando têm determinado tipo de complementos. É o caso de *abrir* quando o seu complemento é uma expressão que denota uma parte do corpo, como em *abrir a mão*: numa frase como *o João abriu a mão*, a mão em questão é necessariamente a da entidade referida pelo sujeito (neste caso, o João).

8.8.3 Relações e predicadores transitivos

Os **predicadores transitivos**¹⁸ são aqueles que selecionam (pelo menos) dois argumentos e permitem a seguinte inferência: se as frases que correspondem ao esquema proposicional $a x b$ e $b x c$ são ambas verdadeiras, então a frase que corresponde ao esquema proposicional $a x c$ também é verdadeira. Muitos dos predicadores que denotam relações espaciais, como *conter*, *(estar) no interior (de)*, *sobrevoar*, são transitivos. Por exemplo, se as frases (a) e (b) dos exemplos abaixo forem verdadeiras, então as frases (c) também o serão¹⁹:

- (39) a. O balão sobrevoa o helicóptero.
 b. O helicóptero sobrevoa o barco.
 c. O balão sobrevoa o barco.
- (40) a. As joias estão no interior da caixa.
 b. A caixa está no interior do cofre.
 c. As joias estão no interior do cofre.

A maioria dos predicadores são **não transitivos**, como *beijar*, *colidir*, *empurrar*, *gostar (de)*, *odiar*, *zangar-se (com)*, etc. Assim, se as frases (a) e (b) dos seguintes exemplos forem verdadeiras, nada implica que as frases (c) também o sejam, embora possam sê-lo, em determinadas situações:

- (41) a. O Carlos gosta da Ana.
 b. A Ana gosta do Pedro.
 c. O Carlos gosta do Pedro.

¹⁷ Paradoxalmente, e mau grado a presença do pronome *se*, as frases com *suicidar-se* não são casos da construção reflexa: cf. a impossibilidade de **ele suicidou-se a si próprio*. Este verbo combina-se “intrinsecamente” com *se*, tal como *afundar-se*, *ajoelhar-se*, *queixar-se*, etc.

¹⁸ É crucial referir aqui que a noção de transitividade discutida nesta subsecção é de natureza lógico-semântica, nada tendo a ver com a noção de transitividade sintática caracterizada nos Caps. 11 e 28.

¹⁹ Para além disso, é necessário que a referência dos argumentos seja a mesma nas três frases, e que as situações descritas se integrem no mesmo “macroevento”. Por exemplo, se o helicóptero mencionado em (39a) for diferente do que é mencionado em (39b), ou se (39a) e (39b) representarem eventos distintos espacial e temporalmente, é evidente que a inferência transitiva não se obtém.

- (42) a. O carro colidiu com o jipe.
b. O jipe colidiu com o muro.
c. O carro colidiu com o muro.

Um **predicador intransitivo**, por sua vez, é aquele que nunca suporta qualquer inferência transitiva; por outras palavras, uma frase que corresponda ao esquema da inferência será sempre contraditória com as frases obedecendo ao esquema das premissas. Um exemplo é o predicador (*ser*) *pai* (*de*): se forem verdadeiras as frases *o Pedro é pai do Luís* e *o Luís é pai do Paulo*, então é necessariamente falsa a frase *o Pedro é pai do Paulo*. No domínio espacial, o predicador (*estar*) (*diretamente*) *ao lado* (*de*) é igualmente intransitivo: assim, se a Maria estiver (*diretamente*) ao lado da Alexandra e a Alexandra estiver (*diretamente*) ao lado da Clara, a Maria não pode estar (*diretamente*) ao lado da Clara.

Contrariamente às relações simétrica e reflexa, não existe qualquer construção gramatical que torne um predicador não transitivo num predicador transitivo.

9	PROCESSOS DE LEXICALIZAÇÃO	215
9.1	Noções gerais sobre lexicalização	217
9.2	Valores semânticos, classes lexicais ou gramaticais e valores funcionais das unidades multilexicais	221
9.3	Identificação das unidades multilexicais e aferição dos seus graus de lexicalização	223
9.3.1	Propriedades formais das unidades multilexicais	225
9.3.1.1	Plano paradigmático	225
9.3.1.1.1	Substituição lexical	226
9.3.1.1.2	Variação flexional	230
9.3.1.2	Plano sintagmático	232
9.3.2	Propriedades semânticas das unidades multilexicais	238
9.3.2.1	Uso das unidades multilexicais em línguas especializadas	241
9.3.3	Aspetos quantitativos das unidades multilexicais	244



9 PROCESSOS DE LEXICALIZAÇÃO

Existem, na língua, sequências de palavras com comportamentos unitários ou tendencialmente unitários, isto é, semelhantes aos de uma palavra única, resultantes de conexões formais e semânticas que se foram estabelecendo entre os seus elementos e que o uso consagrou (cf. *dizer à boca cheia, de forma geral, à sua saúde*, etc.). Essas sequências **lexicalizadas** são aqui chamadas **unidades multilexicais** e ao processo da sua formação chamamos **lexicalização**¹. A lexicalização é, pois, neste capítulo, definida como um processo gradual de fixação de sequências de palavras em grupos formal e semanticamente coesos, com um comportamento semelhante ao de uma unidade do léxico.

Na formação destes grupos, a coesão das palavras entre si pode tornar-se maior ou menor, dependendo, entre outros fatores, do momento mais ou menos recente em que se formaram e do seu grau de fixação no uso. Entre os grupos que apresentam maior coesão, temos itens como *lugar-comum* = 'banalidade'; *de se lhe tirar o chapéu* = 'excelente'; *esticar o pernil* = 'morrer'; *por um triz* = 'quase'; *ainda que* = 'embora'; *de acordo com* = 'conforme'; noutros grupos, o grau de coesão é menor, embora também se vão fixando no uso, por serem muito recorrentes: cf. *mania das grandezas, abandono escolar, absolutamente indispensável, passar fome, ficar a ganhar, fora de série, por mero acaso*. Todas estas sequências serão aqui globalmente consideradas unidades multilexicais².

¹ O termo "lexicalização" tem sido usado com vários sentidos, por vezes muito abrangentes, como seja o da criação, em geral, de novos "lexemas" (cf. Lyons 1977), ou o da criação de palavras lexicais ou gramaticais, simples ou complexas, resultantes de mudanças morfológicas, sintáticas ou semânticas, ocorridas sobre palavras já existentes. Neste último sentido, recobre processos morfológicos produtivos, como a composição e a derivação, e menos produtivos, como a conversão, a redução, a abreviação ou a amálgama. Outras vezes, o termo é usado num sentido muito restrito, como, p.e., o da especialização semântica de uma palavra (cf. Downing 1977). Também há autores que chamam lexicalização ao processo de "desgramaticalização", que se traduz na passagem de uma palavra gramatical a palavra lexical (plena), ou seja, o processo oposto ao da gramaticalização (cf. Cap. 10 sobre processos de gramaticalização).

² Consideramos que estes grupos formam unidades multilexicais, independentemente da sua ortografia. Alguns constam de vocabulários e de dicionários com o estatuto de entradas autónomas e são grafados com hífen. Mas é extremamente hesitante esta grafia, quer nos dicionários, quer na escrita corrente; assim, não trataremos de modo diferente os grupos que alguns dicionários registam com entrada própria, ortografados com hífen (consagrando-os como unidades no léxico) e aqueles que não são habitualmente registados dessa maneira, sendo antes tratados no interior dos artigos dos dicionários, sob a entrada de um dos seus componentes.

Os compostos sintáticos (cf. Cap. 58) como *desportos radicais*, *condomínio fechado* ou *cadeira de rodas* podem ser vistos como expressões sintáticas lexicalizadas, isto é, como unidades multilexicais. O mesmo acontece com as siglas, como *PS*, e os acrónimos, como *ONU*, que também têm na base da sua formação sequências lexicais de natureza sintagmática (respetivamente *Partido Socialista* e *Organização das Nações Unidas* (cf. Caixa [2]).

O processo de lexicalização das unidades multilexicais é um processo gradual, não regular e, em caso de forte grau de fixidez e coesão semântica dos seus elementos, o resultado final tem de ser definido individualmente no dicionário, pois a expressão, no seu sentido atual, não é (ou já não é) analisável semanticamente. Por exemplo, o sentido da expressão *meter os pés pelas mãos* não se deduz a partir do sentido dos elementos que a formam, sendo necessário defini-lo no dicionário como 'confundir-se, contradizer-se'. Quando a interpretação de uma expressão é obtida a partir do significado dos elementos que a formam, diz-se que o seu significado é **composicional** (p.e., o significado da sequência *ver um cão* é obtido a partir do significado das suas palavras constitutivas *ver, um e cão*). Quando isso não acontece, como sucede em *meter os pés pelas mãos*, diz-se que o significado da expressão é **não composicional**.

Existem também muitas sequências com um menor grau de fixidez e de coesão interna que são incluídas, sem definição, nos dicionários que registam grupos de palavras usuais. Por exemplo, na entrada "consumo", estes dicionários podem incluir sequências como *bens de consumo*, *sociedade de consumo*, *impróprio para consumo*, consagradas pelo uso, e que, embora mantendo o significado literal de cada elemento e tendo uma interpretação composicional, adquiriram também, de acordo com a perspetiva aqui seguida, o estatuto de unidades multilexicais.

À lexicalização estão associados processos de fixação sintática, semântica e pragmática que originam uma mudança linguística; diz-se que há **rotinização** quando as expressões lexicalizadas que decorrem desses processos se tornam familiares e rotineiras; um processo relacionado e mais avançado é o da **institucionalização**, que ocorre quando são aceites por toda ou por uma parte significativa da comunidade linguística. Passam, então, a fazer parte do inventário lexical da língua e, como tal, são adquiridas e memorizadas pelos falantes (nativos ou não) e inscritas nos dicionários.

A lexicalização, no sentido que aqui vamos desenvolver, processa-se ao longo do tempo. Existem sempre, em qualquer período da língua, frases ou grupos ambíguos, que podem ser usados quer como "sequências livres", i.e., interpretadas composicionalmente e contendo elementos sintática e semanticamente independentes, quer (em contextos diferentes) de uma forma na qual se observa um processo de lexicalização, em que os seus elementos perderam parcial ou totalmente a independência sintática e semântica, adquirindo, em maior ou menor grau, um caráter unitário. É o caso, por exemplo, da expressão *fazer a cama*, que pode ser usada com a significação literal 'fabricar uma determinada peça de mobiliário' ou então, enquanto unidade multilexical, com o sentido de 'arranjar as roupas da cama' ou com o sentido de 'armar uma cilada' (cf. a discussão desta unidade multilexical em 9.1).

Este capítulo articula-se do seguinte modo: em 9.1 discutem-se noções gerais sobre lexicalização; em 9.2 caracterizam-se as unidades multilexicais quanto às áreas

semânticas (e pragmáticas) em que ocorrem e quanto à sua classe e função gramatical; em 9.3 classificam-se as unidades multilexicais de acordo com o seu grau de lexicalização e discutem-se as suas propriedades formais, semânticas e pragmáticas mais salientes.

9.1 Noções gerais sobre lexicalização

A lexicalização cria unidades multilexicais cuja estrutura apresenta um maior ou menor grau de coesão interna, falando-se, concomitantemente, de “lexicalização (mais ou menos) forte ou fraca”. Revelam um forte grau de lexicalização as unidades multilexicais formadas por compostos morfossintáticos, como *surdo-mudo*, *andar-modelo*, *peixe-espada*, por grupos sintagmáticos³, como *pés-de-galinha* ou *andar à nora*, por siglas, como *TV*, por acrónimos, como *sida*, por fórmulas ritualizadas de saudação, como *passou bem?*, e de delicadeza, como *se faz favor*, e ainda por vários tipos de aforismos⁴, como *mais vale prevenir do que remediar*.

A par destes exemplos com alto grau de unidade formal e semântica, observam-se também, muito frequentemente, outras coocorrências sistemáticas de palavras, com menos coesão interna, que mantêm o sentido literal de alguns ou mesmo de todos os seus elementos e que formam uma unidade de uso (cf. *fogo posto*, *estado lastimoso*, *politicamente correto* ou *induzir em erro*). Estes grupos (considerados aqui também como unidades multilexicais) são chamados **colocações**, **combinatórias** ou **coocorrentes privilegiados**, termos que designam, assim, a tendência particular de certas palavras para ocorrerem em combinação com determinadas outras, em certos contextos⁵.

[1] As unidades multilexicais têm recebido designações genéricas muito diversas e por vezes conflituosas, como sejam, entre outras, “frases feitas, expressões/frases estereotipadas/cristalizadas”, “unidades multilexicais”, “expressões pluriverbais”, “locuções”, “lexias complexas”, “frasemas”, “pragmatemas”, “clichés”, “idiomatismos”, etc. Esta proliferação terminológica e o uso de termos diferentes com o mesmo sentido em nada têm contribuído para clarificar as noções relacionadas com este tema.

O que têm em comum estes diversos tipos de sequências de palavras é o facto de, em todas elas, se terem estabelecido progressivamente relações combinatórias (mais ou menos) fortes entre os seus elementos, relações essas que se foram consolidando formal, semântica e pragmaticamente, em maior ou menor grau, dando origem a grupos (mais ou menos) fixos, mas muito recorrentes no uso, qualquer que seja o seu grau de fixidez. Todos eles são grupos de palavras “pré-construídos”, de que o falante dispõe e que utiliza automaticamente e de forma inconsciente, mesmo quando são analisáveis nos seus elementos constitutivos.

³ Alguns destes grupos são também chamados “compostos sintáticos” no Cap. 58.

⁴ Os aforismos englobam sentenças de origem culta e máximas de cunho popular; adotamos aqui a definição de “aforismo” dada por Prado Coelho (1969:26), como «expressão concisa de um pensamento moral».

⁵ O termo inglês “*collocation*”, que em francês corresponde a “*combinatoire*”, foi definido pela primeira vez por Firth (1957) como «*actual words in habitual company*», «palavras que coocorrem habitualmente» (t.n.).

A consolidação formal da sequência implica, por um lado, que existe pouca (nalguns casos, nenhuma) liberdade de escolha das palavras que a constituem e, por outro lado, que a sua estrutura morfossintática se vai tornando fixa. Dizemos que há **des-sintatização** quando os falantes deixam de ter consciência de uma construção sintática subjacente à sequência e esta passa a funcionar como uma palavra simples: cf. *cara de pau* ou *fim de semana*.

A consolidação semântica consiste na perda do significado composicional dos elementos da unidade multilexical, através de um processo de **semantização** de todo o grupo. Quer isto dizer que o significado da sequência deixa de poder deduzir-se da combinação dos significados de cada uma das palavras que a compõem; a sequência adquire então um significado global, cristalizado e unitário, muitas vezes de carácter metafórico e por vezes completamente opaco. Quando se atinge o grau máximo de fixidez formal e de consolidação semântica, tem-se uma só unidade de forma e de sentido constituída por várias palavras, um processo chamado **idiotização**. Neste caso, os falantes já não encontram nos elementos da expressão qualquer motivação para o seu sentido global. Por exemplo, em *estar com a pulga atrás da orelha* os falantes não veem hoje motivação para o sentido global de 'desconfiar' que a expressão tem.

A consolidação pragmática de uma unidade multilexical, por sua vez, consiste numa habituação (ou **rotinização**) do uso combinado das palavras que a formam. Note-se que, neste caso, as palavras não perdem necessariamente o seu significado individual, podendo manter uma certa "liberdade" semântica. A rotinização é favorecida pela frequente utilização das unidades multilexicais, quer em usos comuns, do dia-a-dia (cf. *cabeça no ar*, *consciência tranquila*, *apinhado de gente*, *banhado em lágrimas*, *esfregar as mãos de contente*, *contrair um empréstimo*, *estar condenado ao fracasso*), quer em usos específicos, de natureza técnica, científica, etc. (cf. *acidente vascular cerebral*, *artrite reumatoide*, *massa molecular*, *ângulo de reflexão*, *consolidação orçamental*, *mercado de capitais*, *carta rogatória*, *ação cível*). Existem também situações de comunicação, sejam elas mais ou menos formais, que favorecem a rotinização das expressões (cf. as expressões de saudação como *bom dia! como está?* ou de início ou fim de discurso oral como *ora viva! até logo!* ou de discurso escrito, como *caro amigo, com os meus melhores cumprimentos*). Em todos estes casos, os falantes memorizam e armazenam as combinatórias como um todo, no seu léxico mental, e dessa forma as utilizam.

Contrariamente às combinações livres de palavras, as unidades multilexicais, enquanto expressões lexicalizadas, para além de não obedecerem à livre escolha dos falantes no que respeita aos elementos que as constituem e à sua estruturação, adquirem, por vezes, um semantismo próprio, institucionalizado pelo uso.

Nomes como *lugar* e alguns dos adjetivos com que se podem combinar ilustram bem estas observações. Por exemplo, a propósito de um lugar onde as pessoas se sentem bem, pode escolher-se, livremente, dizer *que lugar simpático*, *que sítio aprazível*, *que local ameno* ou *agradável* ou qualquer outra combinação entre os nomes *lugar*, *sítio* e *local* e os adjetivos *simpático*, *aprazível*, *ameno* ou *agradável*, em qualquer das duas ordenações nome + adjetivo ou adjetivo + nome. A escolha do nome e do adjetivo, bem como a sua ordem, não implicam uma alteração substancial do significado do sintagma nominal. Estes casos são, pois, exemplos de sequências livres.

Em contrapartida, se se quiser falar de um texto cheio de banalidades, pode dizer-se: *quantos lugares comuns!* Mas, para este mesmo sentido, não se poderá optar por dizer *quantos lugares usuais, quantos sítios vulgares, quantos locais banais* ou qualquer outra combinação dos nomes *lugar, sítio* e *local* com os adjetivos *usual, vulgar* ou *banal*. Do mesmo modo, não é possível inverter a ordem entre o nome *lugares* e o adjetivo *comuns*: cf. a estranheza de *#quantos comuns lugares!*⁶. Ou seja, a sequência *lugar comum* passou a ser uma unidade multilexical, i.e., um sintagma fortemente lexicalizado, que não admite variação na escolha nem na ordem dos seus elementos constitutivos, tendo adquirido o sentido figurado de 'banalidade'; como tal, é produzido e reconhecido pelos falantes e atualmente tem entrada própria, sendo grafado com hífen, na maior parte dos dicionários.

Quer se trate da formação de palavras compostas, como *guarda-chuva* ou *corri-mão*, quer da fixação de unidades gramaticais, como *em vez de* ou *por conseguinte*, e lexicais, como *mercado negro, perder a cabeça* ou *dar com os burrinhos na água*, está-se perante um processo de lexicalização, que pode, na sua origem, ter sido determinado pela necessidade de criação de novas designações, mas que é sempre o resultado da combinação frequente de elementos originalmente autónomos. Estes elementos vão perdendo gradualmente a sua autonomia formal, fixando-se, por vezes, em formas totalmente cristalizadas (*coca-bichinhos* e não *coca-bichos*; *tem-te não caias* e não *tenha-se não caia*); e vai-se tornando também fixa a sua ordem de ocorrência no grupo (*saia-casaco* e não *casaco-saia*; *de viva voz* e não *de voz viva*). Concomitantemente, os elementos vão perdendo a sua autonomia semântica dentro da unidade multilexical, que passa a ter um significado global, o qual tem de ser adquirido e memorizado pelos falantes da mesma forma que o significado das palavras simples; i.e., o falante tem de aprender que, p.e., *perder a cabeça* significa 'descontrolar-se' e *dar com os burrinhos na água* significa 'falhar'. Mesmo nos casos em que os elementos da sequência lexicalizada não perdem o seu significado autónomo, e, logo, o sentido do grupo não é opaco, o seu uso global não se confunde com o uso de uma sequência idêntica não lexicalizada (cf. a discussão a seguir sobre *fazer a cama*).

Sendo a lexicalização o resultado de um processo gradual, é possível observar, tanto diacronicamente como sincronicamente, expressões que atingiram diferentes graus deste processo. Assim, por exemplo, no momento atual, a sequência de palavras *fazer a cama* é ambígua: pode ocorrer como uma combinação livre em que cada uma das palavras mantém o seu sentido individual (cf. (1a)) ou apresentar características que a identificam como uma unidade multilexical, apresentando, contudo, diferentes graus de lexicalização: (i) uma combinatória em que ainda se pode deduzir o sentido global da unidade a partir do sentido dos seus elementos (cf. (1b)), ou (ii) uma combinatória que atingiu o grau máximo de lexicalização, tendo adquirido um sentido não composicional totalmente idiossincrático (cf. (1c)). A mesma expressão pode, ainda, fazer parte de um aforismo (cf. (1d)) que tem também, tal como (1c), um significado global, não composicional:

⁶ Neste exemplo, e noutros deste texto, o símbolo "#" é usado para indicar que a unidade multilexical não é possível na interpretação que se discute.

- (1) a. O marceneiro *fez-lhe a cama* e a cómoda. (= fabricou-lhe a cama)
 b. A mãe *fez-lhe a cama* logo de manhã. (= arranjou-lhe as roupas da cama)
 c. O colega *fez-lhe a cama* e ele foi despedido. (= armou-lhe uma cilada)
 d. *Quem boa cama fizer nela se deitará.* (= terá um bom futuro quem o souber preparar)

Em (1a), as palavras *fazer* e *cama* mantêm o seu significado individual de 'produzir' e de 'peça de mobiliário para as pessoas se deitarem', entrando as duas em relações paradigmáticas com outras palavras da mesma área semântica, como *fabricar* e *construir* (para *fazer*) e *cómoda*, *cadeira* ou *estante* (para *cama*). Estas palavras podem ser substituídas na frase, preservando a sua interpretabilidade e alterando-a apenas em partes bem delimitadas do seu conteúdo (cf. *o marceneiro fabricou-lhe/construiu-lhe a cómoda/estante*). Em (1b-d), no entanto, essas substituições são impossíveis, uma vez que a expressão *fazer a cama* (*a alguém*) funciona globalmente, ou seja, como uma "grande" palavra, com um significado bem diferente do significado composicional que a expressão correspondente tem em (1a).

Como fica patente nos exemplos (1b,c), as expressões lexicalizadas, à semelhança das palavras individuais, podem ter vários sentidos. Em (1b), o sentido da expressão, embora menos transparente do que o de (1a), não é completamente opaco, pois o verbo *fazer* não ficou totalmente vazio de sentido (sendo usado aqui na aceção de 'arranjar, arrumar') e *cama* está usado metonimicamente como 'roupa da cama'. Em contrapartida, em (1c), a expressão *fazer a cama*, usada metaforicamente com o sentido de 'armar uma cilada', é semanticamente opaca, não sendo interpretada de nenhum modo a partir dos seus elementos constitutivos.

Esta diferença semântica reflete-se também na maneira como a expressão interage com determinados fenómenos sintáticos. Assim, *fazer a cama*, em (1b), está mais próxima da sequência livre de (1a) porque admite, tal como esta⁷, variação flexional (cf. (2a)) e sintática (cf. (2b-e)):

- (2) a. As mães *faziam as camas* das crianças logo de manhã. [variação flexional do verbo e dos elementos do SN]
 b. *A cama*, a mãe vai *fazer-lha* logo de manhã. [construção de topicalização]
 c. *Que cama* é que a mãe *fez*? [construção interrogativa]
 d. *A cama* que a mãe *fez* ficou pronta logo de manhã. [construção relativa]
 e. *A cama foi-lhe feita* pela mãe logo de manhã. [construção passiva]

Em (1c), *fazer a cama* admite variação na flexão verbal (cf. (3a)), mas não variação sintática (cf. (3b-d)):

- (3) a. Os colegas *fizeram-lhe a cama* e ele foi despedido. [variação na flexão verbal: pessoa, tempo e modo]
 b. #*A cama*, o colega *fez-lha* e ele foi despedido. [construção de topicalização]
 c. #*Que cama* é que o colega *lhe fez*? [construção interrogativa]
 d. #*A cama foi-lhe feita* pelo colega. [construção passiva]

⁷ As variações de *fazer a cama* como sequência livre, correspondentes às de (2), são, respetivamente, *os marceneiros fizeram-lhes as camas*; *a cama*, *o marceneiro fez-lha* (por bom preço); *que cama é que o marceneiro lhe fez?*; *a cama que o marceneiro lhe fez ficou linda* e *a cama foi-lhe feita pelo marceneiro*.

Do cotejo destes exemplos se conclui que é em (1c) que a sequência atinge um maior grau de lexicalização.

Por sua vez, o exemplo (1d) forma, no seu todo, um enunciado completo, de sentido unitário e idiossincrático, com alto grau de generalidade. Trata-se de «uma unidade de comunicação mínima» (Zuluaga 1980:192, t.n.) e aqui o elemento *fazer a cama* pode ser interpretado como 'preparar o futuro'.

Em síntese, existe lexicalização quando as palavras de um sintagma ou de uma frase, tipicamente, perdem a sua independência semântica, passando a formar um grupo mais ou menos "fixo" ao qual estão associados contextos de uso e/ou significados específicos. Consequentemente, existem limitações fortes na escolha e na combinação das palavras que formam estes grupos; as palavras tendem a "escolher-se" umas às outras de maneira restrita, ou seja, manifestam aquilo que se chama **cosseleção**. Esta situação contrasta com o que se passa na formação de sintagmas e frases "normais" ou "livres", em que as palavras são livremente escolhidas pelos falantes e mantêm o seu significado próprio⁸.

Assim, de acordo com Sinclair (1991), quando se constroem sintagmas e frases "livres", faz-se um uso quase ilimitado das possibilidades de estruturação sintática das palavras e estas são selecionadas de acordo com um "princípio de livre escolha". Em contrapartida, quando se constroem unidades multilexicais, essas possibilidades estão, de um modo ou de outro, limitadas, dando lugar à utilização automática de combinatórias previamente construídas, e a sua seleção deve-se a um "princípio idiomático"⁹.

9.2 Valores semânticos, classes lexicais ou gramaticais e valores funcionais das unidades multilexicais

As unidades multilexicais distribuem-se por várias áreas semânticas e pragmáticas:

- (i) expressões com que se designam realidades extralinguísticas do quotidiano (*pão com manteiga, café com leite, vinho branco, ferro elétrico, serviço de chá, trem de cozinha*);
- (ii) expressões que respondem à necessidade de criar denominações especializadas, científicas, técnicas, artísticas, etc. (*água forte, água-pesada, órgão consultivo, sequenciação do genoma humano, deduzir acusação, natureza-morta*);
- (iii) variantes "expressivas" de palavras ou sintagmas com um significado expressivo mais neutro (*dias a fio por dias seguidos; fino como um coral por muito fino ('esperto'); tratar alguém abaixo de cão por tratar alguém muito mal*);
- (iv) expressões de natureza geral ou específicas de determinadas áreas (*cultura geral, fluxo migratório, drogas ilícitas, sinais exteriores de riqueza, fora do comum, do foro privado, matar à fome, estar em foco*);

⁸ Existem também, obviamente, restrições de natureza sintática, morfológica e semântica na construção dos sintagmas e das frases.

⁹ Estes dois princípios não se excluem mutuamente, podendo, até, coexistir na mesma expressão, como, p.e., em certas unidades multilexicais com um menor grau de lexicalização.

- (v) expressões com uma função exclusivamente pragmática: saudações e fórmulas de cortesia (*bom dia* ou *muito prazer em conhecer*), formas de início ou fecho de discurso (*era uma vez* ou *sem outro assunto*), ditos (*quem sai aos seus não degenera*), máximas (*não deixes para amanhã o que podes fazer hoje*);
- (vi) provérbios (*grão a grão enche a galinha o papo, cão que ladra não morde, dezembro frio, calor no estio*);
- (vii) siglas (RTP – Rádio Televisão Portuguesa, CTT – Correios, Telégrafos e Telefones, EDP – Eletricidade de Portugal);
- (viii) acrónimos (*sida* – síndrome de imunodeficiência adquirida ou AMI – Ajuda Médica Internacional).

[2] A abreviação de unidades sintagmáticas faz parte do processo de lexicalização e inclui a **siglação** (formando **siglas**), a **acronímia** (formando **acrónimos**) e, ainda, o uso de **abreviaturas**, na escrita. Na siglação, combinam-se as iniciais de palavras plenas de um sintagma (cf. AR < Assembleia da República, IRS < imposto sobre o rendimento de pessoas singulares) ou as iniciais dos elementos de palavras complexas (cf. TV < televisão ou ECG < eletrocardiograma). Uma característica importante das siglas (que as distingue dos acrónimos) é o facto de a sua pronúncia ser feita letra a letra. A formação dos acrónimos, por sua vez, é semelhante à das siglas, com a diferença de que os acrónimos têm estrutura silábica, isto é, são pronunciados como uma palavra da língua (cf. OPA < oferta pública de aquisição, IVA < imposto sobre o valor acrescentado, ETAR < estação de tratamento de águas residuais). Finalmente, as abreviaturas consistem em reduções de palavras no registo escrito da língua (cf. a.C. < antes de Cristo, V. Exa. < Vossa Excelência, s.d. < sem data, Sr. Dr. < Senhor Doutor).

As unidades multilexicais pertencem a várias classes, lexicais ou gramaticais¹⁰, dependendo do seu núcleo (cf. Cap. 11), e têm funções variadas na frase ou dentro de um sintagma nominal: podem ser (i) nominais¹¹ (cf. *chamavam-lhe caixa de óculos, só uso sal marinho, é preciso dar-lhes apoio logístico, avaliámos ganhos e perdas*), funcionando como argumentos; (ii) adjetivais (cf. *o rapaz está são como um pero*), funcionando como modificadores nominais restritivos ou como predicativo do sujeito; (iii) preposicionais, com a função de adjunto adverbial (cf. *a dada altura perguntaram-me por ti, gosto de tripas à moda do Porto, por alturas de Coimbra, o carro avariou*); (iv) verbais, funcionando como predicado (cf. *só ontem me dei conta de que tinhas razão, mandaste fazer essas calças, celebraram um contrato de compra e venda de propriedades, eles fazem vista grossa aos erros dos amigos*); (v) orações completivas infinitivas, funcionando como modificador de nome ou como

¹⁰ Designam-se por "unidades multilexicais gramaticais" as locuções gramaticais de vários tipos: conjuncionais, como *ao passo que* e *ainda que*; relativas e/ou interrogativas, como *por que*, *o qual*, etc. Autores como Benson et al. (1986a e 1986b) e Hausmann (1989) só consideram como expressões multilexicais as que são constituídas por itens lexicais (plenos), excluindo, assim, as locuções gramaticais. Neste capítulo não consideramos as locuções gramaticais, que são tratadas nos capítulos dedicados às várias classes a que pertencem.

¹¹ Alguns autores incluem nestas categorias nomes próprios geográficos (*Mar Vermelho, Golfo Pérsico*), de períodos históricos (*Idade Média, Revolução Francesa*), de personalidades históricas (*Rei Sol, Dama de Ferro*), de instituições (*Cruz Vermelha, Legião Estrangeira*).

predicativo do sujeito (cf. *estou com um vestido de trazer por casa, o teu procedimento foi de se lhe tirar o chapéu*).

Em certos casos, a lexicalização reclassifica uma expressão que pertence originariamente a uma classe (lexical ou sintagmática) noutra classe lexical: cf. *as pessoas sem abrigo* (sintagma preposicional atributivo, com função adjetival) → *os sem-abrigo* (nome); noutros casos, a lexicalização envolve fenómenos de gramaticalização das palavras que constituem a unidade multilexical (i.e., a passagem da expressão ou de um dos seus elementos de uma classe lexical para uma classe gramatical. Para vários exemplos desta reclassificação, cf. Cap. 10).

9.3 Identificação das unidades multilexicais e aferição dos seus graus de lexicalização

Firth, no artigo em que escreveu a sua célebre frase «*You shall know a word by the company it keeps*», «Conhecerás uma palavra pela sua companhia» (t.n.) (Firth 1957), chamava a atenção para a importância de se estudarem e caracterizarem as palavras tendo em atenção as outras palavras com que podem coocorrer, designando por “*collocations*” (“colocações”, em português) as possibilidades combinatórias de cada palavra. Marcava, assim, o início do desenvolvimento do conceito de “combinatória de palavras” ou de “unidade multilexical”, o qual tem sido encarado segundo perspetivas diversas centradas em duas propriedades fundamentais:

- (i) a fixidez sintática da sequência de palavras que forma o grupo;
- (ii) a não composicionalidade semântica dessa sequência.

Ambas estas propriedades admitem uma gradação, que decorre da maior ou menor tendência associativa das palavras que formam o grupo¹².

De acordo com a forma como estas propriedades são perspetivadas, assim se têm dado definições mais restritas ou mais latas de unidade multilexical (cf. Bacelar do Nascimento et al. 2006); para alguns autores, estas unidades consistem em associações de palavras com características restritas cujo resultado é semanticamente não composicional (cf. Hausmann 1979, Mel’cuk 1984), enquanto para outros elas abrangem tipos de relações associativas menos restritivos (cf. Sinclair 1991).

Segundo as definições mais restritivas, a unidade multilexical é uma sequência de palavras totalmente coesa e indissolúvel, sem possibilidade de variação das formas que a constituem, e não admite a introdução de outros elementos, em particular no seu interior. Semanticamente, o seu significado é único e não é igual ao conjunto dos significados dos seus componentes (ou seja, não é composicional); antes, é fruto de um processo acabado de conversão de uma estrutura sintática numa unidade de valor lexical ou gramatical, ou numa unidade fraseológica. Pode dizer-se que em unidades deste tipo o processo de lexicalização atingiu o seu termo.

¹² A determinação desta tendência associativa exige o recurso a métodos estatísticos que permitam obter dados de frequência recolhidos com base em *corpora* equilibrados de grandes dimensões. Cf., entre outros, Church e Hanks (1990), Clear (1993), Smadja (1993), Krishnamurthy (1997).

Contudo, não é possível nem estabelecer com exatidão o momento em que se inicia o processo de lexicalização de uma sequência de palavras nem o momento em que está terminado. Por isso, parece-nos demasiado restritivo classificar as unidades multilexicais tomando como critério absoluto a finalização do processo de lexicalização. Seguimos, por isso, uma perspectiva menos restritiva do que a descrita no parágrafo anterior, estabelecendo vários tipos de unidades multilexicais segundo o grau de lexicalização que detêm.

A identificação e a aferição do grau de lexicalização das unidades multilexicais são feitas com base nas suas propriedades formais, semânticas e pragmáticas; o reconhecimento da consolidação pragmática envolve o estabelecimento de dados quantitativos: índices de frequência e estatísticos, incluindo medidas de associação lexical, aqui representadas pelo chamado Índice Combinatório (IC)¹³.

Estas propriedades são mais valorizadas ou menos valorizadas conforme a abordagem seguida, sendo as propriedades formais e semânticas as mais utilizadas, tradicionalmente¹⁴. Dada a abrangência do conceito de unidade multilexical que seguimos nesta Gramática, consideram-se aqui também as propriedades pragmáticas.

Para além disso, consideramos que cada um dos tipos de propriedades, formal, semântico e pragmático, se manifesta de modo independente dos outros; por outras palavras, cada um deles contribui de modo independente para a identificação de uma unidade multilexical e para a determinação do seu grau de lexicalização. Assim, p.e., uma unidade multilexical pode atingir um forte grau de lexicalização do ponto de vista semântico ou pragmático e um fraco grau de lexicalização do ponto de vista formal. Naturalmente, o alto grau de lexicalização de uma unidade multilexical fica tanto mais reforçado quanto mais propriedades se lhe puderem reconhecer.

Para exemplificar este ponto, comparem-se duas unidades multilexicais: *pau mandado* e *fontes fidedignas*. A primeira é fortemente lexicalizada nas três dimensões consideradas: tem uma forte coesão formal, é semanticamente opaca e tem um forte Índice Combinatório (IC = 10)¹⁵. Em contrapartida, *fontes fidedignas* é uma unidade multilexical formada por coocorrentes privilegiados (cf. 9.1), mas não apresenta todas as propriedades identificadoras das unidades multilexicais (cf. 9.3.1): formalmente, aceita a comutação de um dos seus elementos (cf. *fontes seguras*) e a inclusão de uma palavra no seu interior (cf. *fontes bastante fidedignas*), e semanticamente é transparente. Neste caso, a propriedade que permite o seu reconhecimento como unidade multilexical é a forte tendência associativa entre

¹³ O IC (também chamado "Mutual Information") é um índice que resulta da aplicação de uma medida sobre um *corpus* visando aferir a tendência associativa entre duas ou mais palavras (cf. 9.3.3).

¹⁴ Cf., por exemplo, Carter (1987) e Zuluaga (1980).

¹⁵ Recorde-se que o IC é a medida que identifica a tendência associativa de duas ou mais palavras. O valor 10 é o IC desta sequência num *corpus* de 50 milhões de palavras, *corpus* do projeto COMBINA-PT – *Combinatórias Lexicais do Português* do Centro de Linguística da Universidade de Lisboa (<http://www.clul.ul.pt/en/research-teams/183-reference-corpus-of-contemporary-portuguese-crpc>) (cf. Antunes et al. 2008), também referido como *corpus* COMBINA, extraído do *Corpus de Referência do Português Contemporâneo*, projeto do mesmo Centro. Em todos os casos em que é indicado, neste capítulo, o IC de uma unidade multilexical, ele foi extraído deste *corpus*. Considera-se que a partir do índice 7 se torna segura a identificação de combinatórias no plano estatístico. Em estudos sobre este tema (cf. Bacelar do Nascimento 2002, Evert e Krenn 2001, Pereira e Mendes 2002), foi evidenciado que os valores de *Mutual Information* entre 7 e 11 correspondem a uma faixa que tem a maior concentração de combinatórias num dado *corpus*.

os seus elementos, confirmada pelo seu elevado Índice Combinatório (= 12,4), o qual evidencia uma estabilidade de ocorrência conjunta das duas palavras que a formam. Assim, contrariamente a *pau mandado*, o grau de lexicalização de *fontes fidedignas* é praticamente só de natureza pragmática, não envolvendo fatores formais ou semânticos. Trata-se de uma “colocação”, de acordo com a terminologia de Firth, ou seja, uma coocorrência típica entre duas ou mais palavras que, com o uso, se convencionalizou.

9.3.1 Propriedades formais das unidades multilexicais

Descrevem-se nesta subsecção algumas propriedades formais que caracterizam as unidades multilexicais. Estas propriedades manifestam-se no plano paradigmático e no plano sintagmático. As primeiras têm a ver com a possibilidade (ou não) de substituir cada um dos elementos de uma unidade multilexical por outros elementos com um significado semelhante ou próximo, mantendo constante a estrutura da expressão. A variação das unidades multilexicais nesta dimensão é tanto mais importante quanto nos servirá de base para estabelecer a sua tipologia formal básica, articulada em três grupos com propriedades distintas: (i) unidades multilexicais fixas, (ii) unidades multilexicais semifixas e (iii) coocorrentes privilegiados. Esta matéria é abordada em 9.3.1.1. As propriedades sintagmáticas, por sua vez, dizem respeito à maneira como os vários elementos que formam as unidades multilexicais se combinam linearmente e hierarquicamente entre si. Este aspeto é abordado em 9.3.1.2.

9.3.1.1 Plano paradigmático

No plano paradigmático, um critério para avaliar a fixação de uma unidade multilexical consiste em observar a maior ou menor possibilidade de substituição (ou comutação) de cada um dos seus elementos lexicais sem alteração da estrutura interna da unidade multilexical ou do seu significado global. Ou seja, as substituições restringem-se àquelas palavras que detêm com o elemento substituído relações de sentido próximas: sinónimos, hiperónimos, hipónimos e, por vezes, antónimos (cf. Cap. 8)¹⁶.

O número de substituições lexicais que é possível fazer, quer num lugar fixo, quer em diferentes posições dentro da unidade multilexical, apresenta uma relação inversa com o grau de lexicalização da sequência. Por outras palavras, quanto menor for o número possível de substituições maior será o grau de lexicalização da unidade multilexical; inversamente, quanto menos restringida for a seleção paradigmática, mais “livre” é a sequência formada¹⁷. É importante ter em conta, no entanto, que mesmo nas unidades multilexicais com menor grau de lexicalização, o número dos elementos lexicais que podem comutar é tipicamente limitado, formando paradigmas mais ou menos fechados.

¹⁶ A substituição de um elemento de uma unidade multilexical por um antónimo não altera necessariamente o seu sentido global. Veja-se, p.e., *ver com bons olhos* vs. *ver com maus olhos*, em que ambos exprimem a noção geral de ter uma determinada opinião sobre alguma coisa.

¹⁷ Mesmo no “discurso livre” não envolvendo unidades multilexicais, nunca existe liberdade total nas escolhas, porque existem sempre restrições de natureza sintática, semântica e pragmática nas escolhas lexicais que levam à formação dos sintagmas e das frases.

É também no plano paradigmático que se tem em conta a possibilidade ou impossibilidade de variação em número, género, tempo, modo e aspeto dos elementos que compõem a unidade multilexical. Abordamos a substituição lexical e a substituição flexional separadamente.

9.3.1.1.1 Substituição lexical

Para ilustrar este ponto, compare-se a sequência livre *vimos uma estatueta que tinha corpo de gesso e cara de pau* com a unidade multilexical *contrataram uma secretária que tinha cara de pau, mas era muito eficiente*. Enquanto, no primeiro caso, *cara de pau* pode ser substituído por *rosto de pau* ou *cara de madeira* (cf. *vimos uma estatueta que tinha corpo de gesso e {rosto de pau/cara de madeira}*), no segundo tal substituição é impossível (cf. *#contrataram uma secretária que tinha {rosto de pau/cara de madeira}, mas era muito eficiente*). Ou seja, *cara de pau*, neste exemplo, é uma unidade multilexical com um alto grau de fixidez lexical.

Um aspeto muito importante das unidades fixas como *cara de pau*, e que é fundamental para a compreensão do seu funcionamento, é o seguinte: o seu grau de lexicalização (ou seja, a sua fixidez lexical) é tão grande no léxico mental dos falantes, que praticamente já não ocorrem como sequências livres, precisamente porque são automaticamente identificadas como unidades multilexicais. Isto explica que o primeiro exemplo dado acima (*vimos uma estatueta...*) e outros como ele sejam extremamente raros e pouco naturais quando ocorrem (para muitos falantes, a primeira interpretação que vem à ideia é que a estatueta tem “cara de pau” no sentido idiomático).

❖ Unidades multilexicais fixas

Existem unidades multilexicais que não aceitam qualquer substituição dos seus elementos, chamadas, por isso, **unidades multilexicais fixas**. Apresentam-se a seguir alguns exemplos, organizados por classes lexicais e/ou funcionais:

- (4) Grupos nominais
 - a. pau mandado (vs. #vara mandada)
 - b. um belo dia¹⁸ (vs. #um bonito dia)
 - c. golpe de misericórdia (vs. #golpe de compaixão/#golpe de piedade/#golpe misericordioso/#pancada de misericórdia)
 - d. falsa modéstia (vs. #fingida modéstia/#simulada modéstia)
 - e. árvore das patacas (vs. #árvore das moedas)
- (5) Grupos com a função de modificador restritivo de nome ou de predicativo do sujeito (passamos a designar esta função doravante como “função adjetival”)
 - a. de se lhe tirar o chapéu (vs. #de se lhe tirar o boné/#de se lhe pôr o chapéu)
 - b. são como um pero (vs. #são como uma maçã/#são como um fruto)
 - c. de trás da orelha/#de trás do ouvido)
 - d. de força (vs. #de fortaleza/#de fraqueza)
 - e. não contente (vs. #não feliz)

¹⁸ Entende-se aqui *um belo dia* com o sentido de adjunto adverbial de tempo, como em *um belo dia, chegaram os livros que tínhamos encomendado*.

- (6) Grupos com a função de adjunto adverbial
- de pedra e cal (vs. #de rocha e cal)
 - com a pulga atrás da orelha (vs. #com a mosca atrás da orelha/#com o inseto atrás da orelha)
 - a fio (vs. #a linha/#a cordel)
 - a prazo (vs. #a data)

- (7) Grupos verbais
- não se dar por achado (vs. #não se dar por encontrado)
 - meter água (vs. #pôr água/#introduzir água)
 - matar à fome (vs. #assassinar à fome)
 - não ser flor que se cheire (vs. #não ser rosa que se cheire)
 - ficar a ver navios (vs. #ficar a ver barcos)
 - não arredar pé (vs. #não desviar pé/#não afastar pé)

Também não aceitam substituição certas unidades multilexicais que incluem elementos com sufixos avaliativos; em particular, o elemento sufixado não pode ocorrer sem o sufixo:

- (8) a. coca-bichinhos (vs. #coca-bichos)
 b. de casa e pucarinho (vs. #de casa e púcaro)
 c. nervoso miudinho (vs. #nervoso miúdo)
 d. sair de fininho (vs. #sair de fino)

❖ Unidades multilexicais semifixas

Existem unidades multilexicais que aceitam a substituição de alguns elementos da sequência por outras palavras que mantêm com o elemento substituído relações semânticas de sinonímia, de hiperonímia, de hiponímia ou mesmo de antonímia, sem que a estrutura e o sentido genérico do grupo se alterem. Apresentam-se a seguir alguns exemplos:

- (9) Grupos nominais
- ódio mortal/ódio de morte/ódio figadal
 - fome de cão/fome canina/fome de lobo
 - varinha mágica/varinha de condão
 - onda de assaltos/maré de assaltos/vaga de assaltos
 - desenhos animados/bonecos animados
 - habilitações literárias/habilitações acadêmicas
 - altas pressões/baixas pressões
 - consciência pesada/consciência tranquila
 - velho continente/novo continente
- (10) Grupos com função adjetival
- sobejamente conhecido/por de mais conhecido
 - de alto nível/de alto gabarito/de alto coturno
 - fora do comum/fora do vulgar/fora do normal
 - politicamente correto/politicamente incorreto

- (11) Grupos verbais
- armar aos cágados/armar aos cucos/armar ao pingarelho
 - ficar de cara à banda/ficar de cara ao lado
 - fechar a porta na cara/dar com a porta na cara/bater com a porta na cara
 - dizer à boca cheia/dizer à boca pequena
 - dar com o nariz na porta/bater com o nariz na porta
 - ver com bons olhos/ver com maus olhos
 - dar à sola/dar às de vila-diogo
- (12) Grupos com função de adjunto adverbial
- de forma geral/de modo geral
 - a folhas tantas/a páginas tantas
 - por via materna/pelo lado materno
 - em menos de um amém/em menos de um fósforo
 - na mó de cima/na mó de baixo
 - à compita/à porfia
 - a meio gás/a todo o gás
 - com fins lucrativos/sem fins lucrativos
 - de bom grado/de mau grado

Em certas unidades multilexicais, observa-se também variação entre elementos simples e elementos com morfemas avaliativos:

- (13) a. um *fio* de voz/um *fiozinho* de voz
 b. *carro* de linhas/*carrinho* de linhas
 c. *novo* em folha/*novinho* em folha
 d. ter a *papa* feita/ter a *papinha* feita
 e. de *alto* nível/de *altíssimo* nível
 f. de *primeira* água/de *primeiríssima* água
 g. de *corpo* bem feito/de *corpinho* bem feito
 h. com um *grão* na asa/com um *grãozinho* na asa
 i. dar com os *burros* na água/dar com os *burrinhos* na água

Nos exemplos (9)-(13), a variação lexical não é aleatória. De facto, não há total liberdade na escolha dos elementos que aceitam comutação: os paradigmas admissíveis (ou seja, as classes de elementos que podem comutar entre si) são fechados. Por esta razão, consideram-se estas sequências não como expressões livres mas como **unidades multilexicais semifixas**. Os exemplos seguintes ilustram algumas restrições que se verificam na seleção lexical destas unidades:

- (14) a. fome de cão/fome canina/fome de lobo (vs. #fome lupina/#fome de raposa)
 b. varinha de condão/varinha mágica (vs. #varinha de magia)
 c. por de mais conhecido/sobejamente conhecido (vs. #demasiadamente conhecido/#demasiado conhecido)
 d. fora do comum/fora do vulgar/fora do normal (vs. #fora do geral)
 e. fechar a porta na cara/bater com a porta na cara/dar com a porta na cara (vs. #atirar com a porta na cara)

As expressões aforísticas¹⁹ (muitas vezes consideradas como unidades altamente rígidas) também admitem alguma variação paradigmática, mas de modo mais restrito que os casos anteriores, como se ilustra nos seguintes exemplos (destacam-se com sublinhado as palavras registadas como variantes por substituição):

- (15) a. Depois do temporal vem a bonança./Depois da tempestade vem a bonança./Depois da tormenta vem a bonança.
 b. Mestre fora, dia santo na loja./Patrão fora, dia santo na loja.
 c. Longe da vista, longe do coração./Longe dos olhos, longe do coração.
 d. Quem não trabuca não manduca./Quem não trabalha, não come.
 e. {Brigam/Guerreiam/Pelejam/Ralham/Zangam-se} as comadres, descobrem-se as verdades.

Diz-se que duas ou mais seqüências são **variantes** de uma unidade multilexical semifixa quando (i) não têm significados globais diferentes, (ii) não dependem do contexto em que ocorrem e (iii) fazem parte de uma série paradigmática limitada e estável (cf. Corpas 1996:28s).

Este tipo de variação não deve confundir-se com “jogos de palavras” intencionais, a partir de uma unidade multilexical fixa, que por vezes ocorrem no discurso. Assim, por exemplo, no *corpus* COMBINA, ocorrem as seguintes substituições ocasionais na expressão aforística *no poupar é que está o ganho*:

- (16) a. No *anunciar* é que está o ganho.
 b. No *atacar* é que está o ganho.
 c. No *descontar* é que está o ganho.
 d. No *esperar* é que está o ganho.
 e. No *provar* é que está o ganho.
 f. No *comparar* é que está o ganho.
 g. No *economizar* é que está o ganho.
 h. No *prejuízo* é que está o ganho.

[3] Nos casos acima descritos, não se trata de variantes, mas sim de modificações criativas. De acordo com Corpas (*ibidem*), quanto maior é o grau de fixidez de uma unidade, mais possibilidade há de que sofra modificações no discurso e de que essas modificações e o seu efeito sejam reconhecidos pelos falantes. Ou seja, o facto de a unidade multilexical *no poupar é que está o ganho* permitir um tão diversificado número de alterações significa que ela atingiu, efetivamente, um alto grau de lexicalização. É precisamente a sua memorização global que facilita a produção pelos falantes dos “jogos de palavras” em que se mantém a estrutura da unidade multilexical e se conserva a identidade parcial dos seus componentes, mas se altera o semantismo da unidade multilexical. Isto nada tem a ver com os casos de variação da unidade multilexical por substituição do léxico.

❖ Coocorrentes privilegiados

Em (17)-(19), a mesma palavra de base ocorre com modificadores diferentes (sintagmas preposicionais atributivos em (17), sintagmas preposicionais “referen-

¹⁹ Todos os exemplos de expressões aforísticas com variação dados neste capítulo foram extraídos do levantamento apresentado em Chacoto (1994).

ciais" em (18)), ou o mesmo modificador, adjetivo, ocorre com nomes diferentes, em (19b)):

- (17) a. garrafa de água, garrafa de vinho, garrafa de cerveja, garrafa de aguardente, garrafa de whisky, garrafa de leite, garrafa de azeite
 b. pão de mistura, pão de milho, pão de trigo, pão de centeio, pão-de-leite, pão de forma
 c. pão com manteiga, pão com queijo, pão com fiambre
 d. fora de dúvida, fora de causa, fora de propósito, fora de série, fora de moda, fora de serviço
- (18) contrato de adjudicação, contrato de aluguer, contrato de arrendamento, contrato de compra e venda, contrato de promessa de compra e venda, contrato de concessão de serviços, contrato de empreitada
- (19) a. pão branco, pão saloio, pão caseiro, pão quente, pão fresco
 b. porto fluvial, praia fluvial, transporte fluvial, via fluvial

Nestes exemplos, os diferentes modificadores alteram significativamente o sentido do grupo; ou seja, estas sequências não são variantes de unidades multilexicais semifixas, mas cada uma delas é uma unidade multilexical.

Estes grupos são constituídos por combinações de palavras com uma alta frequência de coocorrência, superior ao que seria de prever tendo em conta as suas frequências individuais. Esse facto justifica que sejam considerados unidades multilexicais, chamadas colocações, combinatórias ou coocorrentes privilegiados. Os elementos que compõem os coocorrentes privilegiados mantêm o seu sentido próprio e a sua interpretação semântica é plenamente composicional; por outras palavras, os coocorrentes privilegiados têm um fraco grau de lexicalização.

9.3.1.1.2 Variação flexional

Para além da substituição lexical, contribui também para avaliar o grau de fixidez das unidades multilexicais a possibilidade de flexão morfológica dos seus elementos adjetivais, verbais ou nominais. Neste aspeto, a generalidade das unidades multilexicais comporta-se da mesma forma que as sequências livres de palavras. Assim, muitas delas aceitam a pluralização de nomes e adjetivos de acordo com as regras morfológicas da língua:

- (20) a. ódio mortal/ódios mortais
 b. bom dia!/bons dias!
 c. motivo de força maior/motivos de força maior
 d. arco do triunfo/arcos do triunfo
 e. árvore das patacas/árvores das patacas

O mesmo acontece em provérbios, onde se registam variantes com formas no singular e no plural:

- (21) a. Burro velho não aprende *língua*./Burro velho não aprende *línguas*.
 b. A cavalo dado não se olha *o dente*./A cavalo dado não se olha *os dentes*.

Da mesma forma, na flexão verbal ocorre variação em pessoa/número, para além de variação em tempo e modo (cf. (22a,b)). Repare-se também na possibilidade de ocorrência de um verbo auxiliar aspetual, no terceiro exemplo de (22b) e no segundo de (22c)²⁰:

- (22) a. *Perdi a cabeça./Perdemos a cabeça.*
 b. *Meteste água./Meteram água./Continuam a meter água.*
 c. *Não tem papas na língua./Não têm tido papas na língua.*

Este tipo de variação é menos frequente em expressões de tipo aforístico, embora se encontre nalgumas:

- (23) a. *Maior é o dia, maior a romaria./Quanto maior for o dia, maior a romaria./Se maior fosse o dia, maior era a romaria.*
 b. *Quem parte paga./Quem partiu pagou.*

Em contrapartida, algumas unidades multilexicais cristalizam numa determinada forma flexiva, ou seja, possuem um maior grau de fixidez no que se refere à flexão nominal ou verbal. Veja-se, por exemplo: *ganhar terreno* (unidade multilexical) mas não *ganhar terrenos* (sequência livre), *detrás da orelha* (unidade multilexical) mas não *detrás das orelhas* (sequência livre), *braço direito* (unidade multilexical) mas não *braços direitos* (sequência livre), *quando o rei faz anos* (unidade multilexical) mas não *quando o rei fez anos* (sequência livre), *cão que ladra não morde* (unidade multilexical) mas não *cães que ladraram não morderam* (sequência livre). Os exemplos de (24) ilustram também fixação flexional (em (24a), no plural):

- (24) a. mundos e fundos, primeiros socorros, altas esferas, forças armadas, (ter) macaquinhos no sótão, por artes mágicas, direitos humanos, desenhos animados
 b. caixa alta, ordem do dia, sangue frio

É interessante notar que a fixidez lexical das unidades multilexicais fixas (cf. (4)-(8)) é relativamente independente da rigidez flexional. Assim, p.e., há muitas unidades multilexicais fixas que não admitem substituição dos seus elementos constitutivos mas que admitem variação flexiva: cf. *paus mandados*, *caras de pau*, *golpes de misericórdia*, *falsas modéstias*, *de se lhes tirar o chapéu*, *sãos como um pero*, *não se deram por achados* ou *andam a meter água*.

Para finalizar esta subsecção, mencionamos um tipo de variação morfológica que tem um importante correlato semântico, mais precisamente de natureza referencial. Quando um dos elementos de uma unidade multilexical é um pronome, este pode variar de referência, e, logo, concomitantemente, em pessoa e número. Neste caso, diz-se que a posição que corresponde ao pronome é de “preenchimento livre”:

- (25) a. à {minha/tua/sua/nossa/...} vontade
 b. à {minha/tua/sua/nossa/vossa} espera
 c. sair-{me/te/lhe/nos/...} um peso de cima dos ombros

²⁰ Note-se que em (22) o nome não pode variar flexionalmente: cf. #*perdemos as cabeças*, #*meteste águas* ou #*não ter papa na língua*).

- d. fazer-{me/te/lhe/...} o ninho atrás da orelha
- e. só {me/te/lhe/lhes/nos/vos} faltava mais essa!
- f. que {te/lhe/lhes/vos} faça bom proveito!
- g. não {me/te/lhe/lhes/nos/vos} faltava mais nada!
- h. não {me/te/se/nos/...} dar por achado

Nestes exemplos, pode ocorrer, em vez do pronome, um sintagma preposicional introduzido por *de* quando o pronome é possessivo ou dativo de posse, e por *a* quando o pronome é um dativo ético, como se ilustra a seguir:

- (26) a. à vontade da Maria
 b. à espera do João
 c. sair um peso de cima dos ombros do rapaz
 d. fazer o ninho atrás da orelha do pai
 e. só faltava mais essa ao Pedro!
 f. não faltava mais nada ao Pedro!
 g. que faça bom proveito a todos!

Há outros casos em que a substituição por um sintagma preposicional não é possível: p.e., não é possível a correspondência entre *estar/ficar fora* {*de mim/de ti/de si*} e #*estar/ficar fora de alguém*.

9.3.1.2 Plano sintagmático

No plano sintagmático, a fixação (ou “estabilidade”) sintática das unidades multilexicais tem a ver, essencialmente, com os seguintes pontos: (i) possibilidade de alteração da ordem dos constituintes, (ii) possibilidade de inserção de elementos modificadores ou especificadores (artigos, quantificadores, advérbios, adjetivos, pronomes), (iii) possibilidade de redução ou extensão da expressão e (iv) possibilidade de modificação da estrutura sintática (p.e., em estruturas de nominalização, passivas, interrogativas, relativas, etc.).

❖ Ordem dos constituintes

Nas unidades multilexicais fortemente lexicalizadas, os constituintes ocupam, em geral, uma posição fixa. Quando essa ordem é alterada, deixam de funcionar como unidades multilexicais, o que se ilustra nos seguintes exemplos (primeiro apresenta-se a unidade multilexical e a seguir a sequência resultante da alteração da ordem):

- (27) a. uso e abuso (vs. #abuso e uso)
 b. altas esferas (vs. #esferas altas)
 c. ter lágrimas nos olhos (vs. #ter nos olhos lágrimas)
 d. trazer água no bico (vs. #trazer no bico água/#no bico trazer água)
 e. ter um grãozinho na asa (vs. #ter na asa um grãozinho/#na asa ter um grãozinho)
 f. tirar nabos da púcara (vs. #tirar da púcara nabos/#da púcara tirar nabos)
 g. não saber da missa a metade (vs. #não saber a metade da missa/#da missa não saber a metade)

Também nos coocorrentes privilegiados a alteração da ordem pode levar à perda do estatuto de unidade multilexical, como se ilustra a seguir:

- (28) a. congresso extraordinário (vs. #extraordinário congresso²¹)
 b. prós e contras (vs. #contras e prós)
 c. claro e inequívoco (vs. #inequívoco e claro)
 d. certo e sabido (vs. #sabido e certo)

Em contrapartida, noutras unidades multilexicais com uma lexicalização mais fraca, a ordem de palavras é livre e o estatuto da sequência como unidade multilexical mantém-se, independentemente da ordem, como se ilustra nos seguintes exemplos, retirados do *corpus* COMBINA, em que as duas versões têm um alto Índice Combinatório:

- (29) a. aumento considerável/considerável aumento
 b. silêncio absoluto/absoluto silêncio
 c. conhecimento prévio/prévio conhecimento
 d. consequências graves/graves consequências
 e. consequências inevitáveis/inevitáveis consequências
 f. atual situação/situação atual

Também se verifica alteração da ordem de palavras em variantes de aforismos:

- (30) a. Cedo deitar e cedo erguer, dá saúde e faz crescer./Deitar cedo e cedo erguer dá saúde e faz crescer.
 b. Deus dá nozes a quem não tem dentes./Dá Deus nozes a quem não tem dentes.
 c. Mais vale cair em graça do que ser engraçado./Vale mais cair em graça do que ser engraçado.
 d. A palavras loucas orelhas moucas./Orelhas moucas a palavras loucas.
 e. Pela boca morre o peixe./O peixe morre pela boca.
 f. Alcança quem não cansa./Quem não cansa alcança.
 g. O segredo é a alma do negócio./A alma do negócio é o segredo.

❖ Inserção de elementos novos na unidade multilexical

Este é outro dos critérios que permitem avaliar o grau de lexicalização de uma unidade multilexical. Quando a unidade multilexical é fortemente lexicalizada, a sequência é totalmente coesa, não permitindo a inserção de qualquer novo elemento, em particular no seu interior; se isto acontece, a unidade multilexical torna-se numa sequência livre (em (31), damos primeiro exemplos da unidade multilexical e depois, entre parênteses, exemplos das sequências livres resultantes da inserção de novos elementos):

- (31) a. sangue frio (vs. sangue muito/extremamente frio)
 b. arco do triunfo (vs. arco do grande triunfo)
 c. penas do inferno (vs. penas do terrível inferno)
 d. desenhos animados (vs. desenhos muito animados)

²¹ Neste caso, a alteração semântica deve-se ao facto de o adjetivo em anteposição ao nome adquirir sentido avaliativo (cf. Cap. 31).

- e. fora de moda (vs. fora daquela moda)
- f. linguagem natural (vs. linguagem pouco natural)
- g. fazer prova (vs. fazer uma prova)

Em contrapartida, alguns coocorrentes privilegiados admitem a inserção de novos elementos (em geral, artigos, quantificadores, advérbios, adjetivos) que não fazem propriamente parte da unidade multilexical:

- (32) a. votado ao abandono/votado a um completo abandono
 b. solução consensual/solução totalmente consensual
 c. autores consagrados/autores altamente consagrados
 d. perder a cabeça/perder totalmente a cabeça
 e. não passar da cepa torta/não passar nunca da cepa torta
 f. conhecer os cantos à casa/conhecer bem os cantos à casa

O mesmo acontece em certos aforismos, como em *quem conta um conto aumenta um ponto* vs. *quem conta um conto aumenta sempre um ponto*.

❖ Redução e extensão da unidade multilexical

Certas unidades multilexicais têm uma variante “curta” e uma variante “longa”, sendo relativamente fácil, na maior parte dos casos, identificar qual delas é a básica. Quando a variante básica é a mais curta, diz-se que há **extensão** (cf. *querer à força* vs. *querer à viva força*); quando a variante básica é a mais longa, diz-se que há **redução** (cf. *garrafa de vinho do Porto* vs. *garrafa de Porto*).

Contrariamente aos casos de “inserção” analisados acima, nas unidades multilexicais formadas por extensão, o elemento inserido (geralmente um especificador com características semânticas de intensificador) é intuído pelos falantes como pertencendo à unidade multilexical, e não como um simples “acrescento”, como em (32). Isto é, trata-se, claramente, de variantes curtas e longas, qualquer delas com o estatuto de unidade multilexical, no seu todo. Outra diferença relativamente aos casos de inserção é que, tanto nos casos de extensão como nos de redução, as duas variantes são igualmente reconhecidas pelos falantes como usuais. Dão-se alguns exemplos em (33) e (34) (este último paradigma contendo aforismos):

- (33) a. com o propósito de/com o firme propósito de
 b. estar arranjado/estar bem arranjado
 c. deitar achas para a fogueira/deitar mais achas para a fogueira
 d. doença contagiosa/doença altamente contagiosa
 e. o pão de cada dia/o pão nosso de cada dia
 f. levar uma corrida/levar uma corrida em osso/levar uma corrida em pelo
 g. vai dar uma volta!/vai dar uma volta ao bilhar grande!
 h. ir para a rua/ir para o olho da rua
- (34) a. Gaba-te cesta que vais à vindima./Gaba-te cesta rota que amanhã vais à vindima.
 b. Gostos não se discutem./Gostos e cores não se discutem.
 c. Galinha pedrês não a comas nem a dês./Galinha pedrês nem a comas nem a vendas nem a dês.

- d. Mulher honrada não tem ouvidos./Mulher honrada não tem ouvidos nem olhos.
- e. Abril águas mil./Abril águas mil coadas {por um funil/por um mandil}.

Nalguns casos, a variante da unidade multilexical pode ficar reduzida a um único elemento²²:

- (35) a. espinha dorsal/espinha
- b. ensino secundário/secundário (cf. *a Joana anda no ensino secundário/a Joana anda no secundário*)

A extensão e a redução ocorrem apenas nalgumas unidades multilexicais, estando longe de constituir fenómenos produtivos, como se pode verificar nos seguintes exemplos, em que se comparam unidades multilexicais da mesma área semântica e com significados relacionados, mas em que só uma delas admite a redução/extensão:

- (36) a. Vai dar uma volta!/Vai dar uma volta ao bilhar grande! (vs. *vai passear!/#vai passear {ao jardim/à praia/ao campo}!*)
- b. garrafa de vinho tinto/garrafa de tinto (vs. *garrafa de água mineral/*garrafa de mineral*)
- c. ensino secundário/secundário (cf. *a Joana está no ensino secundário/a Joana está no secundário*) (vs. *ensino universitário/#universitário*²³ (cf. *a Joana está no ensino universitário/*a Joana está no universitário*))

❖ Modificação da estrutura sintática das unidades multilexicais

À semelhança do que acontece com as sequências livres, algumas unidades multilexicais podem ocorrer em construções sintáticas diferenciadas (nominalizações, estruturas passivas, relativas, interrogativas, etc.), mas este fenómeno não é de forma alguma generalizável a todas elas. Descrevem-se a seguir alguns destes casos²⁴.

❖ Nominalizações

Algumas unidades multilexicais de natureza verbal admitem uma versão nominalizada, com um núcleo nominal obtido por derivação deverbal a partir do verbo correspondente, como se ilustra nos seguintes exemplos:

- (37) a. abusar da confiança/abuso de confiança
- b. usar e abusar/uso e abuso
- c. renovar o contrato/renovação do contrato
- d. suspender o contrato/suspensão do contrato
- e. abrir as hostilidades/abertura das hostilidades
- f. abrir a sessão/abertura da sessão

²² Nestes casos, temos uma palavra simples, fruto da redução de uma unidade multilexical da qual é sinónima.

²³ Neste caso, dir-se-á *está na universidade*.

²⁴ Na secção 9.1 já tínhamos chamado a atenção para a importância destas propriedades, dando em (1)-(3) exemplos com a unidade multilexical *fazer a cama a alguém*.

Já as seguintes unidades multilexicais não admitem nominalização:

- (38) a. brincar com o fogo (vs. #brincadeira com o fogo)
 b. andar à nora (vs. #andamento à nora)
 c. custar os olhos da cara (vs. #custo dos olhos da cara)

Não sendo possível generalizar, podemos, no entanto, dizer que a composicionalidade da unidade multilexical parece favorecer a existência de nominalizações. De facto, em (37), tanto o verbo como o nome mantêm os significados que têm fora das unidades multilexicais. Pelo contrário, os sintagmas verbais de (38) são todos eles não composicionais, e o seu sentido global não encontra correspondência nas nominalizações.

❖ Modificadores adjetivais e preposicionais

Algumas unidades multilexicais nominais cujo núcleo é modificado por um sintagma preposicional admitem uma alternativa adjetival para o modificador, como se ilustra a seguir:

- (39) a. fome de cão/fome canina
 b. ódio de morte/ódio mortal
 c. método de contração/método contraceutivo

Outras, pelo contrário, não admitem esta alternativa:

- (40) a. esfera de ação (vs. #esfera ativa)
 b. golpe de misericórdia (vs. #golpe misericordioso)
 c. ordem do dia (vs. #ordem diária)
 d. cara de fome (vs. #cara esfomeada)
 e. fogo de artifício (vs. #fogo artificial)

❖ Pronominalização

A pronominalização de um sintagma nominal dentro de uma unidade multilexical é possível quando esta é formada por coocorrentes privilegiados e mantém um sentido composicional, como se ilustra em (41) (põem-se as unidades multilexicais entre parênteses retos, e em itálico o elemento pronominalizado e o pronome):

- (41) a. Ele ia [celebrar *um contrato*], mas acabou por não [o celebrar].
 b. Ele tentou [contrair *um empréstimo*], mas acabou por não conseguir [contrai-lo].
 c. Ele está sempre a [dar *conselhos*]; na realidade, [dá-os] a torto e a direito.
 d. Adoras [gastar *dinheiro*]! É fácil [gastá-lo], não é?
 e. Tens que [abafar *esse escândalo*]. É preciso mesmo [abafá-lo] já!

Em contrapartida, a pronominalização é impossível em unidades multilexicais fixas, altamente lexicalizadas e sem sentido composicional, como em *à noite todos os gatos são pardos* vs. #*pardos*, [*à noite todos os gatos o são*]²⁵ (cf. (42)), o que faz com

²⁵ Neste exemplo, o elemento que não pode ser pronominalizado é um adjetivo predicativo (cf. *ele é inteligente, sempre o foi*).

que este fenômeno seja um bom critério para distinguir, nestas sequências, aquelas que têm um alto grau de lexicalização:

- (42) a. [Deste com *os burrinhos* na água], #mas oh se [deste com *eles* na água]!
 b. Não [abras *as hostilidades*], #porque quando [*as* abrires], vai ser o inferno!
 c. Isso vai-te [custar *os olhos da cara*], #vai-te mesmo [custá-los]!
 d. Ela [disse *cobras e lagartos*] de ti, #e [disse-os] à pior pessoa!

A pronominalização envolve um empobrecimento lexical, e é provavelmente por esse motivo que ocorre menos facilmente nas unidades multilexicais que dependem, para o seu processamento, do material lexical que as constitui.

❖ Passivização

As unidades multilexicais verbais com menor grau de lexicalização podem ser passivizadas; em contrapartida, as unidades multilexicais verbais altamente lexicalizadas não admitem a passivização. Este contraste é ilustrado pela unidade multilexical ambígua *fazer a cama* (*a alguém*). Como se observou em 9.1, esta unidade multilexical pode ter como significado 'arranjar as roupas da cama (a alguém)', leitura na qual tem um fraco grau de lexicalização, ou 'armar uma cilada (a alguém)', leitura na qual tem um alto grau de lexicalização e um sentido não composicional. Na primeira leitura pode ser passivizada, como se ilustra em (43a,b), mas na segunda não, como se ilustra em (43c,d). Outros exemplos de casos que não admitem passivização são dados em (43e,f) e (43g,h):

- (43) a. A mãe fez-lhe a cama logo de manhã.
 b. A cama foi-lhe feita pela mãe logo de manhã.
 c. Os colegas fizeram-lhe a cama e ele foi despedido.
 d. #A cama foi-lhe feita pelos colegas e ele foi despedido.
 e. O rapaz não tomou chá em pequeno.
 f. #Chá não foi tomado pelo rapaz em pequeno.
 g. Dá Deus nozes a quem não tem dentes.
 h. #Nozes são dadas por Deus a quem não tem dentes.

Outros fatores que contribuem para se avaliar o grau de lexicalização de uma unidade multilexical são:

❖ Ausência de determinante

Quando a unidade multilexical contém um sintagma nominal, a ausência de determinante no SN revela um alto grau de lexicalização, sendo outro bom critério para distinguir as sequências mais lexicalizadas (cf. (44)) das menos lexicalizadas (como *fazer a barba* ou *dar um passeio*):

- (44) a. meter água (vs. #meter a água)
 b. não ser flor que se cheire (vs. #não ser uma flor que se cheire)
 c. ficar a ver navios (vs. #ficar a ver os navios)
 d. não arredar pé (vs. #não arredar o pé)

❖ Subcategorização

Algumas unidades multilexicais de natureza verbal, em geral fortemente lexicalizadas, não admitem um dos complementos que o verbo seleciona numa construção livre:

- (45) a. ajudar (#alguém) à missa
 b. dar (#alguma coisa) às de vila-diogo

❖ Construções com tópico e estruturas interrogativas e relativas

A impossibilidade de os elementos nominais de uma unidade multilexical ocorrerem em posições que não sejam aquelas que correspondem ao seu estatuto argumental normal constitui prova do seu alto grau de fixidez; assim, um sintagma nominal contido numa unidade multilexical não pode ser um tópico marcado nem ocorrer sob a forma de um constituinte interrogativo ou de um pronome relativo, três elementos que tipicamente ocorrem na posição inicial de uma oração (cf. Caps. 12, 48 e 39, respetivamente). Assim, correspondendo a (46a), não se tem (46b-d):

- (46) a. Os colegas fizeram-lhe *a cama* e ele foi despedido.
 b. #*A cama*, os colegas fizeram-lha e ele foi despedido. [tópico]
 c. #*Que cama* é que os colegas lhe fizeram? [constituinte interrogativo]
 d. #*A cama que* os colegas lhe fizeram acabou com ele. [constituinte relativo]

9.3.2 Propriedades semânticas das unidades multilexicais

As propriedades semânticas permitem também classificar as unidades multilexicais em três tipos, numa escala de maior ou menor especialização e motivação semântica: (i) unidades multilexicais (semanticamente) transparentes, (ii) unidades multilexicais (semanticamente) opacas e (iii) unidades multilexicais (semanticamente) semitransparentes. O conceito de “transparência semântica” (cf., adiante, Caixa 3) é usado aqui como indicando a existência de algum grau de motivação ou fundamentação semântica, acessível ao falante, que lhe permite “descodificar” o sentido da unidade multilexical; inversamente, a “opacidade semântica” remete para uma situação em que o falante não encontra qualquer fundamentação semântica para descodificar o sentido de uma unidade.

- (i) As unidades multilexicais transparentes têm um significado literal e composicional (ou seja, obtido a partir do significado dos seus elementos constitutivos), o que as aproxima das sequências livres. Estes grupos são considerados unidades multilexicais porque ocorrem frequentemente e constituem conjuntos estáveis, sendo muito comuns em linguagens de especialidade (cf. *acidente vascular cerebral*, *apneia obstrutiva do sono*, *consolidação orçamental*), embora ocorram também na linguagem comum, como mostram os exemplos de (47):

- (47) consequências inevitáveis, em plena consciência, vaga de incêndios, são e salvo, feito inédito, sigilo absoluto, direitos humanos, entrada livre, opinião pública, por artes mágicas, dívida externa, acima de qualquer suspeita, da pior espécie

(ii) No extremo oposto, as unidades multilexicais semanticamente opacas são as que têm uma “especialização semântica” máxima. Neste caso, os falantes não têm atualmente acesso a qualquer motivação ou fundamentação para “descodificar” o sentido global que adquiriram e que não se deduz a partir dos seus componentes (cf. *dar com os burrinhos na água*, *dar às de vila-diogo*, *ir para o maneta*)²⁶. Consequentemente, esse sentido tem de ser aprendido, tal como se aprende o significado de uma palavra simples. Note-se também que nalgumas destas unidades multilexicais ocorrem elementos que, atualmente, fora delas, não têm qualquer significado, como, por exemplo, os elementos sublinhados das expressões *por um triz*, *fazer jus*, *nem chus nem bus*, *por artes de berliques e berloques*, e do provérbio *lé com lé, cré com cré*. Apresentam-se em (48) outros exemplos de unidades multilexicais semanticamente opacas:

- (48) a. alma do padeiro = buracos do pão
 b. menina dos olhos = pessoa preferida, pessoa em quem mais orgulho se tem
 c. dar às de vila-diogo = fugir apressadamente
 d. dar com os burrinhos na água = falhar
 e. tirar nabos da púcara = tentar saber alguma coisa de forma dissimulada
 f. fazer o ninho atrás da orelha = enganar
 g. estar com a pulga atrás da orelha = estar desconfiado
 h. não passar da cepa torta = não progredir
 i. estar com um grãozinho na asa = estar um pouco alcoolizado
 j. encanar a perna à rã = protelar, atrasar o desenvolvimento de alguma coisa
 k. ter macaquinhos no sótão = ter manias/desconfianças, cismar em alguma coisa
 l. fazer jus (a alguma coisa) = merecer (alguma coisa)
 m. por um triz = por pouco
 n. por uma unha negra = por pouco
 o. tomar chá em pequeno = ser bem-educado

(iii) As unidades multilexicais semitransparentes são aquelas em que alguns elementos mantêm o seu significado literal, enquanto outros adquiriram sentidos figurados, essencialmente por processos metafóricos ou metonímicos (cf. Cap. 8). Estas extensões do sentido, no entanto, são (ainda) familiares para os falantes, o que empresta a este tipo de unidades multilexicais um certo grau de motivação semântica. São exemplos, entre outros, *baixar a cabeça*, *custar os olhos da cara* (formados por metáfora), *ser um bom garfo* (formado por metonímia).

Algumas unidades deste tipo são formadas com os “verbos leves” (cf. Cap. 28) *dar*, *fazer*, *ter* e *tomar* (cf. *dar razão*, *fazer anos*, *ter cabeça/tino*, *tomar banho*, *tomar*

²⁶ Relativamente a algumas destas unidades multilexicais, como *dar às de vila-diogo* ou *ir para o maneta*, por exemplo, é ainda possível encontrar hipóteses sobre a sua possível origem, mas o falante comum não as conhece.

partido). Como é típico dos grupos formados com verbos leves, o complemento direto destas expressões é geralmente um grupo nominal (i.e., um sintagma nominal sem determinante (cf. Cap. 20)), e as expressões são tipicamente equivalentes a um único verbo pleno, o que comprova o seu caráter unitário, do ponto de vista lexical e semântico. Apresentam-se a seguir alguns exemplos:

- (49) a. fazer pressão = pressionar
 b. dar início = iniciar
 c. dar valor = valorizar
 d. pôr fim = finalizar
 e. fazer frente = enfrentar
 f. tomar partido = optar

Apresentam-se seguidamente alguns exemplos de unidades multilexicais semi-transparentes, organizadas segundo o processo semântico que determina a sua interpretação figurada:

- (i) Comparação (*são como um pero, fino como um coral, lindo como os amores, fresco que nem uma alface, magro que nem um carapau, dormir como uma pedra, viverem/darem-se como Deus com os anjos*)
- (ii) Metáfora (*palavras azedas, ideias brilhantes, carta fora do baralho, pau mandado, fruto proibido, arma de dois gumes, passar a pente fino, deitar achas na fogueira, dizer cobras e lagartos, as paredes têm ouvidos, voz de cana rachada, lágrimas de crocodilo*)
- (iii) Metonímia (*fuga de cérebros, ser um bom garfo, ser um bom copo, ser uma boa cabeça, pedir a mão, beber um copo*)
- (iv) Hipérbole (*mil e uma coisas, morrer a rir, morrer de medo*)
- (v) Eufemismo (*ir desta para melhor, despedir-se da vida, ir para o céu, partir deste mundo*)

[3] Embora sejam noções relacionadas, a interpretação não composicional e a opacidade semântica (ou seja, a ausência de qualquer motivação para a descodificação do significado) não são conceitos completamente equivalentes, sendo possível encontrar unidades multilexicais com uma, mas não com a outra propriedade. As unidades multilexicais composicionais são sempre semanticamente transparentes (cf. *renovar um contrato, falta de preparação*) e as unidades multilexicais semanticamente opacas são sempre não composicionais (cf. *dar às de vila-diogo, por uma unha negra*). No grupo das unidades semitransparentes, no entanto, encontram-se casos de não composicionalidade, ainda que a interpretação não seja completamente opaca. Assim, por exemplo, *baixar a cabeça* não tem uma interpretação composicional, pois nenhum dos seus elementos é usado no seu sentido literal. No entanto, muitos falantes podem considerá-la transparente porque existe um processo metafórico facilmente compreensível, que compara o sentido estrito da expressão com o significado global de 'aceitação/desistência' que se lhe atribui; *baixar a cabeça*, no sentido literal, representa uma postura corporal que pode corresponder a submissão, sujeição. Ou seja, o sentido de uma unidade multilexical não composicional tem de ser apreendido como um todo, mas a sua maior ou menor transparência/opacidade depende da

capacidade que o falante tem de recuperar a motivação de base, intuindo o sentido figurado que lhe está associado, em função de condicionantes lexicais, sintáticas, culturais, etc. Note-se também que, ao contrário do que acontece com as propriedades formais, a determinação da transparência ou opacidade de uma unidade multilexical implica uma certa dose de subjetividade. Em contrapartida, é mais fácil (e consensual) determinar se o sentido de uma unidade multilexical é composicional (ou seja, obtido a partir do significado das suas partes constitutivas) ou não (tendo, neste caso, de ser aprendido de forma global).

9.3.2.1 Uso das unidades multilexicais em línguas especializadas

Muitas unidades multilexicais são usadas em domínios do conhecimento especializados (embora, por vezes, possam entrar para o uso comum, como se discute adiante), como elementos das chamadas **línguas especializadas**. Exemplificam-se alguns casos a seguir:

- (i) Artes (*natureza-morta, música de câmara, canto gregoriano, meio soprano, boca de cena, compasso binário*)
- (ii) Economia (*oferta e procura, economia keynesiana, sistemas de valor, estruturas de mercado*)
- (iii) Jurisprudência (*deduzir acusação, decisão sumária, pena agravada, matéria de facto*)
- (iv) Política (*voto de confiança, separação de poderes, minorias parlamentares*)
- (v) Desporto (*livre direto, corrida de fundo, falsa partida, salto à vara, subir a fasquia*)
- (vi) Marinha (*amplitude da maré, ferro a pique, mar chão, mar grosso, à bolina, vela latina*)
- (vii) Energia (*energia solar, energia eólica, energia hidráulica, energia nuclear, central nuclear, gás natural, painel solar*)
- (viii) Zoologia – animais marinhos (*linguado legítimo, linguado da areia, perca do Nilo, perca da rocha, pargo mulato, solha das pedras, falso pampo*)
- (ix) Vestuário (*fato de saia e casaco – também saia e casaco ou saia-casaco –, camisa de dormir, fato de banho, saltos altos, chapéu de palha*)

Tal como os diferentes domínios de especialidade, a diferença de géneros discursivos determina uma especialização das unidades multilexicais. Assim, algumas são próprias do discurso técnico-científico, outras do discurso jornalístico, outras do discurso televisivo ou publicitário, etc. Daremos aqui exemplos de unidades multilexicais usadas no discurso técnico-científico e no discurso publicitário.

Nas línguas especializadas do discurso técnico-científico, é muito comum o uso de unidades multilexicais de natureza nominal e com carácter denominativo, dada a necessidade de se atribuírem designações a novas descobertas científicas e a novas tecnologias ou produtos (cf. *genoma humano, ácido desoxirribonucleico (ADN), tomografia axial computadorizada (TAC), energia solar, antena parabólica, cartões de memória, computador pessoal (PC)*).

Nas línguas especializadas, os termos científicos e técnicos (incluindo as unidades multilexicais) admitem variação conceptual e denominativa (cf. Cabré 1999:120). Assim, contrariamente ao que tradicionalmente se considerava, esses termos (incluindo as unidades multilexicais) admitem sinonímia, como se exemplifica a seguir:

(50) Marinha

- a. mar chão/mar de leite/mar de rosas
- b. mar grosso/mar cavado
- c. ferrar âncora/pôr âncora

(51) Zoologia (animais marinhos)

- a. bicuda gigante/barracuda gigante
- b. escorpião de água doce/peixe-sapo
- c. raia pintada/raia estrelada
- d. tubarão-martelo/peixe-martelo

(52) Física (energia)

- a. terrenos de cobertura/terrenos mortos
- b. carvão classificado/carvão calibrado

(53) Medicina

- a. pressão arterial/tensão arterial
- b. arteriografia coronária/angiografia coronária
- c. doença de Christmas/hemofilia B
- d. esclerose em placas/esclerose múltipla

Esta variação depende, muitas vezes, dos públicos a que se destina o discurso científico e técnico, ou seja, do grau de especialização desse público e também da situação comunicativa. Por exemplo, o mesmo conceito pode ser transmitido por uma unidade multilexical usada por especialistas – um termo – ou por uma unidade multilexical sinónima, mais típica do discurso oral ou escrito dirigido a leigos – uma expressão comum – (cf. *cefaleia* vs. *dor de cabeça*, *melanoma maligno* vs. *cancro da pele*). Pode, ainda, observar-se variação diacrónica na terminologia (cf. *bolina esternalida*, forma antiga de *bolina folgada*, *trissomia XXI*, termo que atualmente se usa para designar *mongolismo*, ou *acidente isquémico transitório*, que hoje substitui o termo *ictus*).

Também é de notar que, tal como as palavras em geral, as unidades multilexicais científicas e técnicas podem passar a ser usadas em áreas de especialidade diferentes daquela em que tiveram origem: p.e., *teoria do caos*, termo inicialmente usado em física, é hoje usado em muitas disciplinas, da medicina à linguística. Levado ao seu extremo, este processo resulta na **vulgarização** da unidade multilexical (também chamada **banalização** ou **vocabularização**), ou seja, na sua passagem ao uso comum a partir de uma língua de especialidade, com a conseqüente mudança de estatuto de termo científico para vocábulo comum. São exemplos deste processo expressões como *anos-luz* e *entrar em órbita*, termos originados na astronomia e correntemente usados, o primeiro, para designar hiperbolicamente grandes distâncias, e o segundo, o facto de alguém ter ficado completamente descontrolado; *subir a fasquia*, termo com origem no desporto e usado com o sentido de 'subir o nível';

mar de rosas, termo metafórico da náutica, usado para designar um momento de calma e felicidade.

Inversamente, há palavras (cf. *rato*, usado em informática) e unidades multilexicais não especializadas (cf. *correr um programa* ou *abrir uma janela*, usados em informática), que passam a ser usadas como termos científicos ou técnicos, num processo chamado **terminologização**.

O discurso publicitário é um domínio que utiliza frequentemente unidades multilexicais de uso comum, “desconstruindo-as”. Assim, p.e., a partir do aforismo *diz-me com quem andas, dir-te-ei quem és*, foi feito o anúncio *dizemos com quem andamos [...] dir-lhe-ão quem somos* (Markimagem)²⁷; a partir das unidades multilexicais *ter um parafuso a menos, dar uma mãozinha* ou *fazer castelos no ar*, encontram-se os seguintes anúncios publicitários: *Você tem um parafuso a menos? Nós não* (Fabory); *Sempre que você dá uma mãozinha, alguém recebe duas* (peditório nacional da Abraço); *Nestas férias não faça castelos no ar... faça-os na areia* (TAP Air Portugal)²⁸.

Inversamente, certas expressões criadas em anúncios publicitários passam, por vezes, para a linguagem comum, como, por exemplo, o *slogan* publicitário da autoria do poeta Alexandre O’Neill *há mar e mar, há ir e voltar*, para uma campanha de socorros a náufragos, que passou para a linguagem comum como um verdadeiro aforismo, ou o *slogan* publicitário para a Coca-Cola, da autoria de Fernando Pessoa, *primeiro estranha-se, depois entranha-se*, que passou a ser usado na linguagem comum relativamente a diversos produtos ou ideias novas.

Também as diferentes situações de comunicação em que se inserem os enunciados determinam a ocorrência de diferentes unidades multilexicais próprias da interação discursiva. Neste caso, estão envolvidas unidades que correspondem a diferentes atos de fala e a distintas rotinas conversacionais, e que se especializam de acordo com a situação: p.e., se os participantes se encontram face a face ou se comunicam a distância, se o fazem oralmente ou por escrito, segundo o maior ou menor grau de formalidade da interação, o tipo de discurso, etc. Alguns destes usos são exemplificados a seguir:

- (i) Fórmulas de saudação (*Bom dia! Boa tarde! Boa noite! Como está? Passou bem? Seja bem aparecido! Até qualquer dia! Até depois! Até à vista!*)
- (ii) Fórmulas de cortesia (*Com licença, Peço desculpa, Se faz favor, Faz favor de dizer, É servido?, Bom proveito!, Estimo as melhoras!*)
- (iii) Fórmulas de organização do discurso (*como ia dizendo, a propósito, por exemplo, por falar em, por um lado, por outro lado, em primeiro lugar*)
- (iv) Fórmulas de início do discurso escrito (*Meu caro amigo, Caro senhor, Excelentíssimo Senhor*)
- (v) Fórmulas de fecho do discurso escrito (*Um abraço, Um beijo, Com os meus melhores cumprimentos*)

²⁷ Exemplo de Teixeira (2007).

²⁸ Exemplos de Coimbra (2000).

- (vi) Fórmulas de modalização (*como sabe, pode crer, de facto, com efeito, em princípio, quer dizer, isto é, de mais a mais*)
- (vii) Estratégias de diálogo (*Não posso acreditar! Não {vai/vais/vão} acreditar! Contado ninguém acredita! Quem havia de dizer! Quem adivinhasse! Qual carapuça! Era só o que faltava! Não faltava mais nada!*)

9.3.3 Aspetos quantitativos das unidades multilexicais

A intuição dos falantes não é suficiente para identificar as inúmeras combinatórias de palavras presentes em vários tipos de discurso como sendo unidades multilexicais; de facto, só o recurso a análises quantitativas de *corpora* vastos e diversificados permite fazê-lo, mediante a identificação das tendências associativas das palavras e das suas cosseleções preferenciais²⁹.

Por outras palavras, os falantes nativos combinam as palavras que formam as unidades multilexicais de forma adequada porque as restrições combinatórias (as “cosseleções”) que as regem fazem parte do seu léxico mental, mas isso não significa que tenham necessariamente consciência do estatuto da combinação que produziram como unidade multilexical, sobretudo quando se trata de coocorrentes privilegiados³⁰.

Em contrapartida, num *corpus*, a existência regular de padrões colocacionais é quantitativamente observável. Por exemplo, para qualquer unidade multilexical arbitrária, é fácil observar num *corpus* que, de entre um conjunto de palavras sinónimas ou quase sinónimas (cf. Cap. 8), apenas algumas (por vezes uma só) ocorrem na sua formação e que as outras não ocorrem, ou ocorrem apenas pouco significativamente. Assim, p.e., de acordo com dados do *corpus* COMBINA, duas palavras que nos dicionários são dadas como sinónimas, *total* e *absoluto*, ocorrem em combinatórias distintas com IC significativos: com *total* mas não com *absoluto* ocorreram, p.e., *descalabro total*, *eclipse total*, *colapso total*, *colesterol total*; em contrapartida, com *absoluto* mas não com *total* ocorreram *prioridade absoluta*, *recorde absoluto*, *monarquia absoluta*, *temperatura absoluta*.

É principalmente nas unidades multilexicais formadas por coocorrentes privilegiados, os quais, como foi dito, detêm um menor grau de coesão formal e semântica, que a análise quantitativa de dados empíricos se torna indispensável para determinar o grau de tendência associativa que existe entre os seus elementos. O estudo quantitativo inclui análises de frequência e outras análises estatísticas, nomeadamente o estabelecimento do Índice Combinatório (IC) (cf. 9.3 e Notas 13 e 15).

A identificação de unidades multilexicais através de dados quantitativos é feita com recurso a *corpora* linguísticos de grandes dimensões. Só assim, de facto, se pode evidenciar a existência de associações lexicais padronizadas e obter informação sobre os seus significados e usos convencionais. Estas associações constituem um

²⁹ Não se trata aqui de “restrições de seleção” de carácter gramatical ou semântico (cf. Cap. 11), mas de associações lexicais que permitem identificar e caracterizar padrões de coocorrência entre palavras, incluindo no plano gramatical e discursivo.

³⁰ A este propósito, veja-se a seguinte passagem de Fellbaum (2007:2): «A idiomatidade de uma língua revela-se talvez melhor através dos erros que cometem os que estão a aprendê-la.» (t.n.).

aspecto importante da estrutura lexical de uma língua³¹. Nesses *corpora*, identifica-se a frequência de ocorrência de cada uma das palavras da sequência em estudo, assim como a frequência com que ocorre a sequência (estes dois tipos de dados permitem estabelecer o Índice Combinatório).

O Índice Combinatório³² é um resultado estatístico que permite avaliar se, num dado *corpus*, a frequência da sequência é maior do que seria previsível relativamente à frequência de cada um dos seus elementos; este resultado, por sua vez, permite também medir a força da associação lexical existente entre os componentes da sequência.

O Índice Combinatório permite, ainda, avaliar se os elementos da expressão têm tendência para ocorrer apenas ou quase exclusivamente juntos ou se também ocorrem com uma grande variedade de outras palavras que não aquelas que compõem a expressão em análise. Se uma determinada palavra ocorrer maioritariamente na expressão em causa e se a sua frequência junto de outras palavras for reduzida, isso aponta para uma alta probabilidade de a expressão ter um grau forte de lexicalização.

[4] A identificação das unidades multilexicais através de uma análise quantitativa torna patentes vários tipos de tendências associativas entre as palavras, de que se dão aqui exemplos para as palavras *abolição* e *mágico*³³:

- (i) a. *abolição*: *abolição da escravatura* (11,6), *abolição da pena de morte* (13), *abolição das fronteiras fiscais* (15), *abolição das portagens* (11,7), *abolição do tráfico de escravos* (14,4)
- b. *mágico*: *artes mágicas* (10,1), *caixinha mágica* (13,3), *como que por artes mágicas* (12,39), *espelho mágico* (8), *flauta mágica* (13,1), *lanterna mágica* (12,5), *lugar mágico* (7,2), *momento mágico* (6,9), *palavra mágica* (7,2), *poção mágica* (14,5), *poderes mágicos* (11,7), *varinha mágica* (15,7)

Outras palavras há que ocorrem significativamente sempre combinadas com um número muito restrito de palavras, como *ínfimo*: cf. *ínfimo pormenor* (13,3) e *ínfima parte* (8,5).

Alguns itens lexicais ocorrem frequentemente em unidades multilexicais, o que decorre da sua aplicabilidade a variadas áreas da realidade extralinguística. Por exemplo, a palavra *contrato* entra numa vasta série de unidades multilexicais:

- (54) a. *contrato a prazo*, *contrato a termo*, *contrato a termo certo*, *contrato sem termo*
- b. *contrato social*, *contrato coletivo*, *contrato coletivo de trabalho*
- c. *contrato de adjudicação*, *contrato de aluguer*, *contrato de arrendamento*, *contrato de compra e venda*, *contrato de promessa de compra e venda*, *contrato de concessão de serviços*, *contrato de empreitada*, *contrato de*

³¹ Cf. Sinclair (1991), Fernando (1996), Biber et al. (1999), Brinton e Traugott (2005), Altenberg (2005) e Bacelar do Nascimento et al. (2006).

³² Cf. Church e Hanks (1990:22ss) e Pereira (1994).

³³ Exemplos retirados do *corpus* COMBINA. Os números entre parênteses são os dos Índices Combinatórios que se registaram neste *corpus*.

fornecimento de serviços, contrato de locação financeira, contrato de manutenção, contrato de prestação de serviços, contrato de *leasing*

- d. regime jurídico do contrato, termos do contrato, cláusulas do contrato, data do contrato, minuta do contrato, modelo de contrato, denúncia do contrato, duração do contrato, execução do contrato, renovação do contrato, final do contrato, incumprimento do contrato, cessação do contrato, rescisão do contrato, revisão do contrato, alteração do contrato, vigência do contrato

As análises quantitativas apontam, ainda, para a recorrência de diversos tipos de unidades multilexicais, do ponto de vista da sua estrutura gramatical. Assim, são particularmente frequentes as seguintes estruturas:

- (i) Nome + adjetivo (*ação prolongada, ajuda humanitária, ano escolar, ano letivo, ar livre, área coberta, boletim meteorológico, cadeia hierárquica, centro comercial, classe média, condomínio fechado, construção civil, cultura geral, direitos humanos, drogas ilícitas, entrada livre, estado lastimoso, estrutura hierárquica, facto consumado, fluxo migratório, fogo posto, fonte fidedigna, força aérea, fórmula mágica, guerra aberta, habilitações literárias, luto carregado, mercado negro, nomes sonantes, opinião pública, orla marítima, ponte aérea, população ativa, suplemento alimentar*)
- (ii) Nome + de + nome (*ar de família, área de serviço, arma de fogo, ato de misericórdia, cabeça de casal, caixa de óculos, caminho-de-ferro, camisa-de-forças, campo de ação, cara de pau, carro/carrinho de linhas, carta de recomendação, dia de anos, direito de antena, dor de alma/dó de alma, esfera de ação, fome {de cão/de lobo}, golpe de misericórdia, motivo de força maior, orelhas de abano, paz de alma, período de garantia, porto de abrigo, prazo de garantia, qualidade de vida, razão de força maior, rede de abastecimento, sessão de abertura, sistema de abastecimento, teia de aranha*)
- (iii) Sintagmas verbais com verbos leves (*dar ajuda, dar autorização, dar início, dar origem, dar resposta, dar um beijo, dar um passeio; fazer a barba, fazer batota, fazer críticas, fazer perguntas, fazer pressão, fazer queixa, fazer uso; meter dó, meter medo; pôr em risco; ter admiração, ter dúvidas, ter medo, ter necessidade, ter paciência, ter pena*)

Concluindo, as unidades multilexicais não são casos marginais, antes constituem uma parte importante da produção lexical³⁴, como diversos estudos sobre *corpora* têm demonstrado: não só são muito recorrentes no discurso, mas também é linguisticamente significativo o seu estatuto na gramática e no léxico mental dos falantes (cf. Fellbaum 2007:2), conforme se procurou mostrar neste capítulo.

³⁴ Moon (1998:57ss) refere a grande recorrência das unidades multilexicais, em correlação também com a tipologia dos textos estudados, e Sinclair (2004), usando uma definição lata de unidade multilexical, afirma que, segundo estimativas feitas sobre *corpora* ingleses, 80% das ocorrências das palavras são cosseleções, ou seja, apenas 20% serão escolhas independentes.