

O Problema em torno do conceito de composição

O grande problema que se observa em relação à composição é o estabelecimento de critérios para sua identificação. Para o processamento automático, esta é uma questão de grande relevância, pois uma palavra composta deve ser analisada computacionalmente como um bloco, para que se possam construir representações adequadas da estrutura sintática e semântica das frases em que se encontra.

Seqüências como *sala de jantar*, *acerto de contas*, *acidente de trabalho*, *caderneta de poupança* devem ser computadas, interpretando-se as três conjuntamente e, portanto, cada uma como uma única ocorrência ou disjuntivamente e, nesse caso, temos ocorrência de três palavras diferentes?

Para responder a esses questionamentos, primeiro faremos uma reflexão sobre o conceito de composição apresentado por alguns lingüistas e gramáticos no item 6.1 e, em seguida 6.2, uma abordagem a partir dos dicionários manuais e, por fim, no item 6.3, uma abordagem pela teoria do léxico-gramática. Esta última é que sustenta todo embasamento teórico desta pesquisa, quanto ao estabelecimento de critérios, que possam identificar uma seqüência, com estrutura Nde N livre de uma seqüência N de N composta.

5.1

Uma abordagem de lingüistas e gramáticos

Nas gramáticas do português (Cunha e Cintra, 1984; Mateus *et al.*, 1994 - *apud* Ranchhod, 1993), os termos usados para designar as categorias constituídas por seqüências coesas de elementos lexicais parecem depender mais da parte do discurso a que dizem respeito do que do diferente comportamento lingüístico dessas seqüências. Assim, os advérbios, as preposições e as conjunções são referidos por 'locuções': adverbiais, prepositivas e conjuntivas (Cunha e Cintra, 1984:540-1 e 586, respectivamente; Mateus *et al.*, 1994:206-7); para os

substantivos e adjetivos, opta-se pelo termo 'composto' (Cunha e Cintra, 1984:187-9 e 252-3; Mateus *et al.*, 1994:185 e 203).

O problema em torno do conceito de palavra composta nas gramáticas é corroborado por Basílio (1999:9), quando afirma que a questão da delimitação de unidades lexicais se manifesta em abordagens tradicionais, apresentando controvérsia sobre várias questões, entre elas a noção de compostos.

De fato, não há consenso sobre o conceito de palavra composta abordado por alguns gramáticos. Bechara (2001:355) afirma que composição consiste na criação de uma palavra nova de significado único e constante, sempre e somente por meio de dois radicais relacionados entre si.

Com esse critério não seria possível explicar, por exemplo, nomes compostos do tipo *lua-de-mel*, *chinelo de dedo*, *chá de panela*, *lágrima de crocodilo*, etc, visto que não é a relação entre os dois radicais que garante o sentido de composição.

Já em Cunha e Cintra (1985:104), encontramos informações que lembram a possibilidade de transparência semântica das partes: uma palavra composta representa sempre uma idéia única e autônoma, muitas vezes dissociada das noções expressas pelos seus componentes. Assim, *criado-mudo* é o nome de um móvel; *vitória-régia*, o de uma planta; *pé-de-galinha*, o de uma ruga no canto externo dos olhos. É curioso observar que Cunha e Cintra admitem a possibilidade de uma associação de sentido com os componentes, porém só apresentam exemplos que não demonstram essa relação de sentido. De fato, o critério da "idéia única e autônoma" dissociada das noções dos componentes não explica, por exemplo, casos de compostos como *mesa de tênis*, *chinelo de dedo*, *relógio de pulso*, etc., pois mesa de tênis é uma mesa, chinelo de dedo é um chinelo e um relógio de pulso é um relógio.

Parece que essa questão é mais clara para Luft (1990:80), pois admite que o significado de uma palavra composta pode estar relacionado ou não ao significado das partes, quando afirma que nessa combinação os elementos primitivos perdem parcial ou totalmente a significação própria em benefício de uma significação nova, global. A palavra composta exprime um conceito novo, mais ou menos distinto, pois em alguns os sentidos primitivos dos componentes persistem com alguma nitidez, o que não acontece em outros.

Nesse caso, parece que a posição de Luft explica a sequência *pai de família* como composta, mas não dá para saber pela exposição do autor se é com perda parcial ou total da significação das partes.

Para esclarecer, Luft cita o exemplo *guarda-chuva*, mostrando que a relação de sentido está no fato de que não há dúvida de que seja um objeto para nos guardar da chuva. Mas observa também que o mesmo não acontece com *manda-chuva* visto que não é um cidadão que manda chuva.

Esse é um ponto que merece cuidado, no sentido de avaliarmos até que ponto a perda parcial ou total do sentido dos elementos componentes da sequência pode definir ou não uma palavra como composta.

De certo modo, Mattoso Câmara (1985:76) tenta resolver o problema, apresentando o mesmo exemplo citado por Luft - *guarda-chuva* - afirmando que composição é a formação de uma palavra pela reunião de outras, cujas significações se complementam para formar uma significação nova.

Para Bechara, não é uma questão de complementação, mas, sim, de dependência, afirmando que a composição consiste na criação de uma palavra por meio de duas ou mais outras cuja significação depende do que encerra os seus componentes.

Como explicar pelo critério de complementação ou de dependência de sentido, por exemplo, *cebola de cabeça* e *chinelo de dedo*?

Rocha Lima (1986:), embora reconheça o processo de composição como um fenômeno único e diversificado em língua portuguesa, não apresenta uma reflexão detalhada sobre o assunto, definindo a composição como um processo autônomo de formação de palavras em português diferente da derivação e da onomatopéia. Juntando-se duas bases preexistentes na língua, o falante pode criar um novo vocábulo dito composto.

De certo modo, essa posição de Rocha Lima encontra ressonância em Basílio, quando ela afirma que na teoria lexical a composição representa o princípio da liberdade, porque podemos juntar qualquer coisa para formar uma nova palavra. Acrescenta ainda que composição é o espaço de todas as combinações para cumprir determinadas necessidades do falante. Mas observa que, por outro lado, formada a nova palavra, pode-se afirmar que uma palavra composta é uma unidade lexical em que nenhum dos elementos tem liberdade, isto

é, o sentido da composição não permite inserção ou substituição de um dos constituintes, caso contrário, a combinação perde o sentido de unidade lexical.

O que caracteriza e define a função do processo de composição em Basílio (1987:29-30) é a sua estrutura, de tal maneira que, das bases que se juntam para formar uma palavra, cada uma tem seu papel definido pela estrutura, como, por exemplo, *sofá-cama*, *livre-arbítrio*, *guarda-roupa*. Assim, a composição é um processo de formação de palavras que utiliza estruturas sintáticas para fins lexicais. Daí permitir a nomeação ou caracterização de seres pela junção de dois elementos semânticos, de existência independente no léxico, em apenas um elemento lexical. Em consequência disso, as formas compostas são freqüentemente desligadas do significado estrito de seus componentes.

Em alguns casos, segundo Basílio, é possível observar a transparência do significado. Isso acontece quando a forma composta descreve as características gerais do objeto de nomeação, que passa a ser o referente da forma composta, a exemplo, *navio-escola*.

Um outro exemplo da própria autora é o de *guarda-vestido*. Não há nada na composição que indique que se trata de uma peça de mobiliário, embora muitos pudessem interpretá-la como uma sequência transparente. Teoricamente, o termo poderia ter sido usado para designar malas, por exemplo.

Ainda é oportuno refletir aquelas palavras que têm a função de nomeação metafórica. Isso se dá quando a descrição de um objeto, ao invés de caracterizá-lo por critérios objetivos, estabelece para este uma descrição em termos de propriedades transferidas em termos associativos. Nesse caso, a interpretação só é possível se já conhecemos o sentido da metáfora, não sendo possível inferir o significado através da simples observação das formas: é impossível, por exemplo, inferir um inseto a partir de *louva-a-deus* ou um docinho a partir de *olho-de-sogra*.

Silva e Koch (1983) preferem retomar a questão, apenas, do ponto de vista semântico, quando afirmam que a composição é o processo de formação de palavras que cria novos vocábulos a partir de palavras já existentes, dando origem a um novo significado. Por meio desse processo, combinam-se dois morfemas lexicais, operando-se entre eles uma fusão semântica que pode ser mais ou menos completa. Assim, por exemplo, em *guarda-chuva* o significado de cada elemento

persiste com certa nitidez; já em *pé-de-galinha* este significado desaparece para dar lugar a outro.

Nesse sentido, Basílio e Silva & Koch estão em harmonia, pois para Basílio (1987) na composição há exclusivamente uma função semântica - função de designação. Essa designação vai ser efetivada por meio da combinação do significado das duas palavras.

Carone (1986:37) insere uma reflexão sobre o aspecto sintático, afirmando que composição é um procedimento pelo qual uma construção sintática se imobiliza, dando origem a uma unidade cristalizada. Em decorrência, forma-se um sintagma bloqueado, com duas características essenciais da palavra: inseparabilidade e irreversibilidade das partes articuladas. E o composto, embora relembre resignadamente os elementos constitutivos, tem um terceiro significado, distinto do significado dos elementos constitutivos

Nesse aspecto, pode-se lembrar Basílio (1987:30), quando afirma que a composição é a utilização de estruturas sintáticas para fins lexicais. Por exemplo, *guarda-roupa* e *guarda-chuva* são dois substantivos formados pela junção de um verbo e um substantivo. O primeiro com função de nome de agente e o segundo, de instrumento.

Para Basílio, a definição de composição, além de apresentar problemas morfológicos e sintáticos, merece também uma reflexão do ponto de vista lexical. A combinação *beija-flor*, por exemplo, tem uma regra de redundância. Regras de redundância são as que expressam relações lexicais que definem informações. *Beija-flor* é a função de um verbo e um nome para formar um substantivo que retoma o referente *pássaro*, mas também a semântica do *beijo*. É, portanto, um nome de agente caracterizado pela ação do verbo sobre o substantivo. É isso que dá a composição uma natureza diferente, porque ela precisa de motivação lexical.

Carone (1986) não menciona a motivação, trata-se de um sintagma bloqueado que se cristaliza e dá origem a uma nova palavra.

Já Sandman (1992:32) afirma que o critério semântico é o melhor critério para distinguir a palavra composta do grupo sintático paralelo. Isso é possível por alguma forma de isolamento ou distanciamento. Essas formas de isolamento podem estar fundamentadas nos critérios semântico, sintático, morfológico ou fonológico. Sandman apresenta ainda uma distinção entre grupos sintáticos fixos,

por exemplo, *tênis de mesa*, "tipo de jogo ou esporte" e grupos sintáticos eventuais, por exemplo, *pé de meia* (que deixei cair) e *copo de leite* (que bebi). De acordo com Sandman, os fixos referem-se a entidades estabelecidas da nossa cultura, como que estereotipadas, como nomes permanentes, com caráter, portanto, de lexemas, enquanto os grupos sintáticos eventuais são sintagmas da frase. Mas, é o critério semântico, segundo ele, que mais nos socorre para distinguirmos compostos de grupos sintáticos. Por exemplo, *copo-de-leite* - tipo de flor - não se distingue de *copo de leite* - copo com leite - pelo critério fonológico, pois a pausa acentual é a mesma, nem pelo critério morfológico, pois a formação do plural é a mesma (copos-de-leite), nem pelo sintático. Para esse mesmo autor, o que nos socorre de fato é a semântica: *copo-de-leite* - tipo de flor - com base na semelhança. O aspecto metafórico representa uma entidade distinta da de *copo de leite* - copo com leite.

O que Sandman se esqueceu de observar é que o sentido de "copo de leite que bebi" se desfaz em relação ao "copo-de-leite que enfeitou a sala", justamente pela restrição de seleção do verbo e pela presença dos complementos. Isso é de natureza sintático-semântica e é o que contribui na interpretação da seqüência. Sandman (1991:50) considera que a metáfora e a metonímia desempenham papel importante para distinguir o significado do composto lexical do significado do grupo sintático. A transferência ou ampliação do significado de uma unidade lingüística se dá por contigüidade ou associação espacial. Em seguida apresenta alguns exemplos, em que fatores metafóricos são responsáveis pela formação do composto, ou seja, a transferência ou ampliação do significado se dá por associação baseada na semelhança, como, por exemplo, *pé-de-galinha* (ruga no canto externo dos olhos), *pé-de-valsa* (dançarino habilidoso), *pé-de-cabra* (nome de uma ferramenta). Em outros casos, a transferência ou ampliação do significado de uma unidade lingüística se dá por contigüidade ou associação espacial.

Para Azevedo (2000:101), as combinações abaixo são exemplos de metonímia em que a entidade referida pelo composto é identificada por sua utilidade ou função, como, por exemplo, *saca-rolha* (utensílio de cozinha), *porta-voz* (pessoa que fala em nome de outra) ou qualquer característica tipificadora, como, por exemplo, *cara de pau* (pessoa sem escrúpulos, que

mantém as feições inalteradas, mesmo diante de situações embaraçosas), *mão-aberta* (pessoa generosa, que gasta dinheiro com facilidade).

Sandman aplica ainda os conceitos de compostos endocêntricos e exocêntricos, por exemplo, a *trem-bala* e *peixe-espada*. Apenas os determinantes "bala" e "espada" estão sendo empregados no sentido figurado. *Trem-bala* e *peixe-espada* são compostos endocêntricos, porque *trem-bala* continua sendo "trem" e *peixe-espada* continua sendo "peixe". Já o composto *perna-de-moça* não é mais uma "perna", porém um "peixe". A extensão de significado é para fora da combinação, por isso é um composto exocêntrico.

Observa-se que a argumentação apresentada por Sandman não leva em conta a possibilidade de inferência de quem não conhece o sentido dessa combinação. *Trem-bala*, por exemplo, poderia não ser a designação de *trem*. A título de curiosidade interrogamos uma pessoa que não tinha essa informação, e ela inferiu que se tratasse de uma máquina de fazer bala.

Para Monteiro (1991:161), a grande preocupação é distinguir o vocábulo composto de uma locução e aponta como primeiro critério levar em conta a ordem fixa dos elementos. Em geral os compostos não permitem a troca de posição de seus componentes, sem que com isso o significado se altere. O segundo critério refere-se à impossibilidade de intercalar qualquer elemento. As locuções admitem essa intercalação. Porém, muitas locuções deixam de admitir essa possibilidade e é por isso que se consideram como nomes compostos locuções do tipo *segunda-feira*, *amor-perfeito*, *obra-prima*, etc. Um outro critério que identifica as locuções é a redução de um dos componentes da combinação, por exemplo, *guarda-chuva*. Se dissermos *o guarda* ou somente *a chuva*, o significado mudará totalmente. Mesmo assim, Monteiro lembra ainda que algumas locuções constituem construções sintáticas fechadas, não facultando por isso a supressão de qualquer elemento. Os nomes *pé-de-galinha*, *pé-de-moleque* e *unha-de-fome*, considerados compostos, para ele não passam de expressões que morficamente não se caracterizam como nomes compostos, e, entretanto, nunca admitem a possibilidade de supressão de qualquer elemento.

É curioso observar que em momento algum essas formas são lembradas por Monteiro como compostos pela extensão de um novo sentido ou pela denominação metafórica, conforme reflexão apresentada em Basílio e Sandman.

Segundo Monteiro, as fronteiras que distinguem o vocábulo composto da locução são bastante imprecisas, e as gramáticas listam entre os compostos muitos exemplos que, segundo o critério acima exposto, constituem em verdadeiras locuções. Estão nesse rol, por exemplo, *segunda-feira*, (segunda), *terça-feira* (terça). Ainda em Monteiro (1991:166), constata-se que a dificuldade em distinguir o vocábulo composto de locução se deve ao fato de que a composição vem sendo interpretada como um processo morfológico. Na realidade, na maioria das situações tem-se um processo de natureza sintático-semântica. Observando a construção *amor-perfeito*, tem-se a combinação de um substantivo e um adjetivo no plano sintático, a julgar pelo processo de concordância nominal. Mas essa articulação gera uma especificação semântica, uma unidade semântica.

Já a explicação de Biderman (1978:21) caminha por outro lado. Para ela, quanto mais o conjunto dos elementos que formam uma palavra complexa se afastar ou isolar semântica, fonológica e morfológicamente dos elementos que o compõem individualmente, mais perfeita será a integração e mais claramente se caracterizará a nova unidade lexical. O conhecimento de uma unidade léxica resulta da configuração dos vários níveis de análise - fonológico, morfológico, sintático e semântico. Porém, não é uma operação simples. Às vezes, os linguistas se defrontam com numerosas ambigüidades de difícil solução. A classificação de uma palavra é resultante do ambiente em que ela se encontra e da sua distribuição relativa. A palavra é um constituinte da frase, ela torna efetiva a significação da mesma; mas ela não aparece necessariamente na frase com o sentido que tem como unidade autônoma. Assim, ela orienta para o teste da substituição e da inserção, para identificarmos uma seqüência já lexicalizada de uma ainda não soldada no nível do sistema léxico. Tomemos um exemplo da própria autora: numa seqüência como *dor de cabeça* não diremos *dor terrível de cabeça*, mas *terrível dor de cabeça* ou *dor de cabeça terrível*. O teste de inserção de um novo item, nessa seqüência, demonstra que a combinação *dor de cabeça* já está lexicalizada.

Biderman critica as gramáticas por considerar palavras como *guarda-chuva*, *terça-feira* como compostos por justaposição quando, de fato, os seus elementos componentes já se aglutinaram e por isso deveriam ser considerados como unidades simples, porque são sintagmas já lexicalizados na língua. Essa

concepção traria um grande problema para o processamento automático porque graficamente essas palavras são constituídas por mais de uma palavra e se uma palavra como *guarda roupa* fosse interpretada como uma palavra simples, considerando-se qualquer possibilidade de erro, em que a palavra *roupa* fosse substituída por *chuva*, por exemplo, todas as relações de sentido do contexto frasal estariam comprometidas e, obviamente, a geração seria de um texto incoerente.

Como se pode observar, o conceito de composição apresenta realmente muitas controvérsias, pois os autores apresentam designações e hipóteses diversas. O grande problema da literatura gramatical é, de fato, a quase ausência de critérios formais para a identificação de nomes compostos, sobretudo para aqueles com a estrutura NdeN.

De modo geral, pode se afirmar que a identificação dos nomes compostos na literatura gramatical é feita com base, principalmente, em critérios semânticos. A concentração dos estudos está muito vinculada à noção de transparência semântica, ou seja, da perda ou não das relações de sentido da seqüência com os elementos constituintes. Poucos tentaram apresentar como alternativa teórica para abordagem dos nomes compostos a "hipótese transformacional", mesmo assim, não deixaram claros, por exemplo, quais são os critérios para se avaliar se uma seqüência, com estrutura Nde N, constitui um sintagma nominal livre ou um sintagma nominal composto. Em geral, os exemplos apresentados são de estrutura Nadjetivo.

Como vimos, há pouquíssimas abordagens com intuito de apresentar critérios para identificação de compostos com estrutura N de N. Para atender os nossos propósitos, ainda há necessidade de um estudo descritivo, o mais detalhado possível, sobre as propriedades das estruturas N de N, apresentando critérios formais que nos permitam identificar um grupo nominal livre de um grupo nominal composto, a fim de que possamos representar essas propriedades e formalizá-las para processamento da linguagem natural. Portanto, a noção de composição que apresentaremos nesta pesquisa está respaldada em critérios formais, que levam em conta as propriedades estruturais internas das estruturas N de N, candidatas a compostos a partir da observação do comportamento morfossintático e semântico dessas seqüências em frases, que nos permitam

avaliar a aceitabilidade ou não das relações de sentido expressas por elas nesses contextos frasais.

5.2

Uma abordagem a partir dos dicionários manuais

Com uma breve consulta ao *Dicionário Houaiss da língua portuguesa*, pode-se constatar o volume de nomes compostos com estrutura NdeN. Mas o que se pode observar é que não existe um critério que mostre a diferença entre uma entrada lexical e uma subentrada, ou seja, entre um verbete e um subverbo, problema também já constatado por Vale (2001:59) em relação às expressões cristalizadas do português do Brasil. Tomemos como exemplos alguns verbetes:

Acidente

1. acontecimento casual, fortuito, inesperado,
 - 1.1 qualquer acontecimento desagradável ou infeliz, que envolva dano, perda, lesão, sofrimento ou morte.

Acidente de percurso

- 1 - acontecimento imprevisto que interrompe a evolução, até então positiva, de um fenômeno, de um processo econômico, social, etc.

Acidente de precaução

- 1- *Mus* acidente colocado numa nota para prevenir erro na leitura.

Peito

- 1- região do tronco que vai do pescoço ao abdome; tórax
- 2- nos vertebrados, porção anterior ou ventral do tórax
- 3- cada um dos seios femininos.

Peito do pé

- 1- parte superior do pé

Cavalo

- 1- mamífero da família dos eqüídeos, nativo das estepes da Europa e da Ásia, mas encontrado em todo o mundo como animal doméstico.

Cavalo de batalha

- 1- fig. argumento principal no qual se insiste.
2- fig. complicação, embaraço, dificuldade.

Fim

- 1- momento ou ponto em que se interrompe um fenômeno, período ou ação.

Fim de semana

- 1 -*Cron* período de tempo que se inicia na noite de sexta-feira e se encerra na manhã de segunda-feira.

Fim de mundo

- 1- lugar longe, de difícil acesso e pouco habitado: cafundó, fim do mundo.

Uma triagem dos compostos que aparecem no Aurélio e no Houaiss com a palavra "língua" ilustra bem a ausência de critérios:

Aurélio (Ferreira, 1999)

Verbetes -*Língua*

1. *Órgão muscular alongado, móvel, situado na cavidade bucal a cuja parede inferior está preso pela base, e que serve para a degustação, para a deglutição e /para a articulação dos sons da voz.*
2. *Designação comum a diversos objetos que têm semelhança com esse órgão.*
3. *Conjunto das palavras e expressões usadas por um povo, por uma nação, e o conjunto de regras da sua gramática; idioma.*
4. *Língua vernácula*

Modo de expressão escrita ou verbal de um autor, de uma escola, de uma época.

língua comprida / língua de trapo / língua do pê / língua extinta / língua flexional / língua secreta / língua suja / língua vernácula / língua viperina / segunda língua / desenferrujar a língua / dobrar a língua / não falar a mesma língua / puxar pela língua de / língua de sujeito nulo.

Verbetes separados: língua-base / língua-de-badalo / língua-de-sola.

Houaiss (2001):

Verbete Língua

1- Órgão muscular recoberto de mucosa, situado na boca e na faringe, responsável pelo paladar e auxiliar na mastigação e na deglutição.

5- Sistema de representação constituído por palavras e por regras que as combinam em frases que os indivíduos de uma comunidade lingüística usam como principal meio de comunicação.

5.1- O idioma nacional.

(...)

7- Estilo de expressão particular a um grupo social, profissional ou cultural

7.1- Estilo de expressão característico de um escritor, uma escola, um movimento, uma época.

Subverbetes

língua aglutinante / língua artificial / língua azul / língua comprida / língua comum / língua cristã / língua de chegada / língua de trapo / língua de palmo / língua negra / língua presa / dar a língua / dar com a língua nos dentes / de língua passada / estar com a língua coçando.

Verbetes separados

língua-alvo / língua-de-costa / língua de sogra / língua-de-trapo.

Não é preciso citar muitos verbetes nem todas as acepções para se perceber que não há um critério para a classificação de uma subentrada. Não são observados os sentidos dos nomes formados por mais de uma palavra. Por exemplo, "peito do pé" é uma subentrada de "peito", mas o seu sentido não é calculado a partir do sentido de "peito", que é a entrada. Já "fim de semana" é uma subentrada de "fim". De fato, existe uma flutuação semântica em relação à palavra "fim", que é a entrada. Mas o sentido de "fim de semana" também não pode ser calculado a partir do sentido das partes. A sequência *fim de semana* tem uma restrição semântica: o *fim de semana* começa na sexta à noite ou no sábado cedo, e acaba no domingo à noite. Portanto, tem uma duração de 48 a 60 horas. As outras formas "fim de Ntempo" não têm esta restrição de duração: não se sabe, por exemplo, quanto tempo dura o *fim de ano* (um dia, uma semana, o mês de dezembro). Outro aspecto a ser observado também é que a sequência *fim de semana* não tem as mesmas restrições entre as propriedades que se podem destacar normalmente em grupos nominais livres NdeN com o apagamento de N2 : (Maria saiu no fim de semana), (*Maria saiu no fim). Podemos ainda recorrer a outra propriedade sintática que as formas "fim de Ntempo" possui e "*fim de semana*" não possui. Observe-se a forma "Detdef fim de Ntempo" - (Detdef = determinante definido), quase sempre é possível substituí-la por "o fim de Detdef Ntempo". Comparando-se os exemplos: (1) "Passamos este fim de semana juntos" e (2) "Passamos o fim desta semana juntos." O sentido de fim de semana tem duração certa em (1) e não em (2). O fim de semana em (2) pode ser da quinta à noite até sexta à noite. Portanto, (2) é um exemplo de construção livre "fim de Ntempo", enquanto que (1) é um nome composto.

O que temos observado, ainda que superficialmente, pois não avaliamos o dicionário na sua totalidade, é que, para que um nome formado por mais de uma palavra constitua uma subentrada, basta, apenas, que o primeiro item seja idêntico, do ponto de vista da forma, à palavra simples que ocupa a posição de verbete, ou seja, de entrada lexical. E para que constitua uma entrada deve ser uma palavra cristalizada há muito tempo. Mas esses critérios não levam em conta a composicionalidade. Também não é observado o sentido de cada subentrada em relação à complexidade das estruturas.

Sabemos que os dicionários convencionais têm objetivos bastante diferentes daqueles dicionários usados em Linguística Computacional, porque descrevem as palavras para quem já fala a língua e conhece a cultura. E que a obrigatoriedade ou não de compostos como entradas implica muitas questões, como, por exemplo, o tamanho e o público alvo do dicionário. Mas, por outro lado, entendemos a composição como um tema muito importante que ainda não recebe a devida atenção por parte de lexicógrafos / dicionaristas. Essa falta de critério se justifica, segundo Vale (2001:59), pela ausência de um conceito de lexia complexa, no que concordamos sem restrições.

Muitos compostos têm o *status* de entrada lexical. Todos que foram observados para fortalecer essa argumentação são grafados com hífen, provavelmente, por obediência à norma apresentada no PVOLP (XXXVI), que tem como princípio geral para o emprego do hífen a seguinte recomendação: *só se ligam por hífen os elementos dos vocábulos compostos em que se mantém a noção da composição, isto é, os elementos das palavras compostas que mantêm a sua acentuação, porém formando o conjunto perfeita unidade de sentido*. No mesmo documento o primeiro critério para o emprego do hífen também recomenda: *emprega-se o hífen nos vocábulos compostos em que os elementos, com sua acentuação própria, não conservam, considerados isoladamente, a sua significação, mas o conjunto constitui uma unidade semântica: água-marinha, arco-íris, galinha-d'água, couve-flor, guarda-pó, pé-de-meia, pára-choque, porta-chapéus, etc.*

Não encontramos nenhuma subentrada grafada com hífen, o que demonstra ser, ainda, o "uso do hífen" um critério para classificação de nome composto, como se pode notar também, na maioria dos exemplos arrolados nas gramaticais tradicionais. Lima (1989:199) afirma que não serve de critério para a identificação de justapostos e aglutinados a presença ou a ausência do hífen, não só por não haver grafia uniforme para as palavras compostas, mas principalmente porque como já acentuara Bréal - o sentimento da composição "tem seu critério no espírito". Porém, apresenta cinco exemplos com hífen e apenas um sem.

Segundo G.Gross (1991:685), os dicionários também não se entendem sobre os elementos lexicais que têm um traço-de-união. No interior de um mesmo dicionário, uma palavra é escrita com ou sem o traço de união. Está sendo feita

para o francês uma lista dos compostos, com traço-de-união, tomando-se como referência dicionários, jornais e revistas, com vistas ao reconhecimento dessas unidades nos textos. Há grandes regularidades: em 45000 compostos do tipo Nome + Adjetivo, somente 200 têm traços de união, dos quais 100 o têm facultativamente. Os compostos do tipo Nome de Nome apresentam uma porcentagem ligeiramente mais elevada. Em oposição, os compostos do tipo Verbo + Nome quase todos têm o traço de união.

Para o português do Brasil, se fosse feito esse levantamento, provavelmente teríamos resultados similares. Silva & Koch & Silva (1983) afirmam que se fosse possível utilizar o hífen como critério, seria mais fácil a identificação na linguagem escrita das ocorrências de justaposição: no entanto, a utilização aleatória deste sinal não permite uma identificação segura.

Certamente, os usuários do português já perceberam que o uso do hífen é uma imposição que não tem nenhuma relação com o sentido de composição.

5.3

Uma abordagem a partir do método de descrição do léxico-gramática

No método de descrição do léxico-gramática, a identificação e a descrição dos nomes compostos vão além das dimensões do léxico, porque leva em conta também a adequação lingüística. A gramática transformacional baseou-se no recurso da intuição dos locutores. Tal método, embora apresente dificuldades, é ainda insubstituível para evidenciar numerosas restrições inesperadas da sintaxe das línguas. É possível observar que certas formas consideradas inaceitáveis com base na intuição não se encontram em textos de grandes dimensões. Sem que isso seja uma garantia de inaceitabilidade, esse processo pode constituir um complemento importante quando da formulação de regras. De igual modo, as dúvidas acerca do emprego de uma dada construção podem ser resolvidas, se essa construção for efetivamente observada num texto.

Considerando-se esses aspectos, segundo Ranchhod (1990) e Baptista (1994), para identificar e descrever nomes compostos de forma fundamentada, é necessário utilizar critérios lingüísticos que vão desde a análise do seu comportamento morfológico até a verificação da total ou parcial perda de

composicionalidade lexical, sintática e semântica. Esses critérios são igualmente necessários para distinguir os nomes compostos de grupos nominais livres, formalmente idênticos, ou seja, aqueles que apresentam a mesma estrutura, N de N, como, por exemplo, *bolsa de festa* e *bolsa de couro*. Do mesmo modo os critérios também servem para distinguir seqüências de palavras ambíguas que tanto podem constituir um composto como corresponder a uma construção livre, como, por exemplo, *chá de bebê* (chá para um bebê ou um tipo de festa) e *colher de chá* (unidade de medida ou nova oportunidade representando facilidade)

O método de descrição do léxico-gramática se contrapõe à literatura gramatical, porque a composição nominal não é definida sob o ponto de vista da homogeneidade como um critério único de definição, como parece ser o caso nas gramáticas tradicionais. No método de descrição do léxico-gramática, vários critérios podem ser usados para avaliarmos se uma seqüência é ou não composta. Uma seqüência será considerada composta, se apresentar restrições quanto às propriedades sintáticas em relação à outra seqüência com a mesma categoria de palavras (Baptista, 1994:8). As restrições podem ser avaliadas, se forem consideradas as relações que ligam a seqüência em frases, a partir de um conjunto de critérios que possam identificar as suas propriedades morfosintáticas e semânticas.

Mas nenhum critério, por si só, é condição suficiente para classificar como composta uma dada seqüência de elementos lexicais. É na intersecção dos vários critérios que se define a composicionalidade de uma dada combinação, que será tanto mais fixa, quanto mais restrições se observarem em relação às propriedades sintáticas que caracterizam um grupo nominal livre formado pela mesma seqüência interna de categorias gramaticais. O grau de composição de um composto não pode, no entanto, ser avaliado apenas pelo número de índices de fixidez que revela, já que os vários critérios sintáticos não são comparáveis entre si:: "les propriétés sont autonomes" Gross (1986:69).

Examinando o *corpus* que constitui objeto de análise desta pesquisa, a partir do conceito definido por Gross (1986:34) - *uma palavra composta é não-composicional* - podemos constatar que esse conceito corrobora com a experiência de nossa análise. Um critério aplica-se a todas as seqüências, porém os resultados dessa aplicação podem ser diferentes, porque as seqüências têm propriedades

distintas. Por exemplo, a seqüência *bolsa de viagem* tem a propriedade N1deN2 ser N1 (uma bolsa de viagem é uma bolsa), já a seqüência *lua-de-mel* não tem, porque *lua-de-mel* (não é uma lua) nem tem a propriedade N1 de N2 ser N2, porque também (não é mel).

Considerando-se a diversidade de propriedades apresentada pelos nomes candidatos a compostos, duas observações merecem destaque: reconhecemos seqüências compostas totalmente opacas cujo significado não pode ser calculado a partir do significado dos componentes, e seqüências em que um dos componentes é transparente, mas o seu significado global também não pode ser deduzido, levando-se em conta essa transparência parcial. Isso quer dizer que uma seqüência com transparência parcial não é necessariamente composicional.

Nesse caso, quando Cunha e Cintra (1985:104) afirmam que uma palavra composta representa sempre uma idéia única e autônoma, muitas vezes dissociada das noções expressas pelos seus componentes, e citam, como por exemplo, *pé-de-galinha*, que denota uma ruga no canto externo dos olhos, pode-se observar que essa noção de nome composto é, por um lado, pouco produtiva e, por outro, só abrange as seqüências em que o significado do composto não apresenta transparência semântica, é não- composicional, ou seja, o significado de uma determinada seqüência não pode ser obtido pela soma do significado das palavras que a compõem. Por exemplo, não há outra palavra composta com a palavra *pé* que seja relacionada ao sentido da combinação *pé-de-galinha*. A palavra *pé-de-moleque*, por exemplo, denota um tipo de doce e não uma ruga no canto externo dos olhos.

Por outro lado, alguns estudiosos admitem que há muitos nomes compostos cujo significado global mantém uma relação com o significado individual de cada um dos seus elementos, o que torna ainda mais difícil a avaliação da composicionalidade de uma combinação. Por exemplo, a combinação *pai de família*, para alguns falantes, pode ter o significado global a partir do significado de um dos seus constituintes. Do ponto de vista da reprodutividade, o conceito de composicionalidade é vulnerável e divergente, porque pode ser uma avaliação de nossa intuição semântica.

Esse é justamente o ponto que destacamos com preocupação para a identificação de compostos, porque parece que a noção de composicionalidade é

avaliada como uma impressão espontânea de cada falante, e essa noção, provavelmente, leva em consideração a avaliação de quem já conhece a língua. Para quem não conhece a língua, uma seqüência como *barriga de aluguel* pode ser interpretada como uma barriga possível de ser alugada e *barriga de freira* como uma barriga que pertence a uma freira e nunca como um tipo de pastel. Ou seja, a transparência parcial não quer dizer que a seqüência seja composicional, isto é, a composicionalidade de uma seqüência vai depender de suas propriedades sintáticas e não da avaliação de quem a observa. Para mostrar essa diferença de julgamento nos apoiamos em Maurice Gross (1986), que tem critérios formais para o tratamento da composicionalidade.

O problema que se coloca só pode ser resolvido, se levarmos em conta as relações entre os elementos a partir de critérios formais. O primeiro critério - o critério geral - que deve ser aplicado para esclarecer a dúvida é o do bloqueio distribucional. Tomando como exemplos as seqüências *testa-de-ferro* e *carne de sol*, não encontramos as formações análogas como **testa de ouro*, **testa de bronze*, **testa de alumínio*. A impossibilidade de estabelecer uma relação paradigmática já é um índice do bloqueio distribucional. Mas esse não é um exemplo que traz dificuldade de avaliação. Qualquer falante do português percebe que essa seqüência revela um novo sentido que não tem nenhuma relação com o sentido dos componentes, porque não temos (**testa de ferro*) ou de qualquer outro material. A seqüência *testa-de-ferro* tem um sentido figurado, significa alguém que faz por nós aquilo que não queremos ou não podemos fazer. Esse sentido é marcado pela invariabilidade do gênero masculino definido pelo determinante *um* - ("Maria é *um testa-de-ferro*"), ("João é *um testa-de-ferro*") em oposição à invariabilidade do gênero feminino definido pelo artigo "a" para "testa" como parte da cabeça - "A testa de João", "A testa de Maria".

O problema maior está na seqüência *carne de sol*, porque a transparência parcial é insuficiente para dar conta de explicar o sentido de toda a seqüência como um processo de preparação de carne. Do mesmo modo que *lua-de-mel* não é uma lua, *carne de sol* poderia não ser carne. Caso contrário, qual seria a função da seqüência "*de sol*" para a comunicação?

Considerando-se a posição de Cunha e Cintra (1985), não poderiam ser incluídos na avaliação dos autores nomes como *cadeira de roda* e *pano de prato*

que, embora apresentem uma transparência parcial, devem ser considerados compostos, porque não há a ocorrência de novos itens para expressar o mesmo sentido dessas palavras. Elas apresentam, portanto, uma distribuição restrita, o que as caracteriza como itens lexicais autônomos. Ou seja, a noção de idéia única defendida pelas gramáticas não deve ser levada em conta apenas pela falta de transparência dos componentes, mas sobretudo pelo bloqueio distribucional. Assim não temos *pano de tigela* para secar tigelas nem *pano de panela* para secar panela. Temos *pano de prato* que denota um tipo de pano cuja função é secar.

Os nomes compostos levantam um segundo problema para o processamento: o de representação. A unidade de representação num léxico linear é, *grosso modo*, a palavra, definida pela sua forma escrita, ou seja: uma seqüência de letras separadas das seqüências vizinhas por meio de lacunas de 'limites'. Em consequência disso, as palavras compostas não podem ser registradas no dicionário de maneira direta, como se faz com palavras simples. É necessário um procedimento de identificação para suas ocorrências em textos. A partir disso, as propriedades lingüísticas formais dos nomes compostos determina o procedimento de identificação nos textos quanto ao tipo de armazenamento que os compostos exigem Gross (1986:20).