

Conclusão

Este estudo apresentou uma descrição lingüística de uma amostra de nomes compostos com a estrutura NdeN. Foram examinadas 1500 seqüências para a observação dos fenômenos léxico-sintático-semânticos.

Os nomes compostos foram analisados a partir de critérios formais, levando-se em conta as propriedades apresentadas pelas seqüências, o que permitiu de maneira sustentável estabelecer critérios de identificação para grupos nominais livres e para grupos nominais compostos que têm a mesma estrutura.

Percebemos que as unidades lexicais compostas não podem ser estudadas se comparadas umas com as outras, mas, sim, inseridas em frases simples para observação do comportamento sintático-semântico dos itens. Esse procedimento segue a hipótese geral do método de descrição do léxico-gramática: uma descrição das unidades de significação não pode ser feita considerando-se a palavra isolada, mas, sim, a palavra em contexto a partir da elaboração de frases simples que apresentam sujeito, predicado e seus complementos.

Todas as seqüências que foram identificadas como compostas são não-composicionais. Segundo Gaston Gross (1990:84), a maioria dos trabalhos sobre a composição nominal se limitam a definir o conceito de composição e propõem um critério único de definição. Esse objetivo pressupõe que se atribui aos nomes compostos uma homogeneidade de estrutura sem a qual a tentativa é votada ao fracasso. Foi o que se passou na literatura gramatical. O conceito de Maurice Gross (1986:34) acolhido nesta pesquisa de que um nome composto deve ser não-composicional é também um elemento de homogeneidade muito forte, mas não impede que o composto sofra variações: redução (bolsa de valores = bolsa), nominalização (acertar as contas = fazer um acerto de contas), variação em gênero e número (amigo/a da onça = amigos/as da onça). Uma vez que cada composto pode apresentar variações diferentes, também existe heterogeneidade nesse modelo.

Nesta pesquisa, observou-se que é preciso considerar todas as relações entre as frases construídas com as seqüências, levando-se em conta as

distribuições sintáticas de seus componentes. Ou seja, é preciso descrever o conjunto de propriedades sintáticas que cada seqüência apresenta. O resultado da análise demonstra que as seqüências têm propriedades diferentes. A diversidade de propriedades encontradas ora em uma, ora em outra seqüência, apontou a necessidade de um estudo detalhado aliado a um conjunto de critérios que permitissem a observação das restrições morfossintáticas da estrutura interna dessa categoria de palavra, isto é, da estrutura NdeN.

Para este estudo os nomes candidatos a compostos foram testados com base nas propriedades sintáticas de cada seqüência, representados da seguinte forma: (N1deN2 ser N1); (N1deN2 ser N2); (*N1deN2 ser N1), (*N1deN2 ser N2); (N2 ter N1). (*N2 ter N1)

As restrições sintático-semânticas foram observadas, avaliando-se a existência de uma ligação sintática entre frases, por meio de alguns critérios: ausência de bloqueio distribucional, ausência de ruptura paradigmática, inserção lexical, apagamento de N2, coordenação, pronominalização, restrições de seleção do nome, variação em gênero e número.

Entre os critérios citados, destaca-se como critério geral para a identificação de uma palavra composta o bloqueio distribucional, seguido da impossibilidade de inserção lexical, ruptura paradigmática e o apagamento de N2.

Evidentemente uma seqüência, na qual a aplicação de todos os critérios citados aponta na direção da composicionalidade, não pode ser considerada composta. Uma seqüência não-composicional, ou seja, composta, é aquela que apresenta bloqueio distribucional e ruptura paradigmática. A ruptura paradigmática não é um critério independente do bloqueio distribucional, ela representa um índice do bloqueio distribucional.

Para sustentar o conceito de não-composicional, a seqüência deve, por meio das suas propriedades, apresentar uma distribuição restrita. Por outro lado, nenhum critério, por si só, é condição suficiente para a classificação de uma dada seqüência, porque a classificação de composto não depende do número de critérios aplicados a cada seqüência Baptista (1994:124), pois a não composicionalidade não é quantificável.

Com base em Silberztein (1990:76), observou-se que o conjunto das propriedades sintáticas dos nomes compostos NdeN não pode ser deduzido a partir das propriedades dos constituintes.

Os critérios usados apontam caminhos para novas pesquisas, porque podem ser aplicados a outras categorias de palavras, tanto do português do Brasil, quanto de outras línguas. Assim, torna-se possível um estudo descritivo comparativo de nomes compostos entre diferentes estruturas e entre diferentes línguas.

Constatamos que é grande o número de compostos com verbo suporte *dar*, *ter*, *fazer*. Mas não foi objetivo desta pesquisa um estudo detalhado dessas construções. Muitos nomes foram examinados em frases construídas com verbo suporte, porém se pode afirmar que não é a presença ou a ausência do verbo suporte que define se uma dada seqüência é ou não composta. Entretanto, há seqüências "aparentemente" com estrutura NdeN cujo verbo suporte faz parte do composto, como em *dar o ar da graça* e *fazer papel de bobo*. Esse tipo de composto aponta a necessidade de um estudo descritivo e de uma codificação específica para essa estrutura. Não o fizemos por reconhecermos que não tem a estrutura NdeN, objeto de análise nesta pesquisa.

A codificação das descrições das entradas lexicais de nomes compostos tem um valor técnico e comercial considerável, por serem utilizáveis na engenharia da linguagem, em aplicações tais como a busca de informações, a correção gramatical e a sumarização automática. Qualquer aplicação que envolva a análise e a geração de textos automáticos por sistema de processamento da linguagem natural, como a tradução automática, a indexação e a geração de sumários, necessita de um dicionário que inclua entrada de compostos. A descrição flexional, em especial, tem uma aplicação direta na indexação automática de textos.

Com essa descrição definiu-se um conjunto de regras e critérios de delimitação de unidades lexicais, constituindo-se uma base para incorporação de novos itens ao léxico. Os nomes definidos como nomes compostos receberam uma representação formal, obtendo-se como resultado a criação de um dicionário eletrônico utilizável em processos eletrônicos.

Chegamos à conclusão de que um estudo descritivo das palavras compostas para elaboração de dicionário eletrônico, se configura como uma condição para a melhoria da qualidade de programas computacionais, que lidam com o processamento da linguagem natural, por duas razões fundamentais: primeiro, porque segundo Gross *et al* (1990), o número de palavras compostas constitui a maior parte do léxico de qualquer língua; segundo, a representação computacional deve explicitar as relações entre frases. Do mesmo modo, é necessária, também, a explicitação das relações entre componentes de palavras, pois em muitas línguas existem muitas palavras com característica de composição. A descrição de nomes compostos pode, possivelmente, solucionar grande parte de um dos problemas clássicos no processamento das línguas - reconhecimento de formas numa sequência linear sem o comprometimento do sentido das informações, eliminando-se ambigüidades, redundâncias, repetições e informações agramaticais.