

ESTUDOS LINGÜÍSTICOS
DOSSIÊ GRAMÁTICAS CATEGORIAIS

AS GRAMÁTICAS CATEGORIAIS

José Borges Neto*

As gramáticas categoriais são um conjunto de formalismos aparentados destinados ao tratamento da sintaxe e da semântica das linguagens em geral, sejam elas línguas naturais, como o português ou o inglês, sejam elas linguagens formais, como as linguagens da lógica, da matemática e da computação.

A origem desses formalismos é o algoritmo definido pelo lógico polonês Kazimierz Ajdukiewicz, em 1935, para determinar a *boa-formação* (*gramaticalidade*, se quisermos usar o termo consagrado pelo trabalho de Chomsky) das expressões de uma linguagem.

Praticamente desconhecidas dos lingüistas – até porque, no confronto com a gramática chomskiana, nos anos 50, a proposta categorial foi vencida –, mesmo no domínio das linguagens formais, por muitos anos as gramáticas categoriais tiveram pouca influência. O quadro muda drasticamente, no entanto, a partir da publicação de um conjunto de trabalhos escritos por Richard Montague, no início dos anos 70, e com o desenvolvimento dos recursos de informática, que passaram a exigir mecanismos para o tratamento computacional das línguas naturais.

Por suas características, as gramáticas categoriais, muito mais do que outras formas de gramática (a chomskiana entre elas), prestam-se muito bem para o tratamento das línguas naturais que o processamento automático exige.

* Universidade Federal do Paraná.

O fato de tratarem, de forma integrada e “matematicamente” rigorosa, os fenômenos sintáticos e os fenômenos semânticos, permite às gramáticas categoriais uma descrição lingüística muito interessante para os lingüistas e muito útil para os pesquisadores da área do processamento de línguas naturais (PLN).

As características principais dessas gramáticas são descritas a seguir:

1. Uma forma extrema de *lexicalismo*. Toda a construção da gramática repousa nas características formais dos itens lexicais. As gramáticas categoriais, no limite, não são mais do que especificações de um léxico categorizado. Em outras palavras, uma gramática categorial é apenas um léxico em que cada item lexical é atribuído a alguma categoria constante num conjunto, infinito e definido recursivamente, de categorias sintáticas. A simples atribuição de um item a uma categoria específica a combinatória sintática do item, dispensando assim qualquer outro mecanismo sintático para a avaliação da gramaticalidade das expressões sintáticas em que o item aparece.
2. A caracterização dos constituintes, sejam eles constituintes sintáticos ou semânticos, como *funções* e/ou *argumentos*, no sentido matemático desses termos. Todo item lexical pertence a alguma categoria e toda categoria ou é básica ou é derivada: as categorias básicas são sempre argumentos de uma função e as categorias derivadas são tanto argumentos quanto funções que tomam outras categorias (básicas ou derivadas) como argumentos. O que é relevante destacar aqui é que todos os elementos constituintes de uma expressão podem ser classificados como argumentos ou funções. Em outras palavras, numa estrutura lingüística qualquer só existem funções e argumentos.
3. A caracterização das relações entre a sintaxe e a semântica, bem como as relações entre os constituintes de ambos os níveis, em termos estritamente *composicionais*, mantendo os tipos sintáticos e semânticos na relação mais próxima possível.
4. Uma tendência para “liberar” a estrutura de constituintes de superfície, garantindo, pelas características vistas, que toda estrutura permitida pela sintaxe da gramática possa ser integralmente interpretada na semântica. O que de fato importa é o conjunto de relacionamentos funcionais que se podem estabelecer entre os elementos e não sua “ordem” na superfície. As várias “ordens” superficiais possíveis manifestam estruturas funcionais; se as “ordens” superficiais manifestam a mesma estrutura funcional, estamos diante de um caso de “ordenação facultativa”; se as “ordens” superficiais manifestam (ou podem manifestar) estruturas funcionais distintas, estamos diante de um caso de ambigüidade (estrutural ou de escopo). Embora nem todas as versões da gramática categorial admitam-na, a “liberação” da

estrutura superficial é um recurso teórico interessante, que vai permitir que a gramática dê conta das “ordens” esdrúxulas que encontramos, por exemplo, na linguagem oral e que tanto perturbam os lingüistas.

O algoritmo básico, definido por Ajdukiewicz, trata “aritmeticamente” as concatenações de expressões lingüísticas como se fossem casos de multiplicação de frações. Para tanto, as categorias recebem “índices” (expressões de uma metalinguagem que nomeiam essas categorias) da seguinte forma: as categorias básicas recebem índices “inteiros” (a categoria dos nomes próprios, por exemplo, recebe o índice N; a categoria das sentenças recebe o índice S) e as categorias derivadas (funtoras) recebem índices fracionários em que, no denominador, encontramos a categoria do argumento tomado pela função e, no numerador, encontramos a categoria do valor resultante da concatenação. Num exemplo simples:

(1) Pedro corre

Suponhamos que já sabemos que a expressão *Pedro corre* pertence à categoria das sentenças (*Pedro corre* é S); suponhamos também que já sabemos que a expressão *Pedro* pertence à categoria dos nomes próprios (*Pedro* é N). Qual poderia ser a categoria de *corre*? Tratemos a expressão *Pedro corre*, aritmeticamente, como se estivéssemos diante de uma multiplicação:

$$(1') \quad \begin{array}{cccccc} \textit{Pedro} & & & & \textit{corre} & \\ \text{N} & & \text{X} & & x & = \text{S} \end{array}$$

Pelo procedimento de isolamento da incógnita (o x fica sozinho e o N que o acompanha passa para o outro lado do sinal de igualdade; como o N está multiplicando o x , passa para o outro lado dividindo), a seqüência dos índices passaria a ser a seguinte:

$$x = \frac{s}{N}$$

Dessa forma, a categoria a que pertence a expressão *corre* seria $\frac{s}{N}$ (S/N).¹

Se já sabemos a categoria das duas expressões atômicas que compõem a expressão maior, o índice categorial dessa expressão, bem como a avaliação de sua “boa-formação”, pode ser obtido por uma operação, também aritmética, de

1 Por razões de ordem tipográfica, as frações que nomeiam as categorias funtoras podem ser representadas por expressões lineares em que o primeiro elemento é o numerador da fração, o segundo elemento é a barra de fração e o terceiro elemento é o denominador.

cancelamento de fração, que funcionaria segundo o processo de simplificação permitido pela aritmética quando da multiplicação de frações. Se temos a seqüência “ $N \times S/N$ ”, podemos “cancelar” o N que é numerador da primeira fração (o “inteiro” é entendido como fração de denominador 1) com o N que é denominador da segunda fração, obtendo o “inteiro” S como resultado. A possibilidade desse “cancelamento” nos indica que a expressão “Pedro corre”, por exemplo, é uma expressão bem-formada e é uma expressão da categoria S . Por exemplo:

$$\frac{N}{1} \cdot \frac{S}{N} = \frac{S}{1} = S$$

Para dar conta da ordenação superficial dos constituintes presentes na cadeia de expressões lingüísticas, podemos indicar, de alguma forma, a posição do argumento em relação a seu functor. No caso visto acima, por exemplo, poderíamos indicar que o argumento N deve estar à esquerda do functor $\frac{S}{N}$, escrevendo-o da forma $N \backslash S$. Desta forma, o índice categorial S/N passaria a indicar um functor que toma um N à direita para resultar num S e o índice categorial $N \backslash S$ indicaria um functor que toma um N à esquerda para resultar num S .

Fica aberta ainda a possibilidade de acrescentarmos novas categorias, e novas regras, ao conjunto utilizado na descrição de fenômenos lingüísticos, desde que essas categorias resultem da definição do conjunto de categorias possíveis e as regras se justifiquem inferencialmente das regras já assumidas. Por exemplo, se podemos concluir que o verbo *corre*, em *Pedro corre*, pertence à categoria $N \backslash S$ (já assumindo um critério de explicitação da posição relativa do argumento) a partir do fato de que *Pedro* é um N e *Pedro corre* é um S , podemos assumir também que *Pedro* poderia ser um $S/(N \backslash S)$ a partir do fato de que *corre* é um $N \backslash S$ e *Pedro corre* é um S . O raciocínio seria o seguinte: Suponhamos que já sabemos que a expressão *Pedro corre* é um S ; suponhamos também que já sabemos que a expressão *corre* é um $N \backslash S$. Tratemos novamente, como fizemos acima, a expressão *Pedro corre*, aritmeticamente, como se estivessemos diante de uma multiplicação:

$$(1'') \quad \begin{array}{ccccc} \textit{Pedro} & & \textit{corre} & & \\ x & \times & N \backslash S & = & S \end{array}$$

Pelo procedimento de isolamento da incógnita (o x fica sozinho e o $N \backslash S$ que o acompanha passa para o outro lado do sinal de igualdade; como o $N \backslash S$ está multiplicando o x , passa para o outro lado dividindo), a seqüência dos índices passaria a ser a seguinte:

$$x = S/(N \setminus S)$$

Se isso é possível neste caso e se a operação utilizada é “aritmeticamente” válida, podemos propor uma regra geral que nos permita passar de categorias como N para categorias como $S/(N \setminus S)$. Essa regra – denominada *Alçamento de Tipo* – poderia ser explicitada como:

(2) Regra de Alçamento de Tipo

$$X \rightarrow Y/(X \setminus Y)$$

$$X \rightarrow (Y/X) \setminus Y$$

X e Y seriam categorias quaisquer e as duas formulações da regra cobririam as possibilidades de o funtor resultante tomar seu argumento à direita ou à esquerda.

Regras como essa poderiam ser acrescentadas ao sistema, tornando-o mais poderoso e mais adequado à descrição lingüística.

Outros mecanismos também podem, de forma legítima, acrescentarem-se a essa “gramática mínima” que apresentamos aqui e alguns deles estarão exemplificados nos textos a seguir. Não vou me estender mais, já que o propósito dessa pequena introdução é apenas o de apresentar o raciocínio básico que comanda o procedimento descritivo contido nas gramáticas categoriais. O leitor interessado em saber mais sobre as gramáticas categoriais poderia procurar informações mais detalhadas em outros lugares. Creio que seriam particularmente úteis os seguintes trabalhos: Borges Neto (1979), Lyons (1979 - cap. 6, item 6.3), Steedman (1993) e Wood (1993).

Os textos que constituem esse dossiê são bastante distintos entre si e exemplificam tanto o tratamento que uma gramática categorial pode oferecer a conjuntos distintos de fenômenos lingüísticos quanto alguns mecanismos teóricos que se podem utilizar no trabalho descritivo. Pela ordem.

O texto que escrevi em parceria com Márcio Renato Guimarães, mais do que tentar descrever algum fenômeno lingüístico em particular, pretende apresentar um mecanismo teórico que poderia ser útil para o tratamento de fenômenos lingüísticos que envolvam relações outras entre expressões que não a relação funtor/argumento. Usamos, quase como uma “desculpa” para introduzir os mecanismos teóricos enfocados, os fenômenos ligados à concordância nominal em português. Não é propósito do texto, no entanto, tratar a concordância. O propósito do texto é apresentar ao leitor a chamada gramática categorial de unificação que, além dos mecanismos teóricos comuns a outras versões da gramática categorial, introduz *traços* nas entradas lexicais e um mecanismo de

“checagem” de traços. Acreditamos que mecanismos dessa natureza poderiam dar conta não só dos fenômenos de concordância em português (e/ou outras línguas), mas também de fenômenos ligados à regência (nominal ou verbal). Possivelmente, um mecanismo de unificação como o que apresentamos poderia permitir um tratamento relativamente simples de fenômenos como a atribuição de casos aos argumentos do verbo (ou do nome) em línguas de casos, como o latim. Poderia também, eventualmente, dar conta de fenômenos de correlação de tempos em sentenças complexas.

Os textos de Gelza Nunes-Pemberon e de Thais Chaves foram resultado de um curso de introdução às gramáticas categoriais que lecionei no segundo semestre de 1997 no programa de pós-graduação em Linguística da USP. O texto de Gelza compara o tratamento dispensado por uma gramática categorial aos fenômenos conhecidos como *strong-crossover* ao tratamento corrente na gramática chomskiana; o texto de Thais compara a solução gerativa e a solução categorial para o problema da extração do sujeito. O interesse maior desses dois textos está na demonstração da capacidade de uma gramática categorial, que não admite nem movimentos nem categorias vazias, em dar um tratamento lingüisticamente razoável a fenômenos que, numa gramática gerativa de extração chomskiana, dependem crucialmente tanto da noção de movimento quanto da noção de categoria vazia.

Finalmente, encerramos esse dossiê com um texto de Luiz Arthur Pagani, que procura discutir a compatibilidade das gramáticas categoriais com teorias do desempenho lingüístico e sua conseqüente utilidade no processamento de línguas naturais. Trabalhando já na interface da lingüística com a ciência cognitiva e com a ciência da computação, Pagani vai discutir a hipótese de que uma gramática categorial, por suas características, permita um processamento automático da linguagem compatível com o que sabemos sobre o processamento psicolingüístico das línguas naturais.

REFERÊNCIAS BIBLIOGRÁFICAS

- AJDUKIEWICZ, K. Die syntaktische Konnexität. *Studia Philosophica*, v. 1, p. 1-27, 1935 (tradução para o inglês em McCALL, S. (Ed.). *Polish logic*. Oxford: Oxford University Press, 1967. p. 207-223; tradução para o italiano em BONOMI, A. (Ed.) *La struttura logica del linguaggio*. Milão: Bompiani, 1973. p. 345-372.
- BORGES NETO, J. Da gramática categorial. *Construtura*, v. 17, p. 1-13, 1979.
- LYONS, J. *Introdução à lingüística teórica*. São Paulo: Nacional/Edusp, 1979.
- STEEDMAN, M. Categorial grammar. *Lingua*, v. 90, p. 121-158, 1993.
- WOOD, M. M. *Categorial grammars*. Londres: Routledge, 1993.