

A LINGÜÍSTICA E A FORMALIZAÇÃO

Elena Godoy *

СОДЕРЖАНИЕ

В настоящей работе поднимаются некоторые вопросы, связанные с проблемой применения формальных /математических/ методов в исследовании фундаментальных языковых закономерностей. Как пример формальной лингвистической модели, кратко анализируется аппликативная грамматика С.К.Шаумяна.

RESUMO

No presente trabalho são levantadas algumas questões relacionadas com o problema de utilização de métodos formais (matemáticos) no estudo de regularidades lingüísticas fundamentais. Como exemplo de modelo lingüístico formal, é apresentada a gramática aplicacional de S. K. Shaumyan.

Existem diferentes pontos de vista sobre o lugar da Lingüística entre outras ciências, que dependem da maneira de encarar a própria linguagem humana. É possível entender a linguagem como parte da cultura; então, o estudo das línguas humanas fará parte da antropologia, da história, etc. Ou, se a linguagem for encarada como uma das faculdades cognitivas do homem, a lingüística será vista como um ramo da psicologia (esta é a postura de N. Chomsky, o pai da conhecida Gramática Gerativa). Agora, se a lingüística for entendida como um ramo das ciências exatas, ela poderá fazer parte ou da própria matemática (e assim a coloca, por exemplo, Richard Montague), ou, então, da cibernética (esta é a postura de Sebastian Shaymyan¹, que nos interessa aqui). Vista desta maneira, a lingüística — a lingüística formal, naturalmente, — torna-se compatível com a teoria dos algoritmos e, conseqüentemente, com a computação, o que permite, por exemplo, trabalhar com os cálculos lingüísticos na construção da Inteligência Artificial, fazer a simulação dos fenômenos lingüísticos no computador, etc.

Partindo do conceito da cibernética que estuda sistemas de qualquer natureza, capazes de apreender, conservar e transformar a informação e usá-la para o controle e a regulação, o sistema de signos de qualquer língua natural pode ser considerado como uma espécie de mecanismo cibernético que transforma um tipo de informação lingüística em um outro tipo de informação (lingüística). Na entrada ("input") dos mecanismos "lingüístico-cibernéticos", encontra-se a informação sobre um número finito de objetos lingüísticos e sobre as regras de ge-

ração de objetos novos a partir dos objetos iniciais. Em um número finito de passos, o mecanismo gerativo fornece, na saída ("output"), palavras e sentenças gramaticalmente corretas e lhes atribui características estruturais.

A teoria lingüística, na concepção de Shaumyan, deve atingir os seguintes objetivos:

- estabelecer os universais lingüísticos, verdadeiros para qualquer língua natural;
- estabelecer os princípios das variações entre as línguas naturais;
- ser capaz de explicar os fatos das línguas particulares;
- prover os mecanismos conceptuais e formais para a construção das gramáticas das línguas naturais.

Shaumyan postula que, para atingir os objetivos estabelecidos, é necessário um sistema abstrato de operadores lingüísticos que revelará as estruturas idênticas subjacentes às construções das línguas particulares. Revelando a mesma estrutura subjacente a fenômenos lingüísticos aparentemente incompatíveis, essa gramática abstrata explicará tais fenômenos como casos particulares da ação dos mesmos operadores.

Para atingir seus objetivos, a teoria lingüística deve possuir uma metalinguagem adequada. Visto que toda metalinguagem necessariamente interfere na descrição da língua-objeto, uma das tarefas importantes da teoria lingüística é procurar reduzir esta interferência até o mínimo. Como as línguas naturais misturam o racional, lógico, com o irracional, Shaumyan considera que as palavras e as sentenças das línguas-objetos não devem ser geradas diretamen-

(*) Elena Godoy. Formada em Letras e Mestre em Lingüística de Língua Espanhola pelo Instituto Pedagógico Estatal A. Herzen (Leningrado, URSS); Mestre em Lingüística de Língua Portuguesa pela UFPR; Doutoranda em Lingüística na UNICAMP (Campinas, SP); Professora do Curso de Extensão de Língua Russa na UFPR.

te pelos mecanismos lógico-matemáticos, e propõe uma *teoria a dois níveis* ("degraus"). Nesta teoria, Shaumyan introduz um nível abstrato (o nível de construtos), um verdadeiro nível de teoria lingüística, em que os universais recebem um status próprio. Shaumyan chama este nível de *genótipo* e o opõe ao *fenótipo* — o nível de objetos lingüísticos reais (o nível de observação).

Para Shaumyan, o genótipo, a língua genotípica, é o sistema semiótico universal comum a todas as línguas humanas, a "carga genética" das línguas naturais. A língua genotípica é entendida assim como uma armação geral, na base da qual as línguas humanas diversificam-se, apresentando características particulares. Neste nível das invariantes, onde é feita a abstração destas características particulares e de todas as irregularidades que afetam a estrutura das línguas naturais, *podem* ser formuladas as regras lógicas que não admitem exceções. Neste, e apenas neste nível, o uso do mecanismo lógico pode ser justificado e pode ser eficaz.

Quanto ao fenótipo, entende Shaumyan que neste nível a gramática transforma os objetos ideais da língua genotípica em objetos reais que pertencem a uma ou outra língua natural. Esta transformação é feita, introduzindo-se dentro do esquema geral (genotípico) toda uma série de limitações próprias da língua em estudo.

A metalinguagem que Shaumyan usa na construção da língua genotípica, é baseada na lógica combinatória desenvolvida por Curry para prover uma base firme na resolução de paradoxos lógicos. Para realizar o cálculo do nível da língua genotípica, Shaumyan escolhe os seguintes primitivos, ou seja, as classes essenciais de expressões lingüísticas:

— nomes de objetos, ou **termos** — α —, que podem ser interpretados empiricamente como substantivos (no sentido lato): Maria, cadeira, ele, cachorro, etc.;

— nomes de situações, ou **sentenças** — β —, por exemplo, 'Maria passeia';

Tomando o termo e a sentença na qualidade de nomes de duas classes lingüísticas primitivas, Shaumyan introduz, também como universais lingüísticos, as classes de **transformadores (operadores)** — funções — que são calculados da seguinte maneira: os objetos de uma classe podem ser transformados em objetos de outra classe, ou, então, em objetos da mesma classe. Deste modo, se obterá mais quatro classes de objetos lingüísticos: $\Delta\alpha\alpha$, $\Delta\alpha\beta$, $\Delta\beta\beta$, e $\Delta\beta\alpha$. No nível empírico, essas classes podem ser interpretadas como:

$\Delta\alpha\beta$ — transformador de termos em sentenças; por exemplo, o predicado de um lugar 'passeia' em 'Maria passeia', ou 'é professora' em 'Maria é professora';

$\Delta\beta\alpha$ — transformador de sentenças em termos; por exemplo, 'que' em 'A avó sabe que Maria passeia';

$\Delta\alpha\alpha$ — transformador de termos em

outros termos; por exemplo, o sufixo '-inha' em 'casinha', ou o adjetivo 'pequena' em 'casa pequena', ou preposição + substantivo 'de campo' em 'casa de campo'. Note-se o isomorfismo existente em português entre os sufixos, adjetivos e as construções 'preposição + substantivo' — é apenas uma rápida observação que salta à vista dos pouquíssimos exemplos que apresentamos;

$\Delta\beta\beta$ — transformador de sentenças em outras sentenças; por exemplo, o negador 'não' em 'Ele não sabe matemática' ou os advérbios sentenciais como 'agora' em 'Agora ele sabe matemática'.

Havendo agora seis classes de objetos lingüísticos, torna-se possível transformar os objetos de qualquer classe em objetos da mesma classe, ou então, em objetos de uma das cinco classes restantes. Teremos assim mais 32 classes de transformadores. Esse procedimento pode ser continuado "ad infinitum". De modo geral, o cálculo de universais lingüísticos é feito da seguinte maneira:

$$\frac{\Delta\alpha\beta \ \alpha}{\beta}$$

ou, generalizando:

$$\frac{\Delta pq \ p}{q},$$

onde Δpq é operador; p — operando; q — resultando. A própria fórmula expressa a **operação de aplicação** que dá o nome ao modelo criado por Shaumyan: Gramática Aplicacional. A fórmula acima pode ser interpretada empiricamente assim: $\Delta\alpha\beta$ é o predicado de um lugar ('passeia'); α é o termo ('Maria'); a aplicação de $\Delta\alpha\beta$ a α resulta em β , que é a sentença 'Maria passeia'. O conjunto de todas as expressões aplicacionais do genótipo é gerado por meio de regras de formação.

Toda língua natural pode ser vista como um conjunto de sentenças composto de dois planos: o de conteúdo e o de expressão. O plano de conteúdo é o significado da sentença e o plano de expressão é o "invólucro físico" da sentença. Assim, a sentença é uma unidade bilateral — tem dois componentes. A situação, entretanto, é uma unidade unilateral que tem apenas um componente: o significado. A gramática pode ser vista, então, como um conjunto de regras que transforma um conjunto infinito de unidades unilaterais (situações) em um conjunto infinito de unidades bilaterais (sentenças).

Para explicar a capacidade dos falantes de identificar o mesmo conteúdo semântico em sentenças diferentes e os conteúdos semânticos diferentes em uma mesma sentença — fenômenos muito comuns nas línguas naturais —, Shaumyan postula a hipótese sobre a divisão ("esfolhação") de cada língua natural em duas sublínguas: 1) a su-

blíngua primitiva, que serve para a notação do conteúdo semântico das sentenças, e 2) a sublíngua expressiva, para a qual são traduzidas as expressões da sublíngua primitiva (o termo "expressiva" é usado no sentido de "que possui meios variados para a expressão do mesmo conteúdo semântico"). A língua genotípica modela esta "esfolhação": a língua genotípica expressiva resulta da primitiva, ampliada com objetos novos. Estes objetos são introduzidos através das regras de tradução, na formulação das quais se usam os operadores lógicos especiais, chamados **combinadores**.

No conjunto de expressões aplicacionais do genótipo existem expressões com e sem combinadores. Algumas expressões sem combinadores podem ser reduzidas (por meio das regras que especificam a ação dos combinadores) às expressões chamadas formas normais, entre as quais se encontram algumas expressões correspondentes a sentenças reais. Estas formas normais são interpretadas como axiomas. A partir dos axiomas, por meio de regras, são derivadas as expressões que correspondem a sentenças da língua natural.

Como foi visto, na língua genotípica existem dois tipos de operandos primitivos: termos e sentenças. Entretanto, a língua genotípica possui vários outros tipos de expressões, cada tipo podendo ser definido em termos dos dois tipos primitivos. A estrutura na base de tipos opõe a língua genotípica (junto com as línguas naturais e as linguagens de programação de alto nível) a outras línguas artificiais baseadas no cálculo proposicional clássico, que são línguas livres de tipos.

Shaumyan entende que uma teoria lingüística, que pretende ser **lingüística** e não apenas **sintática**, não pode separar a sintaxe da semântica. Assim, a tarefa do lingüista será a construção de modelos de sistemas semióticos sintático-semânticos. Nestes sistemas, os níveis sintático e semântico terão "pesos" desiguais, que serão determinados sob dois pontos de vista: o lógico e o comunicativo. Sob o ponto de vista lógico, o nível sintático deve ser considerado o principal, e o nível semântico, o dependente, visto que só o nível sintático pode servir de base formal objetiva para o estudo do significado das unidades lingüísticas. Sob o ponto de vista comunicativo, o nível semântico deve ser considerado o principal, porque a língua desempenha suas funções de instrumento de comunicação através dos aspectos semânticos. Garantida uma correlação rigorosa entre os níveis sintático e semântico, os cálculos abstratos dos sistemas sintático-semânticos poderão servir de instrumento seguro para o estudo das funções comunicativas da língua natural, e a língua genotípica universal, construída por Shaumyan, é precisamente um sistema sintático-semântico.

A continuação da gramática genotípica é a gramática fenotípica, que formula regras

de correspondência entre a língua genotípica e as línguas naturais.

A língua genotípica é uma língua com a ordem fixa de operandos em relação a seus operadores e em relação aos próprios operandos. Esta ordem fixa é conservada no componente semântico da gramática fenotípica aplicacional e apenas o componente morfológico transforma esta ordem em ordem específica da língua natural em estudo. Portanto, a ordem fixa dos operandos serve de uma espécie de padrão em relação à ordem de palavras (constituintes) em línguas naturais.

A Gramática Aplicacional pode parecer complicada, mas, em realidade, ela é muito simples. Esta afirmação soa paradoxalmente. Entretanto, a verdade é que quanto mais simples são, do ponto de vista lógico, os elementos "de partida", mais longa se torna a cadeia dedutiva que une estes elementos fundamentais abstratos com a base empírica.

¹ Sebastian K. Shaumyan é ex-chefe do Departamento de Lingüística Estrutural do Instituto de Língua Russa da Academia de Ciência da URSS. Atualmente, professor do Departamento de Lingüística da Yale University, Connecticut, USA.

BIBLIOGRAFIA

- CURRY, H.B. 1963. *Foundations of Mathematical Logic*. New York.
- SHAUMYAN, S.K. 1965. *Strukturnaja lingvistika* (Lingüística estrutural). Nauka, Moskva.
- SHAUMYAN, S.K. 1974. *Applikativnaja grammatika kak semantičeskaja teorija jestestvennyx jazykov* (A gramática aplicacional como teoria semântica das línguas naturais). Nauka, Moskva.
- SHAUMYAN, S.K. 1984. *Semiotic Laws in Linguistics and Natural Science*. In: Copeland, J. *New Directions in Linguistics and Semiotics*. Rice University Studies, Houston Texas, 231-57.