

3

Pressupostos Teóricos

3.1

Os princípios teóricos de Harris

3.1.1

A aceitabilidade como fonte do conhecimento sintático

A sustentação metodológica de referência para o desenvolvimento desta pesquisa é a da gramática transformacional de Z. Harris (1964, 1961 e 1967 *apud* Vale 2001), que desenvolveu uma teoria lingüística cuja orientação é a aceitabilidade de frases como fonte básica para o conhecimento lingüístico. A gramática transformacional procura representar frases diretamente observáveis, que podem ser julgadas como aceitáveis ou inaceitáveis pelos falantes da língua. Na gramática transformacional de Z. Harris, o objeto central da sintaxe são as relações entre frases. Segundo Z.S. Harris (1952) *apud* Gross 1991:10 propôs para a descrição das variações de sentença um modelo baseado na noção de transformações. As transformações de uma sentença para outra são relações de equivalência que não afetam, ou seja, mantêm invariável o significado básico da sentença: as regras de [passiva], [introdução de modal] e [introdução de negação], entre outras, são exemplos de possibilidades de transformações.

Retomando Laporte (2000:69), para Harris a aceitabilidade em

(1) O presidente aumentou os salários

e a inaceitabilidade de

*o aumentou presidente os salários

são constatadas por julgamentos diretos, embora subjetivos, de um falante da língua. A existência de uma relação de sinonímia entre (1) e a frase seguinte é outro exemplo de fato diretamente observável:

(2) Os salários foram aumentados pelo presidente.

3.1.2

As transformações como elemento central da descrição sintática

A metodologia empiricista de Harris evita a criação e a manipulação de construções abstratas e complexas, bem como de regras que não sejam estritamente necessárias para descrever ou formalizar os fatos observáveis ou para simplificar essa formalização. A limitação à mera descrição combinatória da língua possibilita a construção de gramáticas coerentes. Essa metodologia não é aplicável a todos os campos da Lingüística: por exemplo, no caso da diacronia, a língua não tem forma diretamente observável. Porém, quando é aplicável, orienta o lingüista à procura de um apoio formal às intuições (Laporte, 2000:70).

Um dos recursos dessa metodologia utilizado pelo método de descrição do léxico-gramática Gross (1975) é a noção harrissiana de transformação sintática Harris (1964 e 1968). Existem dois tipos principais de transformações:

- as transformações unárias – que estabelecem a relação de sentido entre as frases conservando o sentido ou mostrando a diferença lexical;
- as transformações binárias, como a coordenação e a subordinação de frases, que combinam duas estruturas em outra estrutura.

A passiva e as pronominalizações são transformações unárias bem conhecidas. Conservam o sentido das frases às quais se aplicam, portanto permitem a constituição de classes de equivalência semântica, como nos exemplos seguintes, descritos por Laporte. O símbolo de “=” marca o fato de que duas frases são ligadas por uma transformação sintática.

(1) O presidente aumentou os salários

= Aumentou os salários (*apagamento do sujeito*).

= O presidente aumentou-os (*clitização*).

= Aumentou-os (*ambos*).

(2) Os salários foram aumentados pelo presidente (*apassivação*)

= Foram aumentados pelo presidente (*apagamento do sujeito*).

= Os salários foram aumentados por ele (*pronominalização do sujeito*).

= Foram aumentados por ele (*apagamento do sujeito e pronominalização*).

A última frase da acima resulta primeiro da aplicação da passiva e depois de duas pronominalizações: uma pronominalização do sujeito e uma do complemento em *por*. Muitas outras transformações podem ser descritas, como, por exemplo:

(1) O presidente aumentou os salários.

= (3) Os salários aumentaram

As transformações têm um caráter regular, isto é, aplicam-se de forma idêntica a numerosas frases. Assim, a relação entre (1) e (3) é observada também entre (4) e (5):

(4) João apagou a luz.

= (5) A luz apagou.

Todavia, essa regularidade está limitada pelas restrições de aplicação das transformações. Por exemplo, esse tipo de transformação não se aplica a (6), porque o resultado da aplicação é inaceitável:

(6) O João tirou o documento da gaveta.

*O documento tirou da gaveta.

As transformações binárias operam sobre duas estruturas. O resultado é uma estrutura complexa, por exemplo, em

uma coordenação:

(7) O presidente aumentou os salários, mas a medida foi cancelada;

uma subordinação adverbial:

(8) A luz apagou porque a lâmpada queimou;

uma construção com pronome relativo:

(9) O presidente aumentou os salários, que não mudavam há vários anos.

3.2

O método de descrição do léxico-gramática

As ferramentas metodológicas e teóricas utilizadas por Harris abriram um programa de pesquisa bastante amplo: o método de descrição do léxico-gramática definida por Gross (1975) nos últimos trinta anos, junto à equipe do Laboratoire

d'Automatique et Linguistique, considerando-se a necessidade de uma descrição lexical efetiva de grande porte por falantes nativos para atender aos objetivos da lingüística.

Segundo Vale (2001:68), o método de descrição do léxico-gramática é uma descendente direta da teoria transformacional de Zellig S. Harris. Embora a palavra "transformacional" possa evocar, em geral, a Gramática Gerativa - que, em algumas de suas várias formulações, foi chamada de Gramática Gerativa Transformacional - a abordagem de Maurice Gross é bem diversa. Com efeito, a abordagem do léxico-gramática é, antes de mais nada, um procedimento empírico. Maurice Gross critica na Gramática Gerativa o seu caráter especulativo: os seguidores de Chomsky raramente fazem apelo a dados reais. Assim M.Gross critica alguns princípios básicos da Gramática Gerativa, como construção de modelos:

... a gramática gerativa opõe a construção de modelos lógico-informático-matemáticos a uma abordagem descritiva, qualificada de procedimental. Vê-se assim serem construídos inúmeros modelos dotados de intenção de poderes preditivos e explicativos, mas em realidade construídos a partir de observações empíricas muito limitadas: são considerados como fatos lingüísticos apenas os fenômenos que permitem a confirmação ou o falseamento de um modelo existente. (...) Só se pode interpretar esse frenesi de construção de modelos como uma interpretação literal do célebre truismo 'a língua é um sistema onde tudo se encaixa'. Este 'axioma' parece legitimar o estudo de interações quaisquer entre fenômenos quaisquer, se é que fenômeno existe. A abordagem é tal que mesmo que os fatos sejam autênticos, eles são tomados ao acaso numa população de fenômenos cujo tamanho nunca foi estimulado (M.Gross, 1976:7-8 apud Vale, 2001:69).

Vale (2001) comenta a esse respeito que o argumento principal de Maurice Gross é o fato de que esses modelos da Gramática Gerativa dão conta apenas dos poucos exemplos que examina, sem levar em conta qual é sua real produtividade na língua. Assim, fatos que têm pouca produtividade são assimilados a outros cuja produtividade é grande e que o princípio segundo o qual a unidade de significado é a frase simples acarreta uma diferença metodológica fundamental: ao invés do método hipotético-dedutivo preconizado pela Gramática Gerativa, o Léxico-Gramática assume claramente uma postura taxonômica. Ou seja, recorrendo às palavras de M.Gross (1976:9), o exame sistemático do léxico constitui um meio, certamente o único no momento atual, de apreender uma língua de maneira global, isto é, de construir uma imagem da língua que tenha uma característica de

generalidade. Somente num quadro desse tipo é possível detectar os fenômenos massivos, opondo-os eventualmente aos exemplos marginais ou exceções".

Desse modo, segundo Silbertzein (1993) *apud* Vale (2001), a taxonomia torna-se um meio para se encontrar as regularidades da língua. Essa taxonomia tem também uma aplicação direta no tratamento automatizado da linguagem. E esses são os objetivos principais do método de descrição do léxico-gramática, que tem como meta investigar os procedimentos lexicais e gramaticais que levam ao reconhecimento de padrões de palavras nos quais está baseado o processo de entendimento para representação das propriedades de seqüências lingüísticas computacionalmente. Por essa razão, o método de descrição do léxico-gramática é o mais adequado e coerente com os objetivos desta pesquisa: descrever as palavras compostas com estrutura NdeN para elaboração de dicionário eletrônico.

O método de descrição do léxico-gramática tem por base o seguinte axioma: a sentença é a unidade lingüística de significado. Essa opção teórica é resultante de dois fatores: primeiro, o estudo de uma palavra isolada priva o descritor da possibilidade de avaliar aceitabilidades, já que o julgamento de aceitabilidade se aplica a frases; segundo, numa frase elementar, o contexto tira muitas vezes a ambigüidade de palavra isolada. Como consequência desse axioma, pode-se afirmar que as palavras não constituem unidades de significado pelas seguintes justificativas: as palavras compostas não adquirem significado mediante as regras gramaticais de combinação de palavras que são aplicáveis às construções produtivas; as palavras simples que constituem as sentenças cristalizadas ou idiomáticas não têm significado próprio. É com base nesse princípio que a separação dos significados é feita a partir de critérios sintáticos baseados na noção de frase elementar constituída por um sujeito, verbo e seus complementos como unidade mínima de sentido. As variações da estrutura sintática básica são definidas pelas propriedades distribucionais e transformacionais: apassivação, redução, pronominalização, coordenação etc. Isso significa que o método de descrição do léxico-gramática lida com procedimentos empíricos para encontrar regularidades da língua. O nível de representação adotado por Gross é, portanto, o nível de representação de Z.S. Harris.

O método de fazer estimativas de aceitabilidade é padrão. A aceitabilidade ou inaceitabilidade de uma seqüência é constatada por julgamentos de falantes

nativos, levando-se em conta o uso na língua. De acordo com M.Gross (1975:225-228), as estimativas feitas com seqüências de palavras deverão ser binárias: as seqüências são aceitas ou não. Na gramática, utilizam-se dois valores: aceitável ou inaceitável. Há algumas razões em favor deste ponto de vista:

- a natureza formal do método de descrição do léxico gramática cujo objetivo é a construção de um sistema característico (uma gramática formal, um autômato finito ou uma função característica);

- a hipótese de que uma primeira abordagem da descrição não deverá incluir dados sociolingüísticos nas representações gramaticais. Assim sendo, não se faz nenhuma referência a noções como, por exemplo, idioleto, gramáticas múltiplas para um falante ou gramáticas distintas para grupos distintos de falantes. Também não se fazem distinções estilísticas como, por exemplo, linguagem culta e gíria.

A distinção entre dados sincrônicos e diacrônicos é aceita, mas somente do ponto de vista operacional das estimativas de aceitabilidade, e não por razões *a priori* teóricas. No caso de muitas seqüências de palavras, há dificuldades para se determinar o grau de aceitabilidade. Alguns procedimentos indiretos utilizados nas estimativas de aceitabilidade constituem padrões de análise. Se, por exemplo, mediante o método adotado geramos uma estrutura dúbia, podemos optar por considerá-la aceitável. A distinção entre aceitabilidade e gramaticalidade pode ajudar, porque é possível construir estruturas que são claramente inaceitáveis, embora sejam gramaticais.

O método de fazer estimativas de aceitabilidade revela, também, por meio das transformações sintáticas a produtividade de uma regra de combinação de palavras, ou seja, pode-se variar os componentes, mas as propriedades se conservam. A observação em relação à reprodutibilidade de uma seqüência é importante para descrição e elaboração de dicionário, porque nos oferece dados para estabelecer critérios para a classificação das palavras.. A aceitabilidade de uma seqüência lingüística só pode ser avaliada por um falante da língua. Em geral, precisa-se de vários falantes da língua para avaliar a reprodutibilidade desse tipo de informação.

Observando-se os exemplos abaixo, *parado* e *organizado*, demonstram bem isso, porque podem ser classificados como participio do verbo ou como

adjetivo, pois apresentam traços que nos permitem afirmar que temos voz ativa e passiva. Portanto, pode-se estabelecer critério de identificação de verbos e considerar *parado* como particípio.

25) O ônibus está *parado*.

26) O ônibus foi *parado*.

27) O ônibus foi *parado* pelo motorista.

28) O ônibus foi *parado* pelo guarda.

Já os exemplos abaixo não se encaixam no critério anterior, pois *parado*, nesse caso, não pode ser passivo.

29) João é muito *parado*.

30) João está muito *parado*.

O reconhecimento dos diversos sentidos e a aceitabilidade do uso de *parado* só podem ser avaliados, por meio das relações sintático-semânticas observadas a partir da elaboração de frases e por quem já conhece a língua.

Outra questão relevante é não entremear as estimativas com fatores estilísticos, semânticos e situacionais que não sejam diretamente relevantes à aceitabilidade sintática da sentença. Contudo, quando um linguísta se vê diante de uma difícil decisão binária, ou seja, ao tentar avaliar uma sentença dúbia, ele deverá considerá-la como aceitável. Assim sendo, as abstrações se apóiam nos dados e não na teoria.

O método de descrição do léxico-gramática propõe critérios de aceitabilidade e reprodutibilidade metodológica para análise das palavras, a partir de frases indicando os seguintes procedimentos: a separação das acepções, a listagem das frases simples, as transformações unárias e as transformações binárias, como se pode observar a seguir.

A separação das acepções

A separação das acepções pode ser observada por meio da distribuição sintática dos itens, dependendo do sujeito, do verbo e dos seus complementos, como se pode observar nos exemplos a seguir.

- 1) É proibido colher flores.

V

Use a sua colher.

N

- 2) Os alunos decoraram o soneto.

V

As vendedoras decoraram a vitrine.

V

A listagem das frases

Quando se faz a constituição das frases, há uma escolha de padrões (que até podem ser arbitrários) e o reconhecimento das frases simples pode depender da existência de transformações, como mostram os seguintes exemplos:

- 3) Pedro partiu o vidro em pedaços .

- 4) Pedro partiu para Paris.

As transformações unárias

As transformações unárias operam com frases simples que se caracterizam por sujeito, verbo e complementos. Essas transformações vão permitir descrever todas as variantes das frases simples, porque a partir de uma frase simples outra frase pode ser criada, como nos exemplos a seguir:

- 5) Pedro partiu o vidro em pedaços.

- 6) O vidro foi partido em pedaços por Pedro.

7) O vidro foi partido por ele.

8) O vidro partiu.

A aplicabilidade das transformações unárias é possível ou não, dependendo da frase simples em que se tenta aplicar, como mostram os exemplos a seguir.

9) Essas medidas oferecem uma solução adequada a esses problemas.

10) Uma solução adequada é oferecida por essas medidas a esses problemas.

11)*Essas medidas fazem a oferta de uma solução adequada a esses problemas.

12)*Uma solução adequada é feita a oferta por essas medidas a esses problemas:

13) Pedro espera o fim da chuva.

14) Pedro está à espera do fim da chuva.

15) *Pedro faz a espera do fim da chuva.

As transformações binárias

As transformações binárias operam sobre duas estruturas. O resultado é uma estrutura complexa que é formada pelos elementos de frases simples e outros complementos que podem ser subtraídos, isto é, as sentenças complexas podem ser descritas como seqüências obtidas a partir de duas sentenças simples, mediante regras de composição denominadas transformações binárias (Z.S. Harris, 1952 *apud* Gross, 1991).

Coordenação

16) Pedro recomendou quatro livros à Maria.

17) Maria leu dois deles.

18) Pedro recomendou quatro livros à Maria e ela leu dois deles.

Subordinação

19) Maria leu dois livros porque Pedro os recomendou.

Pode haver frase complexa que não seja obtida por frase binária, como o exemplo a seguir:

20) Pedro recomendou que Maria lesse dois livros.

Formação de relativa

A transformação de relativização combina duas sentenças como demonstram os exemplos abaixo:

21) Pedro fez barulho. Esse barulho foi ouvido na rua.

22) O barulho que Pedro fez foi ouvido na rua.

A redução do Vsup é uma transformação entre a frase (22) e a frase (23), o que se escreve formalmente como:

23) O barulho de Pedro foi ouvido na rua.

[Red. V sup) = O barulho do Pedro foi ouvido na rua.

V sup =: fazer

O verbo suporte é semanticamente esvaziado, porque o sentido está centralizado na palavra predicativa "barulho" e não no verbo. Segundo Neves (1999), os verbos suportes apresentam um certo grau de esvaziamento do sentido lexical, mas conservam uma acepção cuja contribuição para o significado total da coligação pode ser explicitada.

Ancorado nos princípios da gramática tradicional, estrutural e transformacional, Gross introduziu a dimensão lexical na sintaxe, o que quer dizer que essa proposta engloba tanto um estudo apurado do léxico quanto das regras da gramática. De acordo com Gross, há muitos métodos adequados destinados à construção de gramáticas formais. Entretanto, ainda, não há uma gramática formal

completa, porque existem conceitos lingüísticos não-operacionais.

3.3

Descrições no âmbito da teoria do léxico-gramática

Atualmente, estão sendo construídos métodos léxico-gramáticas para várias línguas. Para o francês, o léxico dos verbos já está mais ou menos completo. Já foram classificados aproximadamente 10.000 verbos de acordo com um sistema de aproximadamente cinquenta categorias, que foram definidas segundo a seqüência de objetos dos verbos. No caso dos verbos, as sentenças constituem unidades básicas de significado: não se leva em conta um verbo sem também levar em conta um sujeito e além dos possíveis objetos a eles atrelados.

Conforme Gross (1986:31), também já foi feito um estudo lexical dos advérbios compostos do francês, bem como o inventário sistemático a partir de vários dicionários. Por meio desse estudo, foi possível observar que, enquanto nos dicionários atuais há aproximadamente 1.500 advérbios constituídos de uma única palavra, a maioria terminados em “ment” (= mente), correspondentes aos advérbios ingleses terminados em “ly”, foram encontrados mais de 5.000 advérbios compostos. Os advérbios compostos são classificados de acordo com a sua estrutura sintática. As formas sintáticas são descritas no nível elementar de seqüências de classe (categorias) gramaticais. São utilizados símbolos de interpretação, por exemplo: *Prep*, *Det*, *Adj*, *N* (= substantivo), *V*, *Conj* e *W* (= variável para os complementos do verbo).

Há pelo menos um terceiro componente do léxico-gramática do francês (assim como de outras línguas): os chamados verbos de suportes (*Vsup*) e verbos operadores (*Vop*) para os quais também há estudos para descritivos já realizados (M. Gross 1981).

Atualmente, também, estão sendo descritas algumas propriedades lingüísticas de terminologia técnica que conduzem à formulação de um algoritmo bastante produtivo e não-atrelado a nenhuma área específica de conhecimento. Esse algoritmo refere-se especificamente a termos técnicos compostos de várias palavras. O vocabulário técnico, que chega a conter vários milhões de palavras, é constituído de substantivos compostos. Essas observações quantitativas foram

confirmadas no estudo das línguas francesa, italiana, espanhola, inglesa e portuguesa.

Os laboratórios e grupos de pesquisas que empreenderam a construção de um método de descrição léxico-gramática constituem uma rede chamada RELEX. A maior parte deles são europeus. Dentre outras, o português europeu (Eleutério *et al.*, 1995; Ranchhod *et al.*, 1999) e o inglês (Marchonis, 1988) são duas línguas que estão sendo descritas neste quadro metodológico e teórico.