

Maria Carlota Rosa

Introdução à morfologia



editora**contexto**

5

Quantas palavras temos num enunciado?

5.1. Introdução

Até aqui usamos o termo *palavra* num sentido pré-teórico, isto é, sem defini-lo, enfim, sem enquadrá-lo no âmbito de uma teoria. Como *palavra* é um termo que existe no uso cotidiano da língua, pode parecer, a princípio, uma tarefa muito simples defini-lo. Não é bem assim.

De acordo com o uso comum do termo, que tem por base nosso conhecimento da escrita, parece trivial definir o que seja uma palavra. Qualquer criança no ensino fundamental já sabe responder quantas e quais são as palavras em *Penélope ama Odisseu*. Talvez já fique em dúvida perante *ama* e *amava*, não sabendo se as contará como uma ou como duas palavras, quando, então, a pergunta parece levar em conta algo que ultrapassa uma determinada sequência de letras ou símbolos. Ou ainda se contará também como palavras elementos como *que*, *se*, *de*, quando entra em jogo o tipo de significado das formas.

Deixada de lado a escrita, porém, tem a palavra, afinal de contas, algum tipo de relevância teórica que a torne interessante para a pesquisa gramatical, ou, ao contrário, poderia ser considerada apenas um epifenômeno, isto é, um fenômeno secundário?

A resposta parece ser sim para a primeira parte da pergunta. A despeito do contínuo sonoro existente na oralidade, os falantes conseguem abstrair parte das características físicas de um enunciado e desenvolver, a partir da mais tenra infância, algum tipo de estratégia que lhes permite segmentar esse contínuo em unidades menores, com base no ritmo do *input* nativo, seja ele acentual, silábico ou moraico.

Pesquisas com línguas diferentes têm demonstrado que há estratégias que não são universais, mas dependentes das características rítmicas específicas da primeira língua, ou da língua dominante no caso de bilíngues, para a apreensão dessa unidade a que estamos denominando intuitivamente *palavra* (Cutler, 1994). Ao ouvirmos o contínuo sonoro, lançamos mão dessa estratégia, e ela é surpreendentemente eficaz mesmo quando os sinais sonoros da fala vêm misturados com outros sons do ambiente, ou pronunciados com sotaque estrangeiro, ou a uma boa distância do ouvinte.

Falantes do inglês como primeira língua, por exemplo, segmentam o enunciado com base no ritmo acentual entre sílabas fortes e fracas (Cutler, *id. et ibid.*). Uma vez que a maior parte das palavras em inglês tem sílaba tônica inicial, os erros de

percepção são mais comuns se produzem a inserção de uma fronteira de palavra diante de uma sílaba forte (como em *achieve* ouvido como *a cheap*) ou o apagamento de uma fronteira de palavra diante de uma sílaba átona (como no caso de *bird in* ouvido como *burgling*). São menos usuais quando produzem a inserção de uma fronteira diante de uma sílaba átona (como em *effective* ouvido como *effect of*) ou o apagamento de uma fronteira diante de uma sílaba tônica (como *were waiting* sendo ouvido como *awaiting*). Por seu turno, falantes de francês usam como estratégia de segmentação o reconhecimento *silaba a silaba*¹, ao passo que falantes de japonês usam como estratégia a *mora* (Cutler, 1994: 92-93)².

Para a linguística, o grande problema em definir *palavra* é ser esse termo passível de receber diferentes caracterizações nas diferentes dimensões do estudo da linguagem, nem sempre resultantes na mesma unidade. Afora o uso na escrita, podemos entender palavra: (a) como uma unidade fonológica; (b) como o elemento mínimo da estrutura sintática; (c) como um elemento do vocabulário da língua. Nas secções que se seguem vamos focalizar essas diferentes interpretações. A começar pela mais usual. Vamo-nos deter um pouco nas justificativas para não levar em conta, aqui, a palavra gráfica.

5.2. A delimitação da palavra

5.2.1. A palavra gráfica

Utilizamos o termo *palavra*, cotidianamente, como uma noção oriunda da escrita, que se aplica também à fala. Por que motivo essa caracterização não será aqui privilegiada? Simples: porque consideramos a escrita como parte de um outro estudo, o dos recursos expressivos particulares que a representação gráfica torna disponíveis (Nunberg, 1990: 7). A escrita tem sua própria “gramática”, isto é, seu conjunto particular de regras.

Aprender a escrever é, em parte, aprender a identificar e a assinalar os limites da **palavra gráfica**, unidade delimitada por **separadores**, *i.e.*, por espaços em branco ou quebras de linha, mas também por sinais de pontuação ou ainda por letras de traçado diferenciado, consoante ocupem ou não a posição final na palavra escrita³.

Embora nas escritas modernas ocidentais que empregam o alfabeto latino a



¹ O ritmo do português do Brasil estaria em mudança: de um padrão silábico como o francês ou o espanhol, para o acentual, como o inglês (Major, 1981).

² Que pode ser uma parte da sílaba que não o *onset*.

³ Vimos exemplo deste último tipo nas primeiras aulas de Grego, quando aprendemos a distinguir o sigma final (ς) do não final (σ).

noção de *palavra gráfica* seja, no mais das vezes, equivalente a uma forma livre, ela é tratada por regras da escrita. Talvez onde melhor possamos percebê-lo, no caso do português moderno, seja na representação de compostos, de preposições e de clíticos. Com a ortografia atual para o português do Brasil por base, por exemplo, temos de considerar fatores como os hifens ao definir uma palavra composta: *pé de anjo* (‘aquele que tem pé grande’), por exemplo, é uma sequência composta por três palavras gráficas, mas *pé de cana* (‘aquele que bebe muito’) é apenas uma palavra, pelo menos até o aparecimento de alguma reforma ortográfica que venha a banir os hifens, ou que, ao contrário, inflacione seu uso, quando, então, o quadro passaria a ser outro. Nosso estudo de morfologia teria de incluir um capítulo sobre o uso de hifens? Seria estranho, uma vez que o uso de hifens é regulamentado, em última análise, por decreto do presidente da República...

Não podemos esquecer que nossa escrita resulta de cerca de quinhentos anos de padronização, que começou a ser efetivamente imposta a partir da invenção, no Ocidente, da imprensa de tipos móveis. A palavra gráfica, como a conhecemos, é um artifício relativamente recente, e escritas mais antigas nem sequer segmentaram o espaço do suporte⁴. A chamada *escrita bustrofedônica* dos gregos (século VI a.C.), por exemplo, fazia-se em linhas horizontais, alternadamente, da direita para a esquerda e vice-versa, algumas vezes de baixo para cima na superfície escrita (Diringer, s.d.: 148), sem qualquer separação entre palavras.

Num tipo de escrita diferente da nossa, a hieroglífica do antigo Egito (*ca.* 5000 a.C. a 100 d.C.), a fronteira gráfica que delimitava o final de uma palavra era marcada por um símbolo especial, o *determinativo*, que podia combinar-se (ou não) com até dois outros. Os 180 determinativos tinham por função classificar as palavras: o rolo de papiro selado () , por exemplo, indicava que aquela palavra pertencia à classe das ideias abstratas (Jacq, 1994: 27). Seria uma nova palavra o que viesse antes do determinativo, ou depois dele, dependendo da direção para a qual as figuras humanas ou de animais se voltavam no suporte. Para os nomes de reis a visualização era mais fácil: o *cartucho*, uma linha ovalada, envolvia o conjunto de símbolos que compunham a palavra, assim, , e sinalizava qual o conjunto a levar em conta em meio a uma sequência sem solução de continuidade.

Para as escritas alfabéticas, talvez possamos considerar os primeiros passos na direção da *palavra gráfica* o desenvolvimento de formatos diferenciados para alguns grafemas, o que viria a ajudar o processamento da leitura de um conjunto compacto de linhas e, nas inscrições monumentais romanas (séculos I e II), o uso de um ponto entre palavras.

Mesmo depois de os separadores começarem a ser empregados, seu uso não coincidiu de imediato com o das escritas modernas. Na Europa medieval, preposições

⁴ Dá-se o nome de **suporte** ao material em que um texto está inscrito. Esse material tem variado ao longo dos séculos: pedra, papiro, pergaminho, papel, vinil, disco magnético, película fotográfica.

e palavras curtas foram, em geral, unidas à palavra seguinte (Bischoff, 1986: 173). Mas, pelo menos até o início do século XVI, obras manuscritas e impressas estiveram repletas de formatos diferenciados para os grafemas, consoante sua posição na palavra, e de *conglomerados gráficos* inesperados pela óptica de um leitor moderno, os quais, por vezes, induziram até mesmo renomados filólogos a erros de leitura.



Um exemplo famoso de erro de leitura, causado por separadores utilizados de forma diferente daquela a que nos habituamos, originou-se de um verso da égloga *Crisfal* (Cristóvão Falcão, 15-): *cantou canto de ledino*. Teóphilo Braga (1875), no *Manual da história da literatura portuguesa*, interpretou esses desconhecidos *cantos de ledino* como '*cantos de romaria*', explicando que eram cantos alegres porque, nas romarias, as moças tinham ocasião de encontrar-se com seus amados. O texto de Braga influenciaria outros filólogos, como E. Monaci, que lhe dedicaria toda uma obra (*Cantos de ledino tratti dal grande Canzoniere portoghese della Biblioteca Vaticana*, 1875).

Carolina Michaëlis de Vasconcelos viria mais tarde a demonstrar que a leitura correta do verso seria: *cantou canto d'ele dino*. Dito de outra forma: *cantou um canto digno dele*. (Para mais detalhes, Roncaglia, 1974-5: 88-89).

Os separadores em textos antigos podiam, mesmo, refletir um estágio da língua que já então era passado: formas como *tal vez, por tanto* e advérbios em *-mente*, por exemplo, continuavam a ser grafadas como duas palavras gráficas tempos depois de já se terem combinado numa só. A título de ilustração, veja-se o excerto abaixo, extraído da *Estoria de muy noble Vespesiano emperador de Roma* (Anônimo, 1496. fol. c3^{r-v}), um dos romances do ciclo do Graal:

E pilatus & el rey archileus com dez caualleyros. se sobirō no muro sem armas. & vestidos de senhos briaes vermelhos [...] O nobre meu padre te encomẽdou esta çidade por que aguardarres & arregarres por elle [...] E de pois de sua morte enuiar te me o trebuto [...] E [...] derprezar teo muyto mal.

Nesse pequeno exemplo, os pronomes átonos ora estão, ora não, formando uma unidade gráfica com o verbo; o advérbio *depois* (< lat. *de post*) aparece como duas palavras gráficas, revelando a consciência do étimo latino. E que fazer quando nos deparamos com termos desconhecidos, como o distributivo⁵ *de senhos*?⁶ Observe-se ainda o traçado diferente para <s>, em conformidade com sua posição na palavra gráfica.

Em suma: a função básica da (orto)grafia é o registro da informação de modo permanente (*verba volant, scripta manent*, já dizia o antigo provérbio) e não a

⁵ Dá-se o nome **distributivo** a um item que refere individualmente cada elemento de um conjunto, como *cada, todo, nenhum*.

⁶ *De senhos* (< lat. *singuli*) significa 'cada um'.

transcrição sistemática e inequívoca de dados para o trabalho de um linguista⁷. Nem se imagine identificar os separadores com as pausas na fala. Quando falamos, não fazemos pausas a cada forma livre ou dependente⁸, mas entre sequências com extensão média de cinco a seis palavras (Chafe, 1992: 25). Deixemos então de lado a escrita e passemos à fala.

5.2.2. A palavra fonológica

O que é denominado **palavra fonológica** – unidade formada por fonemas, sílabas e traços suprasegmentais – pode ser menor do que aquilo que desejamos considerar uma palavra no estudo morfológico. É a um afixo, considerado ou não como uma palavra fonológica, que recorre a análise para distinguir, em italiano: (a) prefixos produtivos que, quando se agregam a uma palavra começada por /s/, não configuram ambiente para a sonorização da fricativa, como em (5.1)⁹; de (b) formações improdutivas que permitem a sonorização de /s/, como em (5.2). No primeiro caso temos duas palavras fonológicas¹⁰, uma que corresponde ao prefixo, outra, à base:

(5.1)	(ri) _p (suonare) _p (a) _p (sociale) _p (prè) _p (sentire) _p	[risuonare] [asoçiale] [presentire]	'soar novamente' 'não social' 'ouvir antecipadamente'
(5.2)	(presentire) _p	[prezentire]	'ter pressentimento'

De outro modo, em italiano, no interior de uma palavra fonológica, o ambiente intervocálico impede o aparecimento de /s/ não geminado:

(5.3)	(rosa) _p (cosa) _p	[roza] [coza]
-------	--	------------------

Em português também há exemplos em que podemos considerar a palavra fonológica menor que a forma livre. É o caso de derivados em *-mente*, *-issimo* e *-inho*, que podemos considerar como constituídos de duas palavras fonológicas. A postulação de uma fronteira entre palavras fonológicas explicaria o não fechamento da vogal média pretônica no derivado, correspondente à tônica no

⁷ Aliás, foi essa mesma conclusão que levou ao estabelecimento de alfabetos fonéticos.

⁸ Quando muito, os espaços em branco da escrita poderiam corresponder a *pausas potenciais*. Se enunciada bem lentamente, no entanto, uma frase poderia apresentar pausas também entre suas sílabas: *a/ bo/ ne/ ca/ de/ pa/ no es/ tá/ su/ ja*.

⁹ Exemplos extraídos de Spencer (1996: 178-179), cuja análise reporta-se a Nespor, M. & Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

¹⁰ _p indica 'palavra fonológica'.

termo derivante, uma exceção ao fato de, no dialeto carioca, as vogais médias abertas não ocorrerem em ambiente pretônico:

(5.4) *forma* ['foxma], mas *formoso* [foxmozu]

Do mesmo modo *clero*, com [ɛ] faz o derivado *clerical*, com [e]. A fronteira entre palavras fonológicas, no entanto, previne o fechamento da vogal média:

(5.5) (sɛri)_p (íssimo)_p (nova)_p (mente)_p (pɛ)_p (zinho)_p

Ao contrário desses exemplos do italiano e do português, há *palavras fonológicas* que não gostaríamos de considerar como uma única palavra num estudo morfológico. É o caso de formas fonologicamente dependentes que se agregam a outros elementos da sentença como se fossem sílabas iniciais ou finais. Uma sequência como *disse-lhe*, foneticamente ['dʒisiɫi] na fala carioca, serve para ilustrar o problema. Formas átonas como o pronome *lhe* do exemplo agregam-se fonologicamente a outra palavra do enunciado a eles contígua e recebem a denominação geral de **clíticos** (do gr. *klitikós*, 'que se inclina ou apoia' pelo lat. *cliticus*).

Um *clítico* é uma *forma dependente*. A classificação *forma dependente* foi criada por Mattoso Câmara Jr. como complemento à distinção bloomfieldiana entre *formas livres* e *formas presas*. O problema classificatório não era novo. Também Sapir constatara, ao tratar do paiute¹¹, que havia formas que nem eram verdadeira afixação, nem se constituíam na justaposição de elementos independentes¹².



O próprio Bloomfield (1926: 26) reconheceu a existência de formas de caráter misto, e classificou-as como *formativos de sintagma*:

Pressuposto S1. Um sintagma pode conter uma forma presa que não é parte de uma palavra. Por exemplo, o possessivo [z] em *the man I saw yesterday's daughter*. ['a filha do homem que eu vi ontem' – MCR]

Def. Essa forma presa é uma *formativo de sintagma*.¹³

¹¹ Língua uto-asteca do N de Nevada, Oregon, Califórnia e Idaho (EUA).

¹² "Enclisis is neither true suffixation nor juxtaposition of independent elements. It has the external characteristics of the former (including strict adherence to certain principles of order), the inner feeling of the latter." (Edward Sapir, 1930. *Southern Paiute, a Shoshonean language*. Apud Klavans, 1982: 1).

E completava: "Esse pressuposto perturba a definição de *sintagma* [...]"¹³ (Bloomfield, *id. et ibid.*).

A questão reapareceria em Nida (1949: 104), já então com a preocupação em justificar por que incluir o genitivo 's do inglês entre os afixos e não entre os clíticos¹⁴. Segundo Nida, os clíticos têm muito maior liberdade que afixos, e s ocorre apenas com nomes, pronomes e sintagmas nominais.

Os *formativos de sintagma* ou *afixos de sintagmas* formam, porém, um conjunto mais restrito que as *formas dependentes* de Câmara Jr., porque estas incluem artigos e preposições, na medida em que, na fala comum, pertencem ao mesmo grupo acentual do núcleo do sintagma.

Os *clíticos* prendem-se fonologicamente a outra palavra no enunciado, que é o seu **hospedeiro** (tradução do inglês *host*)¹⁵. No português europeu, por exemplo, os clíticos de acusativo "são sempre fonologicamente enclíticos, a despeito de qual seja a palavra precedente", o que os impede de estarem no início absoluto da sentença (Nunes, 1992: 5, citando Carvalho, 1989¹⁶). O sinal = nos exemplos abaixo indica a cliticização:

- (5.6) a. *Quem=me vê?*
 b. *Já=te digo*
 c. *Não=te vi*
 d. *João tinha=me dado um livro*
 e. **Me=diga uma coisa*

Afora os clíticos, é interessante observar ainda que a fala espontânea apresenta muitas vezes o "apagamento" de fronteiras entre palavras. Uma sequência como *as*

¹³ Bloomfield (1926: 27): "Uma forma livre não mínima é um sintagma. E.g., *the book*. or *The man beat the dog*; mas não, e.g., *book on* [...], porque é sem significado e, logo, não é uma forma; tampouco *blackbird*, que é uma forma livre mínima."

¹⁴ Nida (1949: 104): "The genitive -s may occur with single morphological units, as in John's, his, and everybody's, or it may occur with phrases, e.g. the old man's (aches and pains) and the king of England's (hat). This morpheme is considered a suffix rather than a clitic because its distribution is limited to occurrence with nouns and pronouns and because the resultant construction belongs to one of two external distribution classes: (1) attributive to nouns, e.g. the king of England's (hat) or (2) a substitute for nouns, e.g. the king of England's was there. [...] The distributional behavior of the genitive -s is quite different from that of the usual clitic, which may be combined with any number of classes of forms and with numerous resultant external distribution class memberships". Ainda no âmbito do estruturalismo norte-americano, a mesma questão seria outras vezes retomada, como, e.g., em Wells (1947: 196ss).

¹⁵ Na dependência da posição que ocupam em relação ao hospedeiro podem ser *proclíticos*, *mesoclíticos* ou *enclíticos*.

¹⁶ Carvalho, J., 1989. Phonological conditions on Portuguese clitic placement: on syntactic evidence for stress and rhythmical patterns. *Linguistics*. 29, 405-436.

amigas [ðza'migðʃ] é um contínuo sonoro, em que a sílaba travada¹⁷ do artigo desaparece em razão de uma vogal iniciar a palavra seguinte, transformando uma sequência VC#V em V-CV¹⁸. Se a *juntura*, por um lado, indica que não houve solução de continuidade na emissão da voz, por outro é ela um fenômeno típico de fronteira entre palavras em português, e que ajuda, por conseguinte, no reconhecimento dessa unidade. Os acentos (representados pelo sinal ') indicam a presença de um nome, verbo, adjetivo, advérbio, que constituem, em geral, unidades também no nível fonológico. Em algumas línguas, mais precisamente aquelas que apresentam o *acento culminativo*¹⁹, como o latim, o checo ou o polonês, a localização do acento é fixa, servindo de auxiliar na identificação de palavras²⁰.

Poremos de lado a *palavra fonológica* neste trabalho. A existência de clíticos, de juntura, de regras como a do italiano a que aludimos anteriormente faz com que essa unidade não coincida, necessariamente, com o que gostaríamos de considerar palavra num estudo morfológico. Em princípio queremos que as formas livres e as formas dependentes façam parte do que entendemos como palavra. Assim, sequências como *os meninos* ou *disse-lhe* serão aqui consideradas como compostas de duas palavras; o italiano *asociale*, por sua vez, como uma única.

Aliás, em estudo recente, Hanlon & Edmondson (1996) sugerem, com base em estudo sobre a afasia de jargão fonêmica²¹, que palavras fonológicas podem manter-se intactas apesar de o paciente não apresentar qualquer outra evidência de que essas unidades tenham significado ou classe gramatical, por exemplo²².

5.2.3. A palavra como unidade sintática mínima

Considerar a palavra como uma unidade mínima para a sintaxe significa dizer que a sintaxe não forma palavras. É o que a literatura gerativa denomina **hipótese lexicalista**. A sintaxe forma constituintes, frases, mas não palavras. Dito de outro modo: a estrutura interna da palavra não é da alçada da sintaxe, mas da morfologia.

¹⁷ Aquela que termina em consoante, representada como (C)(C)VC.

¹⁸ Em que C representa 'consoante' e V, 'vogal'; -, 'fronteira silábica', e #, 'fronteira de palavra'.

¹⁹ Opõe-se ao *acento livre* ou *distintivo* de línguas como o português, por exemplo.

²⁰ Em latim clássico o acento é determinado pela quantidade: em dissílabos acentua-se a penúltima sílaba. Em palavras de três ou mais sílabas, acentua-se a penúltima se esta for longa (ou por apresentar uma vogal longa, ou um ditongo, ou uma vogal breve seguida de duas ou mais consoantes); se for breve, acentua-se a antepenúltima.

²¹ Afasia em que o paciente produz sequências fluentes de fonemas, mas destituídas de qualquer significado. O termo *afasia* foi utilizado primeiramente por Platão, para designar o silêncio daquele que se calava diante de um argumento definitivo, e seria retomado por Armand Trousseau em 1865 com o sentido de défices linguísticos (Guindaste, 1996).

²² Um pequeno exemplo: como resposta à pergunta "Diga-me como dormiu a noite passada, OK?", a paciente responde: 'hæ?ɛ 'd,ɛlɔn 't'ræ?ŋ 't'ræk 'nɔzŋ.

Não é importante para a sintaxe que *reler* seja formado a partir de *ler* pela adjunção do prefixo *re-*²³. Daria no mesmo se estivéssemos diante da forma primitiva *ler*. Importa que *ler* (ou *reler*) seja verbo, uma vez que essa informação é relevante para fenômenos como concordância e regência. A hipótese lexicalista representa o reconhecimento de que as construções morfológicas são reguladas diferentemente da frase. Uma palavra e uma frase não são diferentes porque há mais unidades constituintes nesta que naquela, mas porque têm tipos de coesão interna diferentes.

Numa palavra os constituintes internos não têm mobilidade. A posição de um sufixo não pode ser preenchida por uma raiz; tampouco por um prefixo. Em línguas como o suaíli, que refletem pela concordância o Sujeito e o Objeto na estrutura do verbo, essa ordem interna não tem a mobilidade existente para S(ujeito), V(erbo) e O(bjeto) sintáticos: os formativos têm de apresentar-se na ordem [Sujeito – Tempo – Objeto -raiz]. Assim, a forma verbal que poderíamos traduzir como 'eu lhe darei' é aquela em (5.7):

(5.7) [ni - ta - m - pa]_v
1SG/S - FUT - 3SG/O - dar (raiz)

As unidades da sintaxe têm mobilidade incomparavelmente maior. Numa das chamadas *línguas de ordem livre*, como o latim, por exemplo, o número de posições possíveis para os elementos sintáticos é grande. Uma frase como 'Pedro vê Paulo' pode ser traduzida como (a), ou, por questões de ênfase, como (b), (c) ou (d) abaixo (baseado em Moreland & Fleischer, 1977: 5):

- (5.8) a. Petrus Paulum videt (ordem neutra).
b. Petrus videt Paulum (ênfase no sujeito e no objeto).
c. Paulum Petrus videt ('É Paulo que Pedro vê').
d. Videt Paulum Petrus ('Pedro realmente vê Paulo').

Uma palavra tem expansão limitada e regulada a partir de certo ponto, ao contrário da possibilidade de expansão infinita dos grupos de palavras. Por exemplo, as regras do português que formam derivados em *-al*, *-iz(a(r))*, *-ção* podem aplicar-se em sequência. Assim, podemos derivar de *instituir*, *instituição*, e, sucessivamente, em camadas, *institucional*, *institucionalizar*, *institucionalização*, ??? *institucionalizacional*. A expansão é possível, mas, a partir de determinado ponto,

²³ Os hifens que representam a posição de uma forma presa em relação aos outros constituintes da palavra. Por exemplo: *des-*, *-ção*, *-duz-* indicam, respectivamente: (a) que a forma *des-* precede outra forma; (b) que *-ção* se segue a algo; (c) que *-duz-* (como em *deduzir*) necessita de um elemento anteposto a ele e de outro, que o siga.

as derivações param. A expansão de uma unidade sintática, por sua vez, é, teoricamente, ilimitada: *o meu grande amigo, o meu grande e querido amigo, o meu grande, querido... e inesquecível amigo*.



A hipótese lexicalista veio substituir a chamada **hipótese transformaciona- lista** da gramática gerativa, exemplarmente ilustrada em Lees (*The grammar of English nominalization*, 1960). A hipótese transformacionalista tratava processos gerais de formação de palavras, como a formação de nomes a partir de verbos como um processo sintático, as *nominalizações*: a mesma estrutura sintática geraria *João não quer que Pedro participe da reunião* e *João quer a participação de Pedro na reunião* (exemplos em Basílio, 1980: 26). A diferença entre as duas sentenças resultaria da aplicação de uma regra sintática (a *transformação*) de nominalização no caso da segunda.

Qual a vantagem que essa análise apresentava? A vantagem estava em captar generalizações, como a relação semântica existente entre um verbo e seu derivado nominal²⁴.

Se, no entanto, a estrutura interna da palavra é opaca para a sintaxe, que dizer dos expoentes para as *categorias morfossintáticas*? O Caso e o Número em latim, por exemplo, fazem parte da morfologia dos nomes e são relevantes para a sintaxe. Não há como formar uma oração correta em latim sem esse tipo de informação. O que nos leva a ver essas marcas de Caso e de Número como um tipo de morfologia diferente daquele que temos em formas como *-dade*. Podemos entender que a formação de palavras está em interação com o léxico; a flexão, por sua vez, representa a interação entre a morfologia e a sintaxe, ou morfossintaxe. Por conseguinte, nossa asserção acerca de a sintaxe ser cega para a estrutura da palavra, para não ser facilmente falsificada, necessita de um “ajuste” que distinga os processos flexionais daqueles que formam vocabulário na língua. Se tratamos a flexão e a derivação como tipos distintos de morfologia, podemos considerar seus elementos como entidades de caráter também diverso. Foi o que começamos a apresentar no capítulo 3, ao introduzirmos conceitos como *formativos* e *expoentes*. Os capítulos subsequentes voltarão a essas questões.

5.2.4. A palavra como unidade da morfologia

Como já ficou evidente, o termo *palavra* é ambíguo. Mas e se ficarmos restritos apenas ao seu uso na morfologia? Essa ambiguidade persistiria? A resposta é ‘sim’. Vamos distinguir aqui diferentes significados que o termo *palavra* pode

²⁴ “Como princípios metodológicos” – explica Dillinger (1991: 58) – “preferem-se as generalizações e a eliminação de idiossincrasias. Consequentemente, uma das questões centrais da teoria linguística é: até que ponto informações podem ser eliminadas do léxico sem sacrificar a adequação descritiva?”.

apresentar num estudo morfológico: como sinônimo de *lexema*; como sinônimo de *forma de palavra*; como sinônimo de *palavra morfossintática* ou *gramatical* (vide Matthews, 1972; 1974, 1991).

5.2.4.1. A forma de palavra

A **forma de palavra** é composta de uma sequência sonora. A *forma de palavra* é o uso mais aproximado dos conceitos estruturalistas de *forma livre* (Bloomfield, 1926: 27) e de *forma dependente* (Câmara Jr., 1970: 60).

Fazemos referência à forma de palavra quando dizemos estar diante de palavras homófonas: no português padrão do Brasil²⁵ *amamos* pode representar a *primeira pessoa do plural do presente do indicativo* de AMAR, mas também a *primeira pessoa do plural pretérito perfeito do indicativo* de AMAR.

5.2.4.2. O lexema

O **lexema** é uma palavra considerada como unidade abstrata. Tem significado lexical e pode apresentar variações, caso se inclua entre as *palavras variáveis*. O lexema pertence a uma das classes abertas da língua. Seguimos aqui a convenção de representá-lo com todas as letras maiúsculas.

Em geral, por motivos didáticos, traça-se um paralelo entre um **lexema** (parte do conhecimento de um indivíduo acerca de sua língua) e uma **forma de citação** (unidade de um dicionário em papel, ou na tela de um computador). Quando, na linguagem do dia a dia, referimo-nos a um verbo e não a uma de suas formas específicas, empregamos uma *forma de citação*: procuramos no dicionário o verbo *amar*, ou o verbo *vender*, ou *partir*. Se fosse um verbo latino, a consulta buscaria *amo*, *vendo*, *partior*, por exemplo. Com as formas de citação não fazemos referência somente ao infinitivo não flexionado português ou à primeira pessoa do presente do indicativo ativo latino, mas a todos os tempos, modos e pessoas desses verbos. A forma de citação *amar*, por exemplo, é virtualmente um conjunto de formas e nenhuma delas. O lexema é uma abstração nesse sentido: o lexema AMAR, OU SABER, por exemplo, representa a combinação virtual dos radicais que pode apresentar com todas as propriedades morfossintáticas com que se pode combinar.

5.2.4.3. A palavra morfossintática

E se quisermos falar não do verbo AMAR em geral, mas de um dos membros que formam seu paradigma? Por exemplo: se quisermos focalizar apenas a *primeira pessoa do singular do presente do indicativo*? Estaremos, então, com o mesmo

²⁵ No português de Portugal, a abertura da vogal tônica distingue as duas formas. Já no português não padrão do Brasil, o presente *amamos* distingue-se do pretérito *amemos*.

AMAR, mas agora acompanhado de uma única entre as possibilidades de realização das categorias gramaticais ou morfossintáticas Tempo/Modo/Aspecto e Número/Pessoa. A **palavra gramatical** ou **morfossintática** é o lexema (AMAR, por exemplo) mais determinadas propriedades morfossintáticas, como IND./PRES./1SG.

Nos próximos capítulos nossa atenção estará voltada para o lexema e para a palavra morfossintática. E também para um conceito relevante para os estudos morfológicos: o *léxico*.

Súmula

Os problemas com a análise morfêmica clássica são variados e derivam, basicamente, do modelo de análise adotado, IA. Os problemas são interdependentes e alcançam até mesmo, como aponta Basílio (1974a: 82),

a não abrangência da metodologia, se considerarmos como morfes apenas os elementos aos quais podemos atribuir significado; ou se resumem na não adequação da metodologia às definições básicas de que ela partiu, se considerarmos também como morfes os elementos mínimos que restam após a segmentação em morfes.

Ao se desmembrar, no estudo da morfologia, significado e expressão, abria-se caminho para o retorno da palavra ao cenário, mas não mais como uma forma (*i.e.*, “um traço vocal recorrente que tem significado” – Bloomfield, 1926: 27) livre ou dependente. Por que não? Porque, ao tratar a palavra como uma unidade que se concretiza num enunciado, ela foi definida por um critério de pouco interesse para os estudos morfológicos: se pode ou não ser enunciada sozinha. Para o estudo morfológico, as características gramaticais são mais interessantes do que essa.

Em suma: a palavra é “por definição, uma abstração gramatical” e “qualquer critério fonológico deve permanecer logicamente secundário” (Robins, 1959: 120). O *lexema* e a *palavra morfossintática* são os elementos que nos interessam aqui. Com eles podemos dar conta das relações existentes no vocabulário da língua como também da informação gramatical que carregam.

III

REVISITANDO AS PARTES DO DISCURSO

PALAVRAS-CHAVES

morfologia e léxico;
flexão e derivação;
item lexical;
Regra de Formação de Palavras e produtividade;
bloqueio;
classe gramatical/parte do discurso;
significado lexical/significado gramatical;
classes abertas/classes fechadas;
categorias e propriedades.

6

A formação de vocabulário: o lexema

6.1. Introdução

Quando dizemos que sabemos uma palavra de nossa língua, levamos em conta uma enorme quantidade de informações. Uma das tarefas da linguística é dar conta desse conhecimento, e essa tarefa não é nada fácil.

Um modo de representar o conhecimento que um falante tem da sua língua é propor que a morfologia interage com todos os outros níveis de representação. Uma parte (ou subcomponente) da morfologia, aquela que trata da derivação e da composição, interage com o léxico; outra, aquela que trata da combinação de radicais e propriedades morfossintáticas, interage com a sintaxe; e ainda a parte que focaliza a forma de palavra que resulta da flexão interage com a fonologia (*vide* Spencer, 1991: 422-459 – para um panorama acerca das diferentes visões). No que se segue vamos centrar nossa atenção no léxico.

6.2. O léxico¹

A visão mais tradicional do léxico foi exposta na clássica obra de Bloomfield (1933: 269):

Uma descrição completa de uma língua listará toda forma cuja função não é determinada seja por estrutura, seja por um marcador; incluirá, conseqüentemente, um *léxico*, ou lista de morfemas, que indica a classe de formas de cada morfema bem como listas de todas as formas complexas cuja função seja de algum modo irregular.

¹ Procedemos aqui a uma grande simplificação. O que efetivamente consta do léxico varia bastante mesmo no campo de uma teoria. Para muitos autores na perspectiva gerativa, parte da morfologia está no léxico. Para Anderson (1992), por exemplo, o léxico representa também uma forma de conhecimento, e, por conseguinte, vai além da noção de lista. Assim, além do conhecimento da lista – que inclui minimamente a existência do item lexical, por exemplo *CAMA*, sua descrição fonológica [kama], seu significado, algo como ‘móvel que serve para as pessoas se deitarem’, a característica sintática de ser um nome -, o léxico inclui também “um sistema de regras, que são os modos de relacionar entre si as palavras até o ponto em que essas relações são (ao menos parcialmente) sistemáticas e, assim, parte de nosso conhecimento acerca das palavras” (Anderson, 1992: 182). Assumimos aqui a visão mais tradicional, que separa morfologia de léxico.

Segundo tal proposta, o léxico é uma lista de irregularidades, de fatos imprevisíveis. Em outras palavras: faz-se uma lista daquilo que não pode ser captado (ou descrito) por uma regra, porque é excepcional de algum modo. O que pode ser apreendido por regras pertence à gramática.

Quando falamos em *irregular*, *imprevisível*, temos em mente a *arbitrariedade do signo*. Não há como prever que a sequência fônica *peixe* esteja associada, em português, ao significado que pode ser glosado como ‘tipo de animal que vive na água’. Não há regra que diga como associar sequências sonoras a significados. Mas não é exatamente assim para qualquer palavra. Se tivéssemos algo como *peixinho*, por exemplo, e levando em conta os diminutivos em português, chegaríamos à conclusão de que *peixinho* é regular. Aliás, falar em *diminutivos* é mostrar que podemos estabelecer alguma generalização a respeito do vocabulário do português: *aumentativos*, *agentivos*...

Compreender o léxico como o lugar da irregularidade, do que é imprevisível, nos leva a considerar que contém **palavras existentes** na língua, mas não todas. Os requisitos necessários pedem, pelo menos, que sejam (v. Aronoff & Anshen, 1998): (a) palavras primitivas, que representam a arbitrariedade do signo, como PEIXE, PEDRA; (b) palavras complexas que apresentam alguma parte que não é reconhecida pelo falante: se um falante desconhece CAÇO ou uma regra que forme aumentativos em *-oila*, dificilmente reconhecerá CAÇOILA; (c) palavras cujo significado não é deduzível de sua estrutura, como AMÁVEL, por exemplo. Também serão incluídas no léxico expressões idiomáticas cujo significado não se segue de suas partes formadoras: *chutar o pau da barraca* (‘dizer ou fazer tolices’), *tirar uma pestana* (‘dormir’) e assim por diante, como também formas flexionadas irregulares.

O léxico representa o conjunto de palavras que está disponível para a atuação das regras da morfologia. A morfologia lida, portanto, segundo esta visão, “com a estrutura interna das **palavras complexas potenciais** de uma língua” (Aronoff & Anshen, 1998: 237 – ênfase nossa).

As *palavras* no léxico – ou **lexemas** ou, uma vez que pensamos numa lista, os **itens lexicais** – pertencem a classes abertas, isto é, a classes que, sincronicamente, podem admitir novos membros e apresentam significado lexical. Estamos, portanto, focalizando a interação da morfologia com o léxico como o fundamento da formação de palavras – e nos deteremos aqui apenas na *derivação*. Isto nos leva a uma questão relevante: a *produtividade*.

6.3. Produtividade

Embora “muitas coisas sejam possíveis em morfologia” – como Aronoff (1976: 35) afirmava espirituosamente – “algumas são mais possíveis que outras”. Aronoff fazia referência aos fenômenos derivacionais, considerados “tipicamente esporádicos e apenas semiproductivos” (Chomsky, 1965: 278), nesse sentido em oposição aos fenômenos flexionais.

Mas o que é **produtividade**, afinal? Em termos muito simples, é a formação de palavras novas por determinada regra, que é chamada *Regra de Formação de Palavras* (ou RFP). Por isso dissemos anteriormente que a morfologia lida com palavras *potenciais*: para dar conta de regras produtivas. É muito provável que se forme um novo nome derivado de verbo por meio do sufixo *-ção* (ao contrário de *-ância* ou de *-ança*).

Uma RFP para formações em *-ção* poderia ser representada como em (6.1):

$$(6.1) [X]_V \rightarrow [[X]_V -ção]_N$$

A regra em (6.1) nos diz que se pode formar um nome em *-ção* a partir de um verbo. O verbo tomado como base para a nova palavra é representado pela variável X, que indica podermos ter qualquer membro dessa classe como base.

Cabe notar que uma regra não é produtiva em termos absolutos. Aronoff (1976) apresenta três condições para uma RFP ser produtiva. Em primeiro lugar, uma RFP aplica-se a um determinado tipo de base. O sufixo *-ção* aplica-se a verbos, não a qualquer lexema. Em segundo lugar, regras produtivas apresentam coerência semântica, isto é, o falante tem de poder prever qual será o significado da nova formação; caso contrário, não lançará mão dessa RFP. Voltando ao exemplo em (6.1), um novo nome em *-ção* tem como leituras previsíveis ‘ato de X’ e/ou ‘resultado de X’.

Por fim, o fenômeno do **bloqueio**, definido como “a não ocorrência de uma forma devido à simples existência de outra” (Aronoff, 1976: 43; Aronoff & Anshen, 1998: 239). O bloqueio resulta da interação da morfologia com o léxico: por mais produtiva que uma regra seja, não se aplicará a uma base se já existe uma palavra para aquela exata função.

A produtividade é um contínuo, que vai do muito produtivo ao improdutivo (Aronoff & Anshen, 1998: 242-245). Podemos não mais usar uma determinada RFP para formar palavras, mas ainda assim perceber a estrutura de formações desse determinado tipo.

☞ Aronoff (1976) e Basílio (1980) têm soluções diferentes para focalizar a (im)produtividade. Para Aronoff, as RFPs dão conta de tudo, uma vez que há um contínuo de produtividade. Para Basílio, fenômenos produtivos são descritos por RFPs, ao passo que a análise estrutural (de lexemas formados por regras produtivas ou não) ficam a cargo de **Regras de Análise de Estrutura** (ou RAEs).

Para Basílio, uma RFP tem como contraparte uma RAE. Assim, a par da RFP em (6.1), aqui repetida por conveniência, haveria uma RAE, que representa a possibilidade de o falante reconhecer a estrutura de formações em *-ção*:

RFP: $[X]_V \rightarrow [[X]_V \text{ -ção}]_N$

RAE: $[[X]_V \text{ -ção}]_N$

Quando dizemos que conhecemos uma palavra, dominamos, porém, mais do que a relação entre uma cadeia sonora e um significado, ou, no caso de uma palavra complexa, sua estrutura. Dominamos, também, as variações que abrangem as classes de palavras como um todo e que “completam uma palavra pela marcação de suas relações no interior de estruturas mais amplas” (Anderson, 1985b: 162). Estamos falando da flexão. Mais especificamente: um lexema pode combinar-se com propriedades morfossintáticas. É este o tema dos dois capítulos a seguir.

7

Classes de palavras, tipos de significado e questões relacionadas

7.1. Introdução

A tradição gramatical greco-latina reconheceu na palavra características de três tipos: (a) semânticas, que nos deram definições como *o substantivo é a palavra que nomeia os seres*, ou como questões acerca de quais os elementos que podem ser suprimidos do enunciado mantendo-se, ainda assim, uma estrutura com significado; (b) morfológicas, como, por exemplo, o reconhecimento de que o nome pode flexionar-se em Gênero, Número e Caso, mas não em Tempo, Modo ou Voz; e (c) sintáticas, como a identificação de que o nome, mas não o verbo, pode funcionar como sujeito, além de questões variadas acerca dos fenômenos de *concordância* e *regência*. Em decorrência desse feixe de propriedades semânticas, morfológicas e sintáticas as palavras foram distribuídas em *classes de palavras*, ou, na nomenclatura tradicional, em *partes do discurso*.



O termo **categoria** também costuma ser empregado no sentido de *classe*. Nos trabalhos sobre sintaxe, **categoria** designa os constituintes de uma expressão linguística. É, por conseguinte, um termo mais amplo que *parte do discurso*, porque abrange tanto nomes, verbos, adjetivos, como posições numa estrutura, elementos abstratos, como as *categorias vazias* ou *pro* (que representa o sujeito nulo de línguas como o português). A denominação *categoria* representa ainda propriedades de um sistema flexional, tais como T(empo) ou Agr (abreviação do termo inglês para Concordância).

Nos trabalhos sobre morfologia, **categoria** costuma manter o significado mais tradicional, de *conjunto de propriedades que se associa a determinada parte do discurso*, como Caso, Pessoa, Tempo, Modo, Aspecto, Voz, Gênero, Número...

Em consequência da tradição gramatical, habituamo-nos a considerar as palavras do português como pertencentes a dez *classes de palavras* – **nome, artigo, adjetivo, pronome, numeral, verbo, advérbio, preposição, conjunção,**