

SOBRE A NOÇÃO DE PALAVRA

ABOUT THE NOTION OF WORD

José Borges Neto*
UFPR e UNIOESTE

Resumo: Este texto discute a indeterminação da noção de palavra, mostrando que as tentativas frustradas de a definir – por critérios ortográficos, fonológicos, morfossintáticos ou semânticos – acabam definindo objetos teóricos distintos e incompatíveis entre si. Na medida em que a noção de palavra é central para as teorias da linguagem e que sem uma noção clara de palavra não é possível distinguir morfologia de sintaxe, duas alternativas se apresentam: ou definimos claramente o que é palavra, ou abandonamos essa noção em favor de alguma outra.

Palavras-chave: Filosofia da Linguística. Palavra. Morfologia. Sintaxe.

Abstract: This text discusses the indeterminacy of the notion of word, showing that the unsuccessful attempts to define it - by orthographic, phonological, morphosyntactic or semantic criteria - end up defining distinct and incompatible theoretical objects. To the extent that the notion of word is central to theories of language and that without a clear notion of word it is not possible to distinguish morphology from syntax, two alternatives are presented: either we clearly define what a word is, or we abandon this notion in favor of some other.

Keywords: Philosophy of Linguistics. Word. Morphology. Syntax.

A ORIGEM

Os gregos da Antiguidade clássica criaram uma noção que se tornou a base dos estudos linguísticos até nossos dias: a noção de **palavra**. Uma primeira definição de palavra pode ser encontrada no livro *Da interpretação (Peri hermēneías)* de Aristóteles¹, onde ele define a palavra como um símbolo (oral) das imagens ou afeições (paixões) que estão na alma dos interlocutores². Aristóteles entende a palavra como parte do discurso (“*méros lógou*”), isto é, como a unidade que, combinada com outras, vai resultar nos enunciados linguísticos. Para Aristóteles, na mente humana há imagens das coisas do mundo, juntamente com as paixões, e as palavras são representações linguísticas desses “objetos mentais”. Ou seja, como frequentemente se diz, as palavras são símbolos de **ideias**.

* Professor Titular aposentado da Universidade Federal do Paraná (UFPR), com atuação junto aos programas de Pós-Graduação em Letras da UFPR (Professor Sênior) e da Universidade Estadual do Oeste do Paraná - UNIOESTE (Professor Visitante). E-mail: borgesnetojose@gmail.com.

¹ Uso a tradução espanhola feita por Miguel Candel Sanmartín (ver ARISTÓTELES, 1988). Os trechos citados estão identificados pela numeração convencional para que possam ser consultados em qualquer edição (os números seguidos de letras – “16a”, por exemplo – indicam parágrafos, e os números que se seguem indicam as linhas no interior dos parágrafos). Todas as traduções para o português de citações originalmente em espanhol ou inglês são de minha responsabilidade.

² Aristóteles define a palavra escrita como um símbolo (gráfico) da palavra oral. O termo “símbolo” (*symbola*) significa algo como “contrato” ou “convênio”, no grego clássico, o que revela que Aristóteles atribui um caráter convencional às relações entre escrita e fala e entre fala e pensamento.

Na verdade, há duas expressões em grego que são usadas para os sentidos que damos a *palavra*: *méros lógou* (“parte do discurso” – literalmente, parte do *lógos*) e *léksis*. Jean Lallot (1989, p. 119) diz que as duas expressões “[...] funcionam como sinônimos em variação livre”; Gissele Chapanski (2003, p. 143-144), por outro lado, diz que as *lékseis* são instanciações (realizações físicas) das partes do discurso e que a expressão *méros lógou* seria melhor entendida como **classe de palavras**.

O que podemos dizer com certeza é que, para os gregos, **palavra** é um conjunto de sons que corresponde, de maneira razoavelmente estável, a uma **ideia**. Dessa forma, na ontologia da teoria grega, há uma entidade “híbrida”, de dupla face, que, por um lado, é uma forma fisicamente manifestada como sons, e, por outro lado, é uma imagem ou uma afeição da alma.³

No livro *Da Interpretação*, Aristóteles nos fala apenas de dois tipos de palavras: o **nome** (*ónoma*) e o **verbo** (*rhêma*). O nome é um som significativo por convenção que não indica tempo e que não apresenta partes significativas. Em suas palavras, “[...] com efeito, em *Kállippos*, *ippos* não significa nada por si mesmo, como ocorre no enunciado *kalòs híppos*” (ARISTÓTELES, 1988, *Da Int.*, 16a, 20-22)⁴. O verbo, por sua vez, é um som que significa o que se diz sobre um outro, tem tempo e nenhuma de suas partes tem significado separadamente. Nas palavras de Aristóteles: “Digo que cossignifica tempo no sentido em que, enquanto *saúde* é um nome, *está são* é um verbo: com efeito, cossignifica que se dá agora. E sempre é signo do que se diz acerca de outro, no sentido de que se diz sobre um sujeito” (ARISTÓTELES, 1988, *Da Int.*, 16b, 6-11).

Não podemos deixar de apontar a possível confusão existente entre as **palavras**, que são nome ou verbo, e as **funções** que as palavras exercem no enunciado (sujeito ou predicado). A distinção já era potencialmente confusa para Platão e se mantém assim em Aristóteles. No livro *Da Interpretação*, Aristóteles está principalmente interessado em investigar a relação entre a linguagem e o pensamento. Em sua perspectiva, para que um enunciado possa ser verdadeiro ou falso, uma de suas partes tem de significar a coisa de que se fala (e é *ónoma*) e outra parte tem de significar o que se fala da coisa nomeada (e é *rhêma*): a parte do discurso que nomeia não tem tempo e a parte do discurso que predica deve ter tempo. Como lemos em Robins (1979):

Para Aristóteles a frase (*lógos*) compreende algo mais ao nível semântico, pois, ao contrário das palavras isoladas, ela afirma ou nega um predicado, ou faz uma declaração existencial. Em oposição a *ónoma*, definiu *rhêma* como portador de uma referência temporal e como representante do predicado. Esta segunda parte da definição permitiu-lhe incluir, como fizera Platão, adjetivos como *leukós*, “branco”, e *dikaíos*, “justo”, entre as *rhêmata*, visto que frequentemente funcionam em grego como predicado (*leukós ho híppos*, “o cavalo é branco”). Como a cópula *estí*, “é”, está subentendida e é sempre suscetível de inserção, poder-se-ia dizer que tais adjetivos também são portadores de referência temporal (o presente). Por esse motivo, a tradução de *ónoma* e *rhêma* por *nome* e *verbo* nesta etapa do desenvolvimento da teoria gramatical grega pode ser enganosa. (ROBINS, 1979, p. 21).

³ Conceção que, não por coincidência, reencontramos na noção de **signo** em Saussure.

⁴ *Kállippos* é um nome próprio composto de *kallòs* e *hippos* (respectivamente, “bonito” e “cavalo”). O exemplo de Aristóteles quer mostrar que, embora sendo palavras significativas, *kallòs* e *hippos*, enquanto partes do nome próprio *Kállippos*, não possuem significação (ARISTÓTELES, 1988, p. 37).

Da mesma forma, a identificação de **parte do discurso** com palavra também pode ser enganosa, já que as partes do discurso – a parte que nomeia e a parte que predica – podem ser exprimidas por mais de uma palavra (sintagmas nominais como sujeito, por exemplo).

Outro problema de entendimento está na noção de *lógos* (de que a palavra é parte constituinte). A dificuldade em traduzir *lógos* nas línguas modernas é bem conhecida: para os gregos, o *lógos* era, simultaneamente, “proporção”, “explicação”, “fundamento”, “discurso”, “argumento”, “regra de conduta”, “hipótese”, “razão”, “pensamento”, “raciocínio”, “narrativa”, “enunciado”, “debate”, “assunto de um poema ou narrativa”, “proposição”, “frase” e ainda algumas outras possibilidades⁵.

Enfim, simplificando muito a questão, podemos dizer que o pensamento linguístico grego se constrói a partir da noção de **palavra**: nada há de significativo abaixo da palavra e é somando significados de palavras que construímos proposições⁶.

Uma gramática baseada na palavra compreende três procedimentos principais: a identificação da palavra como entidade linguística isolável; o estabelecimento de um conjunto de classes de palavras [...]; a apreensão de categorias gramaticais adequadas à descrição e à análise da morfologia dos vocábulos reunidos em paradigmas e das relações sintáticas que surgem entre as palavras na construção de frases. (ROBINS, 1979, p. 20).

E é esse o roteiro que a construção da gramática grega obedece. Identificam-se as palavras como **unidades significativas atômicas** (conjuntos mínimos de sons que correspondem a ideias); estabelecem-se **classes de palavras** a partir dos tipos de ideias de que são símbolos ou dos papéis que desempenham nos enunciados (**nomes, verbos, pronomes, advérbios** etc.); estabelecem-se categorias gramaticais que descrevem a morfologia “típica” de cada classe (**fleções, declinações e conjugações**); e estabelecem-se tipos de relações sintáticas na composição das frases (**concordância e regência**).

O tratamento que Dionísio Trácio (cerca de 100 a.C.) dispensa ao **nome** e ao **verbo** são bons exemplos. Para Dionísio:

[O nome] é a parte da frase sujeita à variação de caso, que designa um corpo ou uma coisa abstrata – um corpo como em “pedra”, uma coisa abstrata como em “educação” –, que pode ter sentido geral ou específico – geral como em “homem” ou “cavalo”, específico como em “Sócrates”. (CHAPANSKI, 2003, p. 28).

[O verbo] é uma palavra não sujeita à variação de caso, que admite tempo, pessoas, números e exprime atividade ou passividade. (CHAPANSKI, 2003, p. 32).

⁵ Cf. Henry George Liddell, Robert Scott. *A Greek-English Lexicon*. Disponível em: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.04.0057%3Aentry%3Dlo%2Fgos>. Acesso em: 6 dez. 2019. A escolha das acepções e suas traduções para o português é minha.

⁶ Essa primazia da palavra não é uma necessidade e houve tradições que, se não a evitaram, ao menos a relativizaram. Mesmo no interior do pensamento grego, encontramos quem colocasse em questão a equação “uma palavra/um significado”. Segundo Robins (1979, p. 17), para os estoicos, “[...] os significados das palavras não existem isoladamente e podem variar de acordo com a situação contextual”.

O **nome** é identificado por ser um designador de “coisas”, concretas ou abstratas, próprias ou comuns, e por apresentar variação de caso. O **verbo**, por sua vez, exprime atividade ou passividade (**voz**) e apresenta variação de tempo, pessoa e número.

A “morfologia” de nomes e verbos tem a ver com as diferentes formas gramaticalmente relevantes que essas palavras podem apresentar. Dionísio já nos fala que os nomes variam em **gênero** e **número**, mas não em **voz**, **tempo** e **pessoa**, que seriam próprios do verbo. Assim, definidas as classes de palavras, é necessário estabelecer as categorias gramaticais (“atributos consequentes”) que podem se aplicar a cada uma delas.

A teoria linguística grega dá centralidade à **morfologia**⁷ e cria um primeiro modelo de análise morfológica: o modelo que Hockett (1954) vai chamar de “palavra-e-paradigma”.

A NOÇÃO DE PALAVRA NA TRADIÇÃO GRAMATICAL

Essa concepção de palavra sobreviveu na tradição gramatical ocidental até, pelo menos, o final do século XVIII. Como se lê no verbete *Unidades Significativas* do *Dicionário das Ciências da Linguagem*, de Oswald Ducrot e Tzvetan Todorov:

Até o fim do século XVIII, a maior parte dos linguistas ocidentais está tacitamente de acordo em que a mais pequena unidade linguística que, simultaneamente, tem uma realidade na cadeia falada e é portadora de significação é a *palavra*: a frase é composta por proposições, compostas elas mesmas por palavras. Só se decompõe a palavra em unidades não-significativas (as sílabas, as letras). Aliás, a definição de palavra continua a estar geralmente implícita. É que a divisão do enunciado em palavras parece gozar de uma espécie de evidência, que dispensa qualquer determinação explícita. Com efeito, esta divisão não se apoia apenas numa tradição gráfica solidamente estabelecida desde o Renascimento, mas também em fenômenos de pronúncia incontestáveis: a palavra é a unidade de acentuação. (DUCROT; TODOROV, 1976, p. 245).

A noção grega de **palavra** adentra ainda a primeira metade do século XX. Segundo Robins (1979, p. 21), “[a] definição de Aristóteles é notavelmente semelhante à de Meillet: ‘associação de um dado significado com um dado grupo de sons suscetível de emprego gramatical’”⁸. Da mesma forma, encontramos na *Gramática Metódica da Língua Portuguesa*, de Napoleão Mendes de Almeida (1911-1998), a seguinte afirmação: “[...] a linguagem indica o pensamento; as palavras, como partes que são da linguagem, indicam as partes do pensamento, ou seja, as ideias” (ALMEIDA, 1979, p. 17, §3)⁹.

A associação entre **palavra** e **ideia** permanece firme. Embora já apareçam, desde o século XVIII, movimentos na direção de considerar partes significativas na palavra, a concepção vigente da organização interna da palavra impede que surjam noções como a de **morfema**. Ducrot e Todorov citam o caso de Jacques Turgot (1727-1781) que, no artigo “*Étymologie*” da

⁷ Os estudos sintáticos só aparecem tardiamente – cerca de 200 anos depois do tratamento morfológico de Dionísio – com Apolônio Díscolo.

⁸ Trata-se do linguista francês Antoine Meillet (1866-1936).

⁹ A *Gramática* de Napoleão Mendes de Almeida teve 47 edições: primeira edição é de 1943 e a última, póstuma, é de 2002. Cito aqui a 28ª edição, de 1979.

Encyclopédie, diz que, no caso da palavra ser derivada de outra, é necessário “[...] devolvê-la à sua raiz, despojando-a desse aparelho de terminações e de inflexões gramaticais que a disfarçam; se é uma composta, é necessário separar as suas diferentes partes” (DUCROT; TODOROV, 1976, p. 245).

O modelo morfológico tradicional – palavra-e-paradigma – entende que as flexões não criam novas palavras e que, portanto, a ideia central permanece a mesma: trata-se de terminações distintas e de inflexões **de uma mesma palavra**. Provavelmente, Turgot está considerando que o verdadeiro significado da palavra está em sua raiz e que a etimologia só pode realizar sua tarefa se privilegiar esse significado, digamos, nuclear. De fato, não se trata de reconhecer partes significativas na palavra, mas de localizar seu significado essencial.

Esse modelo morfológico considera a **palavra** uma entidade abstrata que pode se manifestar concretamente sob diversas formas. A cada palavra pode-se associar um **paradigma**, que é o conjunto de formas alternativas em que as palavras se manifestam. A relação entre as formas do paradigma recebe o nome de “flexão”.

Em um exemplo simples: temos, em português, a palavra “*mesa*”; essa palavra pode se manifestar concretamente por meio de duas formas alternativas: **mesa** e **mesas**. Essas duas possibilidades de manifestação da palavra “*mesa*” constituem o paradigma associado a ela e a relação entre as formas é uma relação de flexão. Podemos falar, então, que a palavra “*mesa*” se flexiona, ora aparece como **mesa**, ora aparece como **mesas**. A expressão “*mesa*” (entre aspas) é apenas uma forma de citação do conjunto que constitui o paradigma e, em consequência, não é uma expressão linguística propriamente dita. A forma “*mesa*” jamais é pronunciada: se tentarmos pronunciá-la, diremos ou **mesa**, ou **mesas**. O significado “essencial”, aquele que constitui a ideia que dá identidade à palavra, existe no conjunto das formas do paradigma, sempre acrescido de significados “secundários”: no nosso exemplo, o significado de singular, em **mesa**, e o significado de plural, em **mesas**. Para encontrar o significado de “*mesa*”, é preciso despojar as formas do paradigma dos significados secundários. Dessa maneira, “*mesa*” é alguma coisa (alguma “ideia”) que não é singular ou plural. Creio que Turgot faz um raciocínio muito semelhante a esse.

O aspecto central do modelo palavra-e-paradigma reside na consideração de que as formas do paradigma são formas de uma mesma palavra, que – suponho – será sempre identificada nocionalmente, isto é, com base nas características da ideia a que corresponde (se corresponde a objetos, é nome; se corresponde a eventos, é verbo; se corresponde a circunstâncias do evento, é advérbio etc.). Assim, na medida em que o paradigma associado à palavra “*menino*” inclui as formas **menino**, **menina**, **meninos** e **meninas**, essas formas são manifestações alternativas de uma mesma palavra; as mais de 70 formas alternativas presentes no paradigma de “*cantar*” (**cantei**, **cantaria**, **cantamos**, **cantasse**, **cantado** etc.) são flexões da palavra “*cantar*”, e assim por diante.

Como há palavras cujo paradigma só contém uma forma – como “*hoje*”, “*agora*”, “*sobre*” etc. – podemos dizer que há palavras que não se flexionam (aparecem sempre de uma mesma e única forma) e que são chamadas de “palavras invariáveis”. Assim, o modelo palavra-e-paradigma conduz-nos diretamente à distinção entre palavras **variáveis** e **invariáveis**: “*mesa*”, “*cantar*” e “*menino*” seriam palavras variáveis, já que seus paradigmas apresentam mais de uma forma alternativa, enquanto “*ontem*”, “*de*” e “*apenas*” seriam invariáveis.

ALTERNATIVAS AO ENFOQUE TRADICIONAL

Somente a partir do século XIX, a natureza atômica da palavra começa a ser colocada em questão. Como dizem Ducrot e Todorov (1976, p. 245): “Foi o aparecimento da linguística comparativa que impôs uma dissociação da palavra em unidades significativas mais elementares”.

O reconhecimento, digamos, “oficial” de que a palavra tem partes significativas só vai aparecer no final do século XIX, em 1895, com o trabalho do linguista polonês Jan Baudouin de Courtney (1845-1929) que criou o termo “morfema” para designar as partes integrantes da palavra que podiam ter significados independentes (raízes, afixos e desinências). E a incorporação da noção de morfema ao arsenal teórico dos linguistas levou ao surgimento de dois modelos morfológicos concorrentes com o modelo palavra-e-paradigma, que Hockett (1954) chamou de modelo “item-e-processo” e modelo “item-e-arranjo”.

O modelo item-e-processo (IP) supõe palavras “básicas” e processos que se aplicam a elas para a obtenção de novas palavras. Por exemplo, em vez de supor que **mesa** e **mesas** são flexões da palavra “mesa”, podemos supor que **mesa** é uma palavra “básica” e que, por um processo de sufixação (que “pluraliza” a forma básica), podemos obter a palavra **mesas**. Da mesma forma, a palavra **menino** seria uma palavra “básica” e, por dois distintos processos de sufixação, obteríamos **meninos** (menino+s) e **menina** (menino+a)¹⁰. A forma **menina**, obtida a partir da forma **menino**, poderia sofrer um novo processo de sufixação, que resultaria em **meninas**. No modelo IP, a ênfase é dada nas relações entre as palavras, particularmente nos modos pelos quais conjuntos de palavras são derivadas de bases comuns pela aplicação de certos processos (que se manifestam sob a forma de adições de material fonológico, como em **mesa/mesas**; subtração de material fonológico, como em **irmão/irmã**; ou alternâncias, como em **avô/avó**). No fundo, é um modelo em que todas as relações entre as palavras são “derivacionais”, não havendo distinção entre **derivação** e **flexão**. Hockett afirma que o modelo IP, embora adequado ao tratamento das mudanças linguísticas, não é o modelo mais interessante para as análises sincrônicas – e, por isso, perdeu prestígio com o aparecimento do estruturalismo no início do século XX.

O modelo item-e-arranjo (IA), por sua vez, entende que as palavras são cadeias de elementos (como “trenzinhos”), cada elemento portando seu próprio significado. Assim, tanto **mesa** quanto **mesas** são constituídas por dois morfemas: um primeiro morfema, raiz, que seria “mesa=”, comum às duas formas, e um segundo morfema, indicador de número, que, em um caso, seria um ZERO, indicando número **singular**¹¹, e, no segundo caso, seria um -s, indicando número **plural**. Dessa forma, a palavra **mesa** seria analisada como “mesa+Ø” e **mesas** seria analisada como “mesa+s”.

O que precisa ficar claro desde já é que esses dois modelos morfológicos assumem noções claramente diferentes de **palavra**, entre si e com relação ao modelo palavra-e-paradigma tradicional. Os modelos IP e IA não usam a noção de paradigma e, em consequência, a noção de flexão deixa de fazer sentido.

¹⁰ Alguma regra morfofonêmica eliminaria a vogal temática final da forma básica **menino** na formação de **menina** (ver CÂMARA JUNIOR, 1970, p. 63).

¹¹ Em um sistema com poucas alternativas, a ausência de sinal pode ser significativa. Meu primeiro professor de linguística – o Prof. Eurico Back – usava o seguinte exemplo para explicar esse fato: suponha um carro trafegando numa rua; se o motorista acende a seta da esquerda, sinaliza que vai virar à esquerda; se acende a seta da direita, sinaliza conversão à direita; se não acende nenhuma seta, sinaliza que seguirá em frente. A ausência de sinal pode ser tão informativa (tão significativa) quanto a presença de sinal.

Para o modelo item-e-arranjo, a palavra não é uma unidade, mas um conjunto de morfemas (organizados em uma estrutura “sintática” específica). A palavra **meninas**, por exemplo, é um conjunto ordenado de três morfemas: menino + a + s. Na ordem, um morfema raiz (**menino**), que pertence a um conjunto aberto, um morfema *-a*, que pertence a um conjunto fechado e que ocupa uma posição restrita a morfemas indicadores de gênero (a outra possibilidade é apenas o ZERO), e um morfema *-s*, que pertence a um conjunto fechado e que ocupa uma posição restrita a morfemas indicadores de número (a outra possibilidade de ocupante dessa posição é também o ZERO). Dessa forma, podemos estabelecer uma estrutura sintática para essas palavras, estrutura que é a mesma para muitas outras palavras, que só se distinguem pelos morfemas que ocupam as posições na estrutura. Em um esquema:

Raiz	Gênero	Número
menino	∅	∅
garoto	-a	-s
gato		
...		

Dada a estrutura (o *arranjo*), se pegarmos sempre um elemento de cada coluna (os *itens*) e os juntarmos em uma palavra, iremos obter: **menino**, **menina**, **garota**, **gato**, **gatas** etc. Nas palavras de Hockett:

A essência de IA é falar simplesmente de coisas e dos arranjos em que essas coisas ocorrem [...]. Assume-se que qualquer enunciado de uma dada língua consiste inteiramente de um número mínimo de elementos gramaticalmente relevantes, chamados morfemas, num certo arranjo relativamente aos outros. A estrutura do enunciado é especificada pelo estabelecimento dos morfemas e do arranjo. (HOCKETT, 1954, p. 212, tradução nossa).

No modelo IA, portanto, a distinção entre derivação e flexão também desaparece. Se a palavra é um conjunto de morfemas organizados em um arranjo, e se o conjunto de morfemas de **menino** não é o conjunto de morfemas de **meninas**, necessariamente temos de considerar que **menino** e **meninas** são duas palavras distintas. No caso de **meninice** e de **meninada**, temos a mesma situação: trata-se de duas palavras distintas entre si e, mais importante, distintas de **menino** e **meninas**. No entanto, são palavras distintas porque são novos conjuntos de morfemas. O que não existe mais é o critério para dizer que **menino** e **meninas** são flexões de uma mesma palavra – a palavra “menino” – enquanto **meninice** e **meninada** são palavras derivadas. Ou tudo é derivação, como assume o modelo IP¹², ou tudo é flexão, atribuindo um novo sentido ao termo “flexão”. O único modo de recuperar a distinção entre flexão e derivação é algum tipo de **estipulação**. Se for possível achar algum critério, suficientemente consistente, para distinguir flexão de derivação (e Gonçalves (2011) nos mostra que não há esse critério), certamente ele nos levará a resultados muito distintos do que se obtém no modelo palavra-e-paradigma.

Embora as coisas não sejam tão simples como parecem à primeira vista, como o próprio Hockett reconhece em seu texto, o modelo morfológico item-e-arranjo teve grande influência

¹²Todas as formas de palavras seriam resultado da aplicação de operações (processos) a um item básico, que poderia ser, a depender da proposta teórica, um morfema ou uma palavra considerada básica.

nos estudos morfológicos que se fizeram no século XX. E, surpreendentemente, também na morfologia de nossas gramáticas escolares¹³.

VÁRIAS NOÇÕES DE “PALAVRA”

Uma primeira conclusão que podemos tirar do que vimos anteriormente é que há mais de uma noção de **palavra** em jogo, a depender do modelo de análise que assumimos: como o conjunto das formas de um paradigma, como um arranjo de morfemas ou como uma “base” sobre a qual se aplica um processo. Entretanto, há ainda outras noções de **palavra** envolvidas, que não resultam exatamente dos modelos morfológicos assumidos.

Vejamos o que diz o *Dicionário de Linguagem e Linguística* de Robert L. Trask (1944-2004), traduzido e adaptado por Rodolfo Ilari (TRASK, 2004). Trask abre o verbete definindo palavra como “[...] uma unidade linguística tipicamente maior que o **morfema**, mas menor que um **sintagma**”¹⁴ (TRASK, 2004, p. 218), e segue com o seguinte comentário: “O termo **palavra** poderia parecer familiar e suficientemente transparente, mas, na realidade, há pelo menos quatro maneiras de definir a palavra, e essas maneiras não se equivalem” (TRASK, 2004, p. 218).

Na verdade, o texto de Trask é enganador, na medida em que, aparentemente, ele supõe que exista uma unidade linguística **palavra**, entendida como uma entidade ontológica presente nas línguas, e que essa entidade pode ser definida de quatro maneiras distintas. Quando olhamos suas definições, no entanto, o que encontramos são quatro entidades distintas, cada uma delas com seus próprios critérios definidores: a **palavra ortográfica**, a **palavra fonológica**, o **item lexical** (ou **lexema**) e a **palavra morfossintática**.

Trask define a palavra ortográfica como “[...] algo que se escreve com espaços brancos de ambos os lados, mas sem espaços brancos em seu interior” e a palavra fonológica como “algo que se pronuncia como uma única unidade” (TRASK, 2004, p. 218). Ignorando a vagueza das formulações, posso dizer que elas são muito problemáticas, mesmo quando tentamos torná-las mais precisas.

É interessante ver o que diz Luiz Carlos Schwindt sobre palavras ortográficas e palavras fonológicas:

No mais das vezes, [...], os falantes fazem uso de um critério ortográfico para identificar uma palavra, [...]. Sabemos, contudo, que esse conhecimento, além de dependente da cultura escolar (letrada) [...] é bastante controverso, já que qualquer um de nós já titubeou ao decidir sobre a segmentação de um vocábulo (p.ex., em português, *de repente* e não *derrepente*, mas *devagar* e não *de vagar*).

¹³ É relevante destacar que, no âmbito da sintaxe, encontramos modelos de análise que adotam tanto a perspectiva item-e-arranjo quanto a perspectiva item-e-processo. A gramática gerativa chomskiana (ao menos em suas versões anteriores aos anos 1990) tem, claramente, o modelo item-e-arranjo como inspiração (há regras que geram estruturas sintagmáticas, ou “arranjos”, e há regras de inserção lexical que preencham os nós terminais com itens lexicais). As gramáticas categoriais, por sua vez, assumem a perspectiva item-e-processo: a gramática consiste simplesmente de um léxico categorialmente especificado e algumas poucas operações, como a **aplicação funcional**, que permitem construir os enunciados. Existe, também, uma **morfologia categorial**, que trata as estruturas das palavras por procedimentos próximos do modelo IP (ver, por exemplo, Hoeksema e Janda 1988).

¹⁴ Martin Haspelmath (2011, p. 61, tradução nossa), sobre afirmação equivalente, diz: “[...] essa afirmação é na verdade fraca demais se não for especificado qual o critério que pode ser usado para identificar tal unidade. De fato, [...] podemos selecionar, para a maior parte das línguas, um grande número de unidades intermediárias, se quisermos”.

Ao serem inquiridos sobre o que é uma palavra, os falantes podem ainda fazer uso de critérios de natureza fonológica, apelando também para espaços – só que agora caracterizados por vazios sonoros (ou pausas) – entre cada palavra falada. (SCHWINDT, 2014, p. 128-129).

O quadro desenhado por Schwindt é real. Se perguntarmos aos falantes o que é uma palavra, vamos receber uma resposta retirada da experiência dos falantes com a escrita (“palavra” definida como palavra ortográfica) e, eventualmente, considerando o grau de escolaridade do falante, com alguns acréscimos de natureza fonológica. O que não fica claro na exposição de Schwindt é que as respostas dos falantes são respostas que podemos classificar de **linguística ingênua** (“*folk linguistics*”). O problema aparece quando fazemos a pergunta aos linguistas e o que recebemos como resposta não ultrapassa o nível de uma linguística ingênua.

De qualquer forma, vamos olhar os quatro conceitos de palavra trazidos por Trask. Começo com a palavra “ortográfica”. A ortografia do português se construiu historicamente a partir da ortografia do latim, foi inicialmente formatada pelos editores e não pelos gramáticos¹⁵, sofreu inúmeras “reformas” (algumas oficiais e muitas outras oficiosas), passou por intensos debates sobre qual deveria ser seu fundamento – se a pronúncia ou a etimologia – e acabou sendo matéria de lei. Nesse percurso, muita coisa foi fixada de qualquer jeito e o sistema final apresenta pouca consistência lógica. Quando do debate entre a proposta de ortografia da pronúncia (que dizia que a escrita devia registrar exclusivamente o que se falava) e a ortografia etimológica (que dizia que a escrita devia registrar também a origem histórica da palavra), nos séculos XVIII e XIX, Jerônimo Soares Barbosa aponta o seguinte:

Entre estas duas *ortografias* caminha a *usual*, assim chamada, porque não tem outra auctoridade senão a do uso presente e dominante; ja para seguir as etimologias, e introduzir arbitrariamente escripturas mui alheas da pronunciação presente; ja para não fazer caso da dirivação mesma, e incoherente em seus procedimentos escrever, por ex.: *he, huma* com *h*, que não ha na origem Latina; e *filosofia*, e *física* com *f* e *z*, que não ha nas palavras Gregas. (BARBOSA, 1822, p. 57).

Essa **autoridade do uso**, como diz Soares Barbosa, cria inúmeras inconsistências no sistema ortográfico do português. O próprio Soares Barbosa registra em sua gramática a palavra **qualquer** e, juntamente a ela, registra a palavra **quemquer**. Ora, nosso sistema ortográfico considera **qualquer** uma palavra ortográfica, mas recusa reconhecimento igual para **quemquer**. Poderíamos perfeitamente considerar que em ambos os casos teríamos o que deveriam ser duas palavras ortográficas, o que explicaria o plural bizarro de **qualquer** (= quaisquer). Quem ainda não se perguntou por que **embaixo** é uma palavra ortográfica e **em cima** são duas?

Note-se, também, que formas como **papelzinho**, **anãozinho** e **bastãozinho** são palavras ortográficas, embora suas formas plurais e femininas correspondentes fossem imensamente mais bem tratadas se considerássemos que estamos diante de sintagmas (um substantivo seguido de um adjetivo): o feminino de **anãozinho** é **anãzinha**, o plural de **papelzinho** é **papezinhos** e

¹⁵ Independentemente das propostas dos ortógrafos, que eram todas muito próximas do latim, mesmo quando reconhecidas as diferenças fonológicas (Fernão de Oliveira é um bom exemplo), os livros eram impressos pelos editores com os tipos (latinos) que já existiam. Dessa forma, os editores, tacitamente ao menos, limitaram o grau de afastamento do latim das ortografias propostas para as línguas vernáculas.

o plural de **bastãozinho** é **bastõezinhos**. Parece estranho que palavras terminadas em “zinho/zinha” marquem duas vezes o gênero e o número, ou seja, os fatos parecem indicar que estamos diante de plurais e femininos regulares de DUAS palavras¹⁶. E se alguém ainda tem dúvidas, note que **zinho/a** é uma forma livre que pode funcionar também como substantivo: “Aquele zinha está de olho no namorado da Maria”.

Em suma, a noção de palavra ortográfica é muito importante no processo de alfabetização, é muito valorizada pelos falantes em geral (que dão extrema importância à correção ortográfica), mas não é uma noção que mereça especial atenção dos estudos linguísticos, exceto quando se estudam aquelas noções oriundas da linguística ingênua.

Além disso, até há cerca de mil anos, a prática corrente era a da *scripta continua*, isto é, da escrita sem separações com espaços em branco. Como aponta Haspelmath (2011, p. 36, tradução nossa), “[...] o uso dos espaços na moderna ortografia é em alguma extensão guiado pela estrutura das línguas, mas não de modo que a soletração convencional possa ser usada para decidir situações contenciosas”. Desse modo, a palavra ortográfica não pode ser critério para uma definição geral de **palavra**.

Passemos agora à palavra fonológica. A formulação de Trask de palavra fonológica como “algo que se pronuncia como uma única unidade” é excessivamente vaga. Provavelmente, ele está pensando em algo semelhante à definição bloomfieldiana de palavra como uma **forma livre mínima**, em termos fonológicos. Contudo, se pensarmos na cadeia de sons que constitui um enunciado como “As asas azuis das aves” – fonemicamente algo como /a.’za.za.za.’zuys.da.’za.ves/ – e se procurarmos localizar as palavras na cadeia sonora, sem apelar para a escrita ou para a semântica, teremos sérias dificuldades. Os estruturalistas identificavam as palavras fonológicas a partir dos acentos – procedimento que se mantém na maioria das teorias fonológicas atuais – e se olharmos para a cadeia sonora, vamos encontrar três proeminências de intensidade (as sílabas com o início marcado pelas aspas simples), mas essa descoberta não nos ajuda muito. Quantas palavras fonológicas temos aqui? Três?¹⁷ Onde está a palavra fonológica que corresponde à palavra ortográfica **asas**? A palavra ortográfica **das** é, de fato, uma única palavra ou são duas: **de** e **as**? E, independentemente da resposta, onde estaria o **de** ou o **as** na cadeia sonora?

Obviamente, alguém poderia produzir o enunciado anterior em uma cadência e com uma velocidade tal que todas as palavras ortográficas corresponderiam a palavras fonológicas: algo como /as::’a.zas::a.’zuys::das::’a.ves/. Tenho dúvida de que essa pronúncia poderia ter alguma relevância para os estudos fonológicos, exceto talvez se estivermos interessados em estudar a influência do letramento na pronúncia, ou seja, a influência da escrita na fala¹⁸.

¹⁶ Cf. Schwindt (2014, p. 129, grifos do autor), que diz: “[...] *palavra* seria, então, aquela estrutura que admite flexão na borda direita do tema; desse modo, num sintagma como *esses cachorros magros*, por exemplo, estaríamos diante de três palavras, já que as três suportam o elemento flexional”.

¹⁷ Parece que temos, no enunciado, cinco palavras ortográficas e apenas três palavras fonológicas. Mattoso Câmara, por exemplo, procurava identificar os **vocábulos fonológicos** a partir dos acentos, mas supunha simultaneamente, em seus exemplos, palavras morfossintáticas correspondentes, identificadas certamente por algum outro critério – talvez mesmo pelo critério ortográfico (ver CÂMARA JUNIOR, 1970, p. 34-39).

¹⁸ Lembro de quando meu filho, então aluno alfabetizando, descobriu que a palavra /’lo.ko/ se escrevia “louco” e que podia ser enunciada com um ditongo. Ele passou um dia inteiro brincando de dizer /’low.ko/, com destaque na primeira sílaba, e rindo com a descoberta. Escrevi, na época, um pequeno texto sobre a tal “redução de ditongos” – que, na verdade, é um fenômeno de **ditongação**, desencadeado pelo letramento – a partir dessa experiência (ver BORGES NETO, 1980).

A conclusão mais importante, no entanto, é que dificilmente poderíamos sustentar que os termos “palavra ortográfica” e “palavra fonológica” denotam um mesmo fenômeno ou uma mesma entidade teórica da língua.

Passemos, então, para as outras duas “entidades” que Trask associa à noção de palavra: o item lexical e a palavra morfossintática. As definições propostas por Trask são as seguintes: “O *item lexical* ou *lexema* é uma palavra do dicionário, uma unidade para a qual se espera que exista uma entrada própria no dicionário” (TRASK, 2004, p. 218), e “A *forma gramatical de palavra* (“gramatical word-form”, GWF) ou *palavra morfossintática* é qualquer uma das formas que um item lexical pode assumir para fins gramaticais” (TRASK, 2004, p. 218).

Deixando de lado a estranheza que desperta a consideração de que as escolhas dos dicionaristas possam valer como critério para a delimitação da noção de **palavra**, já que os dicionaristas **usam** a definição estabelecida pelos gramáticos ou pelos linguistas e não **estabelecem** a definição¹⁹, vamos assumir que as escolhas dos dicionaristas correspondam a alguma conceituação de palavra teoricamente justificada. Na verdade, essa assunção é necessária, já que a própria definição de palavra morfossintática depende crucialmente da noção de item lexical. Eventuais imprecisões relacionadas à noção de item lexical contaminam mortalmente a noção de palavra morfossintática.

O que chama atenção, porém, é a proximidade que podemos perceber entre as noções de item lexical e palavra morfossintática, da linguística moderna, e as noções de palavra e de paradigma do modelo morfológico tradicional: o item lexical corresponde exatamente à forma de citação de um paradigma (a **palavra**)²⁰, e a palavra morfossintática é qualquer forma que pertença ao paradigma de uma dada palavra. Dessa forma, Trask (e boa parte dos linguistas que concordariam com a distinção entre **item lexical** e **palavra morfossintática**, nos termos colocados por Trask) está raciocinando a partir do quadro teórico do modelo palavra-e-paradigma e simplesmente alterando os termos: o modelo palavra-e-paradigma, para Trask, passaria a se chamar modelo “item lexical-e-conjunto de palavras morfossintáticas”: a palavra/o lexema “mesa” estaria associado a um paradigma que conteria duas palavras morfossintáticas – **mesa** e **mesas**. O que significa que estamos diante de um retorno ao modelo tradicional, ocultado por uma mudança de nomes – em uma paráfrase do refrão do personagem Dom Fabrizio, no romance *O leopardo* de Tomasi di Lampedusa (1896-1957): **mudam-se os termos para que as noções permaneçam as mesmas**.

A consequência dessa incorporação do modelo palavra-e-paradigma em um tratamento que poderíamos chamar de “moderno”, em oposição ao tratamento tradicional, é que o item lexical (ou o lexema) deixa de ser uma expressão linguística e passa a ser apenas uma abstração, com existência apenas em uma metalinguagem.

Trask²¹ tem consciência das limitações de suas quatro definições de palavra para a tentativa de chegar a alguma noção consistente de **palavra**.

¹⁹ Eventualmente, o dicionarista é também um gramático ou um linguista. Nesse caso, é **enquanto** gramático/linguista que estabelece a definição, não **enquanto** dicionarista.

²⁰ Essa natureza de “forma de citação” da **palavra** no modelo de análise morfológica tradicional e o caráter de “entrada de dicionário” do item lexical, certamente, não é mera coincidência.

²¹ Mais provavelmente, Rodolfo Ilari, que traduziu e adaptou o dicionário. Não conheço o texto original de Trask (em inglês), mas certamente os exemplos todos são produzidos por Ilari.

O item *grana preta* consiste em duas palavras ortográficas, mas é uma única palavra fonológica (é pronunciado como uma unidade), um único item lexical (é objeto de uma entrada à parte no dicionário) e uma única forma de palavra gramatical, pois não tem uma outra forma: parece difícil imaginar contextos em que faria sentido usar *granás pretas*, no plural.

O singular *cão* e o plural *cães* são duas palavras ortográficas diferentes, duas palavras fonológicas diferentes e duas formas gramaticais de palavra diferentes, mas representam o mesmo item lexical (...). O mesmo vale para *tomar*, *toma*, *tomou*, *tomado*, *tomando*; cinco palavras ortográficas, cinco palavras fonológicas, cinco formas gramaticais de palavra, mas apenas um item lexical. As entradas de dicionário dos dois itens lexicais usados aqui como exemplo seriam *cão* e *tomar*; são essas as formas usadas para citar esses itens lexicais, isto é, as formas que usamos para nomeá-los ou para falar deles.

Na contração *àqueles* (*para aqueles*), há uma única palavra ortográfica e uma única palavra fonológica, mas há dois itens lexicais (a preposição *a* + o demonstrativo *aqueles*). A perífrase verbal *passar sem* (*sobreviver apesar da falta de*, como em *Posso passar sem comer açúcar por algum tempo*) compreende duas palavras ortográficas, duas palavras fonológicas, mas apenas um item lexical: devido a seu sentido imprevisível, precisa ser objeto de uma entrada separada no dicionário. E tem várias formas gramaticais de palavras: *passar sem*, *passei sem*, *passa sem* etc. O sentido bem diferente de *passar sem* exemplificado por *O dia passou sem incidentes* seria considerado pela maioria dos linguistas como um item lexical diferente do anterior, exigindo uma entrada lexical independente, mas esse item lexical apresenta as mesmas formas ortográfica, fonológica e gramaticais que o primeiro.

Em suma, quando estivermos falando de palavras, é essencial que especifiquemos exatamente que sentido temos em mente, e pode ser preferível usar uma das denominações mais específicas apresentadas neste verbete. (TRASK, 2004, p. 218-219, grifos do autor).

Estamos, assim, diante de quatro definições que, na verdade, definem entidades diferentes. Não são quatro definições de um mesmo objeto teórico, mas quatro objetos distintos, cada qual com sua definição, como já dissemos anteriormente.

O mais interessante é que essas quatro entidades pertencem a universos diferentes. Como vimos, a palavra ortográfica só tem existência na escrita, e a noção que lhe corresponde é fruto de uma linguística ingênua; a palavra fonológica só tem existência na fala e quaisquer correspondências que possamos propor com as palavras ortográficas, os itens lexicais ou as palavras morfossintáticas – e mesmo com entidades “misteriosas” como *ideias* – estão fadadas ao fracasso (na melhor das hipóteses, vamos ter de inventar dezenas de critérios secundários para que alguma aproximação seja possível). O par de conceitos, item lexical e palavra morfossintática, ressuscitado pela morfologia gerativa, na medida em que, no fundo, reproduz o modelo tradicional palavra-e-paradigma, vai carregar consigo todas as virtudes e todos os defeitos do tratamento tradicional.

Só para finalizar esse item, creio que vale a pena mostrar que a noção de palavra morfossintática ainda precisa ser mais bem compreendida. Como vimos anteriormente, a noção de **palavra** não é a mesma conforme o modelo de análise morfológica adotado. A noção de palavra que nos surge no modelo palavra-e-paradigma não corresponde (ou só corresponde

parcialmente) à noção de palavra decorrente do modelo item-e-arranjo ou do modelo item-e-processo. E a diferença que observamos aqui reside justamente nos itens lexicais e nas palavras morfossintáticas. Explico.

No modelo item-e-arranjo (IA), a unidade é o morfema. Logo, os verdadeiros “itens lexicais” são os morfemas. A **palavra** (que não é mais um nível necessário para a teoria) passa a ser apenas um arranjo específico de morfemas (uma estrutura sintática cujos elementos são morfemas que ocupam posições em um arranjo, em uma construção que não é muito diferente de uma estrutura sintagmática). No caso em que a teoria reconheça a palavra como um nível de análise, a distinção entre palavra variável e invariável perde completamente o sentido: todas as palavras seriam invariáveis. A ideia de que as palavras morfossintáticas são variantes alternativas de algum item lexical (= **palavra**, no sentido do modelo palavra-e-paradigma) vai depender exclusivamente de estipulações do linguista: o linguista precisa **estipular**, com alto grau de arbitrariedade, que dois arranjos de morfemas são variantes de uma mesma palavra, já que não me parece haver critério de decisão confiável.

No modelo item-e-processo (IP), a unidade será sempre um elemento básico, que tanto pode ser uma raiz quanto uma palavra (a depender de a teoria²² conseguir delimitar uma denotação não-ambígua para o termo **palavra**). Esse elemento básico sofre uma operação que dá origem a uma nova forma – não básica – que, por sua vez, pode sofrer uma nova operação, e assim por diante em um processo recursivo. Se o elemento básico é uma palavra, o resultado da operação também é uma palavra (não-básica); se o elemento básico é uma raiz, o resultado da operação poderá ser chamado de **palavra** (embora não necessariamente, já que poderíamos precisar de mais de uma operação para chegar a algum resultado que se aproximasse do que se chama normalmente de palavra).

A NOÇÃO DE PALAVRA É UM SÉRIO PROBLEMA PARA A LINGUÍSTICA

Como se não bastasse a ambiguidade do termo **palavra**, para teorias linguísticas de caráter universalista, como a gramática gerativa, por exemplo, a noção de palavra teria de ser definida, uniformemente, **para todas as línguas**.

Logo no início de seu texto de 2011, Martin Haspelmath, professor do Instituto Max Plank, de Leipzig, diz:

Neste artigo, eu afirmo que não temos uma boa resposta à questão de como definir a noção de palavra de modo claro e consistente que corresponda a nossas intuições e à prática convencional, a despeito de décadas de pesquisa que se voltaram à questão (cf. Krámský 1969, Juilland & Roceric 1972 para trabalhos mais antigos). Os linguistas geralmente empregam um conjunto de critérios diferentes, mas esses critérios não são uniformemente aplicáveis em distintos contextos ou distintas línguas, e quando são aplicáveis, eles nem sempre convergem. [...].

²² O modelo IP surgiu no quadro teórico da linguística histórico-comparativa e destinava-se a expor como as **raízes** iam sofrendo, historicamente, processos sucessivos que as transformavam nas palavras das línguas descendentes. Nas versões mais recentes (estruturalistas ou categoriais), a preocupação deixa de ser histórica e a preocupação fundamental são as relações entre as **palavras** – particularmente, como vimos anteriormente, nos modos pelos quais conjuntos de palavras são derivadas de bases comuns pela aplicação de certos processos.

Mas se as palavras não podem ser facilmente identificadas, por que nós trabalhamos com elas o tempo todo? Há duas explicações possíveis para isso: por um lado, pode ser que as palavras existam, apesar de tudo, e apenas nós ainda não encontramos um jeito de identificá-las consistentemente. Mas outra possibilidade que deve ser tomada seriamente é que a ideia de palavras universais se deva a indícios tomados da linguagem escrita e da forte influência do hábito de separar palavras com espaços nas línguas ocidentais que tem estado conosco por cerca de mil anos. Em todas as línguas que usam alfabetos derivados do grego (Grego, Latim, Russo etc.) e em muitas outras línguas influenciadas por elas, espaços em branco tornam as palavras ortográficas unidades muito salientes da língua escrita. No entanto, os linguistas superaram as influências da língua escrita em outros aspectos (e.g. distinguindo cuidadosamente sons e letras desde o século 19), assim, é tempo para que reestabelecamos as evidências para a noção de palavra de modo livre de influências. É certamente possível que possamos eventualmente descobrir evidências para alguma coisa como uma noção de palavra válida para todas as línguas, mas, no presente estágio de nosso conhecimento, não temos essas evidências, como veremos. (HASPELMATH, 2011, p. 32-33, tradução nossa).

Com base em dados de mais de 30 línguas e em bibliografia substancial, Haspelmath mostra que as propostas de definição de palavra encontradas na literatura são inconsistentes entre si e todas parecem incapazes de dar uma resposta adequada à questão do que é uma palavra. Mesmo quando algum critério parece adequado para a identificação de palavra em alguma língua particular, de modo geral é incapaz de identificar a noção em outras línguas. Haspelmath mostra, também, que o uso de critérios combinados não resolve o problema, já que a seleção dos critérios é particular de cada pesquisador e não há, em consequência, nenhuma possibilidade de comparar as diversas combinações propostas. Em suas palavras:

Alguns autores percebem a não coincidência dos critérios, e.g. Börjars (1998:44): *“The behaviour of elements is often not totally consistent. This means that in order to arrive at the conclusion that an element is either a clitic or an affix, certain criteria must be assumed to be less crucial.”* [O comportamento dos elementos é com frequência não totalmente consistente. Isso significa que para chegar à conclusão de que um elemento é um clítico ou um afixo, certos critérios devem ser assumidos como menos cruciais]. Mas, se diferentes autores assumem que diferentes critérios são ‘menos cruciais’, nenhum acordo pode ser alcançado. (HASPELMATH, 2011, p. 59, tradução nossa).

Aparentemente, a noção de palavra tem de ser definida, alternativamente, (i) de forma particular para cada língua, sem que haja possibilidade de uma noção universal²³, ou (ii) como uma noção prototípica (nebulosa, “fuzzy”), em que haveria membros típicos do conjunto denotado pelo termo (membros com alta probabilidade de pertencimento) e membros com probabilidades menores de pertencimento, todos situados em um *continuum*. Nessa segunda alternativa, teríamos, além de expressões que seriam claramente “palavras” e expressões que seriam claramente “não-palavras”, coisas como **semi-palavras**²⁴, que poderiam ser mais próximas dos morfemas,

²³ Como pensa Martinet (1972, p. 117), por exemplo, quando diz que “[d]e balde se procuraria definir com maior rigor a noção de ‘palavra’ em linguística geral”.

²⁴ O conceito de semi-palavra e os dois exemplos apresentados são atribuídos por Haspelmath a Wolfgang Wurzel (cf. HASPELMATH, 2011, p. 61).

como é o caso do islandês *hund-s-in-s* [cão-GEN-DEF-GEN] ‘do cão’, que apresenta concor- dância interna de caso, ou mais próximas dos sintagmas, como o inglês *the Queen of England’s mother* [A mãe da Rainha da Inglaterra], com flexão localizada na fronteira de um sintagma que funciona como uma palavra²⁵.

As duas alternativas são ruins. Considerar que as línguas particulares estabelecem noções próprias, específicas, de palavra (sempre com a possibilidade de que alguma língua não tenha nada que se possa chamar de **palavra**) significa renunciar ao caráter universal da gramática, ao menos em alguns aspectos importantes. Significa, também, entender que a distinção entre **morfologia** e **sintaxe**, uma vez que a noção de palavra é crucial para essa distinção, teria de ser estabelecida de forma particular para cada língua e, conseqüentemente, não poderiam ser módulos de uma gramática universal. Considerar que a noção de palavra é prototípica signifi- caria colocar a **morfologia** e a **sintaxe** como polos de um *continuum*, limitando sensivelmente a possibilidade de estabelecer uma distinção clara entre os dois módulos. Não vejo que essas alternativas sejam impossíveis, mas tenho claro que a tentativa de implementação de qualquer delas implicaria uma reformulação substancial da arquitetura da gramática (e, até onde sei, não existem propostas dessa reformulação).

CONCLUSÃO

Quero concluir este texto com a discussão de uma pergunta que Haspelmath faz no trecho que citamos anteriormente: “Mas se as palavras não podem ser facilmente identificadas, por que nós trabalhamos com elas o tempo todo?”. Haspelmath dá uma dupla resposta à questão. De um lado, ele admite a possibilidade de que existam, de fato, palavras nas línguas e que o problema é que ainda não conseguimos chegar a uma definição satisfatória. Trata-se de uma resposta de caráter realista: coisas, como **palavras**, existem no mundo e nossa tarefa é descobri-las, defini-las e tratá-las teoricamente. Por outro lado, ele admite a possibilidade de que a noção de palavra se deva a “[...] indícios tomados da linguagem escrita e da forte influência do hábito de separar palavras com espaços” (HASPELMATH, 2011, p. 32, tradução nossa). Nesse caso, temos uma resposta de claro caráter antirrealista: somos nós, nossos hábitos, que construímos as coisas do mundo. Eu diria que Haspelmath se equivoca e que a questão que está por detrás do uso extensivo da noção de palavra se deve à **naturalização das teorias tradicionais**, cuja ontologia é, de forma geral, assumida acriticamente²⁶.

Os antigos gregos e latinos viram nas línguas o que podiam ver, o que suas perspectivas ideológicas, seus sistemas de pensamento, permitiam que fosse visto. Ou mais precisamente, o que lhes parecia certo, justo e útil ver nas línguas. E uma noção como a de *palavra* certamente foi útil para a construção de uma primeira teoria das línguas. Essa teoria greco-latina das línguas, embora não tenha passado incólume pelo tempo, foi em grande parte mantida até os nossos dias. Mais do que mantida, a teoria greco-latina foi **naturalizada**. Com isso, quero dizer que a pos- teridade considerou que as entidades assumidas pela teoria greco-latina, independentemente de suas definições, eram de fato entidades do mundo e que, portanto, precisavam ser apresentadas, definidas, classificadas, tratadas, enfim, por qualquer outra teoria que se debruçasse sobre as

²⁵ Se não entendêssemos que “Queen of England” funciona como uma unidade (uma “palavra”), a tradução da expressão para o português deveria ser “a rainha da mãe da Inglaterra”, com o possessivo ligando exclusivamente “mother” à expressão “England”.

²⁶ Sobre a naturalização, ver Borges Neto (2011, 2012a, 2012b).

línguas. Os modistas medievais discutiram as definições e as classificações que Prisciano dava às entidades gramaticais do latim, mas jamais discutiram a existência ou não dessas entidades. Os Senhores de Port-Royal reviram o sistema gramatical e estabeleceram novos critérios de definição, mas nunca propuseram a possibilidade de não existência de sílabas, palavras, nomes e verbos, flexões e derivações etc.

Até mesmo Haspelmath (2011, p. 32, tradução nossa) cai na armadilha da naturalização ao dizer que “[...] não temos uma boa resposta à questão de como definir a noção de palavra [...] que corresponda a nossas intuições e à prática convencional”. O que se poderia entender por “nossas intuições”, senão o que aprendemos sobre palavras em nosso processo de escolarização?²⁷ E como definir **palavra** de modo que cheguemos à “prática convencional” senão assumindo que a entidade ontológica que buscamos definir é exatamente a que herdamos dos tratamentos tradicionais (ou de uma linguística ingênua)? Parece claro que, no fundo, o coração de Haspelmath pende para a perspectiva realista.

Na medida em que a noção de palavra foi naturalizada, todas as abordagens teóricas da linguagem deverão considerá-la um dos elementos constituidores da gramática. E o mais interessante é que mesmo aqueles linguistas que dispensam a noção de palavra acabam por dar um jeito de recuperar os seus efeitos por meio de hipóteses auxiliares, estipulando que formas seriam **de uma mesma palavra**, por exemplo.

Enquanto uma noção de palavra, digamos, “tradicional” seguir sendo útil para a proposição de regras gramaticais, os linguistas simplesmente esquecerão sua indeterminação e continuarão a propor construtos teóricos frágeis em sua fundamentação. Infelizmente!

REFERÊNCIAS

ALMEIDA, N. M. **Gramática Metódica da Língua Portuguesa**. 28. ed. São Paulo: Saraiva, 1979.

ARISTÓTELES. **Tratados de Lógica (Órganon)**. Tradução Miguel Candel Sanmartín. Madrid: Gredos, 1988. v. II.

BARBOSA, J. S. **Grammatica Philosophica da Lingua Portuguesa**. Lisboa: Typographia da Academia, 1822.

BORGES NETO, J. Um caso de redução de ditongos em português. *In*: UNESP. **Estudos Linguísticos III** – Anais de seminários do GEL. Araraquara: Unesp; GEL, 1980. p. 227-233.

²⁷Note-se que Haspelmath (2011, p. 34-35) tem consciência disso, já que afirma: “[...] de acordo com Aronoff & Fudeman (2005:36), ‘speakers – literate and illiterate – have clear intuitions about what is and what isn’t a word’ (similarmente, Langaker 1972: 36, Bauer 1988: 45, Himmelman 2006: 255). Mas não é clara a base desse otimismo. Falantes não-letrados em alguma língua (o caso para o melhor teste) estão progressivamente se tornando raros, e os linguistas de campo que não esperam encontrar palavras não são comuns também. Além disso, os falantes precisam ser capazes de comunicar suas intuições, mas como notam Dixon & Aikhenvald (2002:3), ‘the vast majority of languages spoken by small tribal groups . . . have a lexeme meaning “(proper) name”, but none have the meaning “word”’. Nas línguas europeias, também, podemos ver que as palavras correntes para ‘palavra’ (e.g. *word*, em inglês, *mot*, em francês, *slovo*, em russo, etc.) tiveram originalmente significados mais gerais (atos de fala) e parece terem adquirido o sentido mais restrito do grego *léxis* apenas por meio da formação escolar, particularmente escrevendo e ensinando gramática (ver Haebler 2002). É ainda, certamente, uma questão aberta quais as intuições que os falantes não-letrados têm, mas é claro que os falantes letrados de línguas com sistemas de escrita com separação de palavras não têm intuições que sejam independentes das regras de escrita que aprenderam”.

BORGES NETO, J. Morfologia: conceitos e métodos. *In*: Lima, M. A. F.; Alves Filho, F; Costa, C. S. C. (org.). **Colóquios linguísticos e literários: enfoques epistemológicos, metodológicos e descritivos**. Teresina: Edufpi, 2011. p. 53-72.

BORGES NETO, J. Gramática tradicional e linguística contemporânea: continuidade ou ruptura? **Todas as Letras**, São Paulo, v. 14, n. 1, p. 87-98, 2012a.

BORGES NETO, J. Flexão e derivação: será que os tratamentos dados a esta distinção não adotam uma perspectiva “enviesada”? **Cadernos de Estudos Linguísticos**, Campinas, v. 54, n. 2, p. 7-317, 2012b.

CÂMARA JUNIOR, J. M. **Problemas de linguística descritiva**. Petrópolis: Vozes, 1970.

CHAPANSKI, G. **Uma tradução da Tékhne Grammatiké, de Dionísio Trácio, para o português**. 2003. Dissertação (Mestrado em Letras) – Setor de Ciências Humanas, Universidade Federal do Paraná, Curitiba, 2003. Disponível em: <https://acervodigital.ufpr.br/handle/1884/24818>. Acesso em: 17 nov. 2020.

DUCROT, O.; TODOROV, T. **Dicionário das Ciências da Linguagem**. Tradução Antônio José Massano, José Afonso, Manuela Carrilho e Margarida Font. 3. ed. Lisboa: Dom Quixote, 1976.

GONÇALVES, C. A. **Iniciação aos Estudos Morfológicos: flexão e derivação em português**. São Paulo: Contexto, 2011.

HASPELMATH, M. The indeterminacy of word segmentation and the nature of morphology and syntax. **Folia Linguistica**, [s. l.], v. 45, n. 1, p. 31-80, 2011. DOI: <https://doi.org/10.1515/flin-2017-1005>

HOCKETT, C. F. Two models of grammatical description. **Word**, [s. l.], v. 10, n. 2-3, p. 210-234, 1954. DOI: <https://doi.org/10.1080/00437956.1954.11659524>

HOEKSEMA, J.; JANDA, R. D. Implications of process-morphology for categorial grammar. *In*: OHERLE, R. T. *et al.* (ed.). **Categorial grammars and natural language structures**. Dordrecht: Reidel, 1988. p. 199-247.

LALLOT, J. **La grammaire de Denys le Thrace: traducción anotée**. Paris: Éditions du Centre National de la Recherche Scientifique, 1989.

MARTINET, A. **Elementos de linguística geral**. 4. ed. Tradução Jorge Morais-Barbosa. Lisboa: Sá da Costa, 1972.

ROBINS, R. H. **Pequena história da linguística**. Rio de Janeiro: Ao Livro Técnico; Brasília: INL, 1979.

SCHWINDT, L. C. Morfologia. *In*: SCHWINDT, L. C. (org.). **Manual de Linguística: Fonologia, Morfologia e Sintaxe**. Petrópolis: Vozes, 2014. p. 109-154.

TRASK, R. L. **Dicionário de Linguagem e Linguística**. Tradução e adaptação Rodolfo Ilari. São Paulo: Contexto, 2004.

Recebido em: jul. 2020.

Aceito em: jul. 2020.