



## CLASSIFICAÇÃO MORFOLÓGICA DAS LÍNGUAS: DA TIPOLOGIA HOLÍSTICA ÀS DEFINIÇÕES CIRCULARES

### MORPHOLOGICAL CLASSIFICATION OF LANGUAGES: FROM HOLISTIC TYPOLOGY TO CIRCULAR DEFINITIONS

*Eudes Barletta Mattos<sup>1</sup>*

*João Paulo Lazzarini Cyrino<sup>2</sup>*

#### RESUMO

No presente artigo, realizamos uma investigação teórico-histórica de duas grandes inconsistências na tipologia linguística, inseridas na interface morfologia/sintaxe: o uso da noção de palavra (DIXON; AIKHENVALD, 2003) e a classificação morfológica em tipos holísticos (SCHWEGLER, 1990), ambas relacionadas e interdependentes. Intentando avaliar suas origens e superação – ou não –, fazemos uma revisão bibliográfica, acompanhando os primórdios da classificação morfológica (VON SCHLEGEL, 1818), surgida em meio a discussões acientíficas sobre o espírito dos povos e pureza racial. Em seguida, aludimos à inadequação de imensa parte do cabedal teórico europeu às diferentes línguas, com atenção ao caso do mandarim entre as línguas siníticas. Daí, debruçamo-nos sobre propostas mais recentes de estudos no campo da tipologia. Percebemos claramente a persistência do conceito de palavra e seu caráter estruturante para a linguística, mesmo nas propostas mais recentes, bem como utilização disseminada das classificações morfológicas da tipologia holística. Por fim, apontamos alternativas ao uso da palavra como item de análise, bem como outras maneiras de comparar línguas que não as subsumindo inteiramente a tipos morfológicos discretos.

**Palavras-chave:** Tipologia linguística; Morfosintaxe; Palavra.

#### ABSTRACT

In this paper, we present a theoretical-historical investigation on two major inconsistencies in linguistic typology, both right at the interface between morphology and syntax, such as using the notion of word (DIXON; AIKHENVALD, 2003) and the morphological classification by holistic types (SCHWEGLER, 1990), both closely interconnected, to evaluate their origins and whether these notions have been overcome or not. For this purpose, we peer into the beginning of morphological classifications (VON SCHLEGEL, 1818), which arose amidst unscientific discussions on the spirit of peoples and racial purity; afterward, we allude to the inadequacy of a substantial share of the European theoretical body as to the linguistic diversity of the world, paying special attention to the case of Mandarin as representative of Sinitic languages; then, we follow onto analyzing some more recent proposals on linguistic typology studies. We clearly realize the persistence of the concept of word and its structuring character for linguistics even in recent proposals, as well as the pervasive use of the morphological classifications of holistic typology. Finally, we point to some alternatives to the use of the word as an analysis item, as well as other ways of comparing languages without entirely subsuming them to discrete morphological types.

**Keywords:** Linguistic typology; Morphosyntax; Word.

1 Graduando em Letras, Universidade Federal da Bahia, eudesbarlettam@gmail.com

2 Professor, Instituto de Letras, Universidade Federal da Bahia, jpcyrino@gmail.com

## Introdução

O estudo comparativo das línguas na tradição científica europeia teve sua existência marcada por fortes vieses religiosos, culturais, étnicos e políticos que até hoje permanecem finamente entremeados ao corpo teórico da linguística, produzindo efeitos a despeito de o fazer científico assumir publicamente uma face – discutível – de isenção. O nascimento deste campo de conhecimento deu-se a partir do encontro de dois vetores: o primeiro, a gramática greco-latina e sua perspectiva apoiada na palavra como unidade de análise e no paradigma como forma de organização; e o segundo, mas não menos influente, a tradição judaico-cristã. Datam do primeiro século de nossa era as especulações sobre a “língua adâmica” – supostamente falada por Adão e Eva no mito bíblico do Gênesis. Este se conservaria tema central para a discussão sobre língua na Europa por dezessete séculos, até a chegada do século XVIII (MAZZOCCO, 1993; PARRY, 2015). Com o colonialismo e a instituição do Raj Britânico na Índia, o aparecimento do sânscrito nos círculos intelectuais europeus do século XIX fomenta uma série de estudos em linguística comparativa, envolvendo o agrupamento das línguas do mundo, primeiramente por ancestralidade, e em seguida, por atributos em comum.

A tentativa científica de categorizar línguas com base em características tipológicas toma os primeiros contornos fortemente atrelada a sentimentos nacionalistas e etnocêntricos, em especial os despertados pelo romantismo alemão. É nesta episteme comprometida por ideologias de pureza racial e superioridade étnica que se consolida a classificação linguística com base em um perfil morfológico holisticamente aplicável. É dizer, cada língua em sua totalidade pertenceria a uma determinada categoria, pois que a afiliação a um tipo seria a expressão visível, no plano morfológico, de uma suposta *essência* não só da língua, como também de sua comunidade de falantes (MAZZOCCO, 1993).

A despeito do rechaço das ideias de pureza e superioridade racial do século XIX, bem como da tipologia holística na esfera da produção científica mais especializada, permanece entre muitos acadêmicos da linguística a ideia vaga de que as línguas apresentariam consistentemente tais ou tais características morfossintáticas em sua totalidade. Não raro, estudantes, pesquisadores e interessados em linguística no geral expressam das mais variadas formas a noção de que há algo de *essencialmente* diferente entre os seguintes exemplos clássicos de tipos morfológicos aos quais as línguas poderiam se encaixar:

- (1) Mandarim (sino-tibetana: sinítica)  
lóng zài yún lǐ fēixiáng  
dragão em nuvem dentro voar  
“O dragão está voando nas nuvens”
  
- (2) Turco (túrquica: oguz)  
Burak Doğuş-a balıĝ-ı tut-tur-du  
Burak Doğuş-DAT PEIXE-ACC PEGAR-CAUS-PST  
“BURAK FEZ DOĞUŞ PEGAR O PEIXE”

- (3) Francês (indoeuropeia: itálica)  
il y a pas de nuage-s dans le ciel  
3SG LOC haver.3SG neg de nuvem-PL em MASC.SG céu  
“Não há nuvens no céu”
- (4) Latim (indoeuropeia: itálica)  
non est nub-es in cael-o  
NEG ser.3SG nuvem-NOM.PL em céu-ABL.SG  
“Não há nuvens no céu”

Da observação comparativa de dados como (1) e (4), linguistas têm feito classificações das línguas em *tipos morfológicos* há dois séculos. Nas palavras de Edward Sapir,

“há algo de irresistível quanto a um método de classificação que comece com dois pólos exemplificados por, digamos, chinês e latim, agrupe o que convenientemente puder a estes pólos, e jogue todo o resto num ‘tipo transicional’” (SAPIR, 1921, p. 130) [tradução nossa].

Sob a aparência relativamente descomplicada destas operações classificatórias, porém, repousam uma série de conceitos cujas existências foram postuladas com base em tradição, conhecimento religioso ou ideologia externa à linguística, e cujas validades merecem ser postas à prova.

Portanto, este artigo faz uma revisão bibliográfica sobre o percurso da classificação linguística, desde seu início como alegada expressão da *essência* de uma língua – e por isso, da inteligência coletiva de seus falantes – até um sistema classificatório que, embora reconfigurado, carrega em seu cerne conceitos que demandam ser discutidos: as definições circulares de palavra e a oposição morfologia/sintaxe. Organiza-se em quatro seções: na primeira, acompanhamos o histórico da **tipologia morfológica**, de Dante Alighieri no séc. XIV, passando por nomes como August von Schlegel e Edward Sapir, a Joseph Greenberg, já na segunda metade do séc. XX; na segunda seção, discutimos a problemática do conceito de **palavra**, sobre o qual muito da classificação das línguas tradicionalmente se baseou; na terceira, é apresentada a importante contribuição do conceito de **formativo** para a análise dos fenômenos gramaticais de forma não-holística, muito embora ainda calcada na palavra; e na quarta e última, apontamos outras perspectivas.

## Breve histórico crítico da tipologia morfológica

### Os primórdios da classificação no ocidente

Dante Alighieri é um dos expoentes do eruditismo europeu que, no século XIV, reforça o debate sobre a língua adâmica, descrevendo-a como divina, imutável, estável e universal, independente dos caprichos dos falantes humanos (MAZZOCCO, 1993). Desta língua divina,

todas as demais línguas do mundo teriam surgido como corrupções degeneradas depois do incidente da confusão babilônica – a queda a torre de Babel, também ficção bíblica. Segundo Dante, após tal episódio, a língua adâmica teria dado origem a três outras, a germânica, a românica e a bizantina, que subsequentemente teriam passado por subdivisões internas próprias (ALIGHIERI, 1996), gerando a diversidade de línguas conhecidas até então. Em sua obra *De Vulgari Eloquentia*, Dante estabelece o conceito de *gramatica* – a face imutável da língua, construída para a comunicação padronizada. De forma análoga, muitos proponentes do conceito de língua adâmica, eruditos ou não, acreditavam que tal língua, por divina, seria capaz de exprimir a *verdadeira essência* das coisas de uma maneira que as línguas humanas então contemporâneas, sendo degenerações, não poderiam (PARRY, 2015). As discussões quanto à natureza da língua adâmica prosseguiriam até o século XVIII, quando foram em sua maior parte abandonadas. As noções de essência e de corrupção de um passado ideal, no entanto, estavam plenamente enraizadas no pensamento sobre a linguagem, e assim prosseguiriam até o século XIX.

### O século XIX e Von Schlegel

O século XIX marcou de forma profunda as ciências humanas. Disseminados entre o fazer científico e a discussão política europeia da época, encontram-se diversos elementos oriundos do pensamento erudito alemão. Um grande eixo ideológico, norteador do grosso das produções intelectuais daquele país, foi o conceito de *Volksgeist*, ou espírito de um povo. Para além dos círculos acadêmicos e literários alemães, a noção de *Volksgeist* influenciou sobremaneira a epistemologia hegemônica desde o momento de sua concepção. Baseado no *Nationalgeist* de Justus Möser e Johann Gottfried Herder, pensadores do século XVIII, é no século subsequente que o conceito se populariza entre os acadêmicos de vários campos do conhecimento. A despeito de particularidades de cada área que influenciou, pode-se afirmar de forma geral que o *Volksgeist* é a ideia de que haveria um gênio em cada povo. O espírito de um povo seria, portanto, sua *essência* psicológica, presente em todos os seus membros, ao mesmo tempo servindo de motriz para sua história e se revelando em suas várias expressões – tais quais língua, mitologia, religião e costumes. Do direito, com Friedrich von Savigny, à filosofia, com Friedrich Hegel, passando por Wilhelm Wundt, Moritz Lazarus and Heymann Steinthal sob a forma da *Völkerpsychologie* – a psicologia de um povo – (STOCKING, 1996), também ganha corpo entre a linguística e a literatura por meio dos trabalhos de expoentes como Wilhelm von Humboldt, Friedrich e August Schlegel, Jakob e Wilhelm Grimm e Franz Bopp.

Armin Schwegler (1990) analisa minuciosamente a discussão analiticidade/sinteticidade desde sua proposição por expoentes do pensamento alemão até a contribuição de Joseph Greenberg, já no século XX. Para compreender as subsequentes propostas e o impacto destas ideias na linguística moderna, observemos aqui seu percurso investigativo a partir de August von Schlegel no início do século XIX. Em sua obra *Observations sur la langue et la littérature provençales*

(1818), von Schlegel propõe classificar as línguas por características morfológicas; tem-se aí um dos primeiros impulsos de categorizar as línguas de forma não genética com base em similaridades de comportamento morfológico. A classificação de von Schlegel levava em consideração a estrutura das palavras de uma dada língua, sendo tríplice; em primeiro lugar, estariam as línguas que, segundo o autor, seriam desprovidas de estrutura gramatical. Suas palavras seriam compostas por raízes puras, e as relações entre elas se dariam puramente por meio da sintaxe. O exemplo canônico deste tipo, desde a época de sua proposição, são as línguas siníticas, como o mandarim.

O segundo tipo de línguas seria o das línguas que codificariam as relações gramaticais por meio do emprego de afixos, a exemplo do turco; as descrições, então recentes, das línguas indígenas das Américas também serviriam de exemplo deste tipo morfológico:

- (5) Quéchua meridional de Conchucos (quéchua: quéchua I)  
wamra-qa                      aywa-ku-sha    tullu-n apa-ri-sh  
criança-TOP    ir-MID-PST.R3    osso-3    pegar-PUNC-PTCP  
“A garota escapou carregando os ossos” (HINTZ, 2007)

Por fim, viria o terceiro tipo de von Schlegel, as línguas com flexão; nestas, as relações gramaticais se dariam por modificações de “letras” [termo do autor] nas raízes, bem como adição de sufixos derivacionais. O autor divide este tipo em dois subtipos, fundando uma distinção que é bastante difundida até hoje nos estudos em morfologia: a de línguas sintéticas e analíticas. As línguas flexionais **sintéticas** seriam, para von Schlegel, o tipo mais “perfeito” de expressão linguística humana, do qual todos os outros tipos seriam corrupções; nele, as relações gramaticais se estabelecem com base em modificações nas palavras em si. O exemplo típico é o das línguas clássicas europeias, como o grego antigo e o latim. Já as **analíticas** representariam uma degeneração das línguas sintéticas, muito embora preservassem em parte o caráter flexional – e por isso ainda fossem mais desenvolvidas, segundo ele, que as línguas dos outros dois tipos. O autor define tais línguas como as que utilizam artigos antes dos substantivos, pronomes pessoais antes dos verbos, recorrem a auxiliares na conjugação verbal, a advérbios para modificar o grau de adjetivos, entre outras características típicas de línguas europeias modernas, como as línguas românicas.

Tal classificação não foi proposta sem apresentar um juízo de valor extremo, em consonância com o nacionalismo da época. Para von Schlegel, as línguas do primeiro tipo, como o chinês, seriam “estéreis”, desprovidas da vida orgânica típica das línguas mais desenvolvidas, havendo para ele uma relação direta entre estrutura morfológica da língua e inteligência humana coletiva. Ainda segundo o autor alemão, a mudança ocorrida nas línguas europeias, do subtipo sintético para o analítico, teria se dado por conta do declínio cultural advindo com o fim da era clássica e a queda do Império Romano, já que as línguas sintéticas pertenceriam “a outra fase da inteligência humana” (VON SCHLEGEL, 1818). Portanto, não é exagero afirmar que desde sua concepção, a classificação das línguas com base em tipos morfológicos foi eivada de julgamentos de natureza pouco científica, de cunho etnocêntrico e discriminatório. Mais do

que uma preocupação puramente linguística, a proposta classificatória deixa entrever ideais de superioridade étnica, racial e cultural, e uma busca pelo estado puro, incorrupto, das línguas. Tal quadro ecoa os primeiros esforços, inextricavelmente ligados ao cristianismo, de determinar a primeira língua falada pela humanidade, mas não é balizado em fenômenos naturais.

O processo de construção de uma identidade nacional pelo qual a Alemanha do século XIX passava deu o tom ao ambiente acadêmico daquele país, e von Schlegel, como expoente do romantismo alemão, teve sua produção científica profundamente investida neste contexto. Desde o renascimento, os humanistas alemães buscavam uma imagem de si como uma nação pura, forte, em contraste com a decadência de Roma. É no século XIX, porém, que a identidade alemã vai sendo estabelecida em oposição à França napoleônica; é dito que muitos alemães, intelectuais ou não, definiam-se como sendo o que os franceses não eram (VICK, 2003). Vê-se aí o tema da pureza original, caro ao pensamento europeu fundado como um meio-termo entre a herança clássica e a tradição judaico-cristã. Não é por acaso, portanto, que a dicotomia sintética/analítica, como proposta por von Schlegel, tratasse de línguas clássicas, representantes de um estado anterior puro, *versus* suas descendentes, frutos de uma suposta degeneração – entre estas últimas, o francês e demais línguas românicas, contra as quais os sentimentos nativistas alemães se levantavam. Também não é à toa que von Schlegel situe as línguas germânicas (à exceção do inglês, por ele considerada analítica) como intermediárias – sintéticas na origem, embora “tendendo à analiticidade”.

Digno de nota é que Wilhelm von Humboldt, muito embora igualmente afinado ao romantismo alemão e às influências da concepção de *Volksgeist* à discussão linguística, discorda da proposta de von Schlegel. A discordância não é só quanto à dimensão evolutiva proposta – Humboldt rejeita a evolução das línguas flexionais a partir de línguas com morfologia tida por menos rica –, como também quanto à própria necessidade de tal classificação. Para ele, uma mesma língua poderia apresentar construções sintéticas e analíticas lado a lado, ponto de vista também defendido por Georg von der Gabelentz; ademais, ainda para Humboldt, a classificação não-genética das línguas seria impossível (SCHWEGLER, 1990).

### **A primeira metade do século XX – Sapir e a reintrodução dos conceitos**

No século XX, a proposta classificatória de von Schlegel, esquecida por praticamente meio século, foi revivida por Edward Sapir. Em sua obra *Language: An Introduction to the Study of Speech* (1921), Sapir torna a não só abordar a tipologia morfológica na classificação das línguas, como também a usar as palavras *analítica* e *sintética* propostas por von Schlegel – mas não sem uma postura crítica quanto ao espírito acientífico que tinha dominado o século anterior.

Sapir afirma que manter noções preconcebidas de superioridade de uma ou outra língua – e com base nisto, classificar as línguas do mundo – é um esforço infrutífero, além de possivelmente o entrave mais poderoso de todos ao pensamento lúcido na linguística. Em seu texto, ele afirma diretamente que há um preconceito evolucionário influenciando a classificação científica das

línguas, por conta da admiração europeia pelas línguas flexionais clássicas – grego e latim –, equivocadamente tidas como a etapa mais elevada do desenvolvimento linguístico. A linguagem, como expressão simbólica da intuição humana, pode se apresentar, segundo ele, nas mais variadas formas, independentemente de avanço ou primitivismo materiais do povo que articule a língua.

A despeito desta análise crítica, Sapir faz sua própria proposta classificatória, com a ressalva de que tais tipos não deveriam ser aplicados nem de forma exclusiva – já que as línguas apresentam construções de mais de um tipo –, nem de forma absoluta – pois que uma construção é mais ou menos analítica ou sintética apenas em relação a outra –, sob pena de superficialidade da análise. Sua classificação então é elaborada numa grande matriz de possibilidades combinatórias entre três parâmetros principais. O primeiro é o **grau de síntese** da língua, variando de analítico – línguas que não combinariam conceitos em palavras únicas, como o chinês, ou o fariam de forma muito econômica, como o inglês; sintético – na qual os conceitos se agrupariam de forma mais definida em palavras, que seriam mais “ricamente intrincadas” [termo do autor], como no latim, árabe ou finlandês; e polissintético – que apresentaria elaboração extrema da palavra, nela codificando as relações sintáticas, a exemplo de muitas das línguas indígenas americanas. Outro parâmetro é a **técnica** empregada pela língua – isolante, sob o qual sempre há correspondência entre raiz e palavra; aglutinante, sob o qual palavras seriam formadas por junção de afixos às raízes; e fusional – sob o qual elementos relacionais se fundiriam às raízes. Por fim, o último parâmetro, o **tipo fundamental**, distinguiria as línguas simbólicas, que possuem o poder de mudar o significado dos radicais por meio de mudanças internas, das línguas relacionais, nas quais tais mudanças não operariam (SAPIR, 1921; SCHWEGLER, 1990). Sua matriz oferece mais de 20 combinações, exemplificadas por línguas desde o chinês, turco, francês e latim ao sioux, chinook e khmer.

O posicionamento de Sapir reitera a advertência feita desde von Humboldt quanto à imposição da classificação por tipos morfológicos a toda uma língua. Para as alegações de atribuição de uma dada língua em sua inteireza a um tipo, não nos é difícil imaginar contra-exemplos; as próprias línguas românicas não raro possuem uma forma analítica e uma sintética para exprimir o mesmo tempo verbal – pensemos nas formas *vou olhar* e *olharei* do português e seus equivalentes nas outras línguas aparentadas. As críticas à classificação de Sapir, porém, se concentraram na pouca clareza dos critérios para se determinar a analiticidade/sinteticidade – se deveriam ser semânticos, sintáticos, morfológicos, ou uma combinação destes; além disto, sua escala de síntese situava as línguas por meio de notações altamente vagas, como as de “levemente polissintética”, “notavelmente sintética” ou “muito próxima de complexa relacional pura” (SCHWEGLER, 1990).

### **A lista de Tauli**

Em um artigo publicado no ano de 1945, Valter Tauli oferece sua contribuição para a discussão dos conceitos de analiticidade e sinteticidade. Na publicação, expõe que os termos “analítico” e “sintético” são extensivamente utilizados na linguística, a despeito de não haver

definição amplamente aceita quanto ao que tais termos representariam. Um ponto crucial abordado por Tauli, porém, é a noção de que analiticidade/sinteticidade estão inevitavelmente ligadas ao conceito de **palavra**, um conceito que, nas palavras do autor, “é tão vago e difícil de se definir quanto os [próprios] conceitos de análise e síntese” (TAULI, 1945, p. 80). Seu trabalho, infelizmente, não obteve muita projeção, sendo ignorado por muitos dos que tomariam parte na discussão de analiticidade/sinteticidade (SCHWEGLER, 1990).

Ele elenca onze critérios para definir o grau de analiticidade:

1. linearidade, sendo uma forma linear mais analítica que uma forma não-linear;
2. presença ou não de alomorfa em bases, sendo formas invariáveis mais analíticas que formas alternativas;
3. presença ou não de alomorfa em morfemas, sendo formas invariáveis mais analíticas que formas alternativas;
4. ocorrência independente das bases, sendo as independentes mais analíticas;
5. ocorrência independente dos morfemas, sendo os independentes mais analíticos;
6. força fonética do morfema, sendo um morfema foneticamente forte - silábico, por exemplo - mais analítico que um não-silábico;
7. se os elementos de uma forma são separáveis - sendo portanto mais analíticos - ou não;
8. se a ordem destes elementos pode ser trocada - sendo assim mais analíticos - ou não;
9. se no caso de coordenação, o morfema aparece em apenas uma palavra, sendo mais analítico - ou em todas, mais sintético;
10. se no caso de subordinação, há concordância ou não, sendo expressões sem concordância mais analíticas;
11. ordem dos elementos, sendo ordem progressiva mais analítica que a regressiva.

Tauli se alinha ao posicionamento de que construções, e não línguas inteiras, seriam analíticas ou sintéticas – e de forma relativa, sempre em comparação a outras construções (muitas vezes da mesma língua), posicionáveis em uma escala de acordo com sua lista de critérios. Da análise destes, não é difícil se depreender que, por exemplo, formas não-lineares sejam mais sintéticas que forma lineares afixais que tenham raízes e morfemas invariáveis. É daí que Tauli tira uma observação curiosa: formas “aglutinantes” – sequências de morfemas lineares e invariáveis, como as que comumente aparecem em exemplos tais quais diversas línguas bantu –, teriam um caráter analítico relativamente forte. O que diferenciaria, portanto, palavras de línguas tradicionalmente grafadas como afixais das palavras de línguas tradicionalmente grafadas como isolantes, já que ambas poderiam apresentar morfemas linearmente sequenciais e sem alomorfa – e não raro, línguas historicamente consideradas isolantes como o chinês mandarim possuem ordem dos constituintes (dita sintática) tão rígida quanto a de morfemas no interior de uma palavra?

## A segunda metade do século XX – de Greenberg em diante

Joseph Greenberg (1960) buscou reduzir o grau de imprecisão na classificação morfológica das línguas, lançando mão de métodos quantitativos. Uma série de parâmetros quantificáveis – e, portanto, prontamente comparáveis entre línguas diferentes – foi proposta, com destaque para o **grau de síntese**, definido por Greenberg como o quociente  $m/w$  entre morfemas ( $m$ ) e palavras ( $w$ ). Este índice, que forneceria o grau de complexidade das palavras, situaria as línguas como analíticas – aquelas cujo quociente seria de 1,00-1,99 –, sintéticas – 2,00-2,99 – e polissintéticas – 3,00 ou mais. Adotando tais medidas, classificações prévias de línguas no espectro analítico-sintético foram em geral corroboradas: uma língua dita polissintética, o inuit, apresenta grau de síntese de 3,72; uma língua como o sânscrito, tradicionalmente vista como sintética, 2,59. O vietnamita, sempre classificado como analítico, apresenta índice de 1,06<sup>3</sup>.

Imprescindível apontar que, como aludido por Tauli, frequentemente em um sistema linguístico co-ocorrem construções mais “analíticas” e outras mais “sintéticas”, e a análise sob os parâmetros de Greenberg, depreendida de *corpora* escritos, poderia resultar em valores que encobrissem desproporções nas estratégias empregadas por uma mesma língua. Uma série de linguistas, seguindo a proposta de Greenberg, passa a propor ajustes metodológicos ao longo da segunda metade do século XX, visando corrigir tal desvio (SCHWEGLER, 1990). A maior das inconsistências, porém, é outra: embora uma boa definição de **palavra** seja axial aos cálculos que sua proposta implica, Greenberg não a define satisfatoriamente – também não o fizeram seus predecessores, e tampouco o farão os subsequentes refinamentos sobre sua proposta. Como veremos na seção seguinte, este conceito questionável é onipresente ao longo do corpo teórico da linguística, ameaçando a validade de muito do que se produz sobre o fenômeno da linguagem.

Schwegler (1990, p. 26), a respeito de subsequentes adaptações do modelo de Greenberg, afirma que após quase dois séculos de pesquisa e investigação em tipologia morfológica, resta claro que “os parâmetros sintético e analítico não podem ser aplicados a línguas em sua inteireza de nenhuma forma significativa”. Aos poucos, o enfoque holístico, ao menos nos círculos mais diretamente especializados ao estudo da tipologia, se desassociou da discussão analiticidade/sinteticidade. Outros fenômenos passam a ocupar papel central como generalizações aplicáveis à totalidade de uma língua, a exemplo da ordem canônica das frases (VSO, SVO, SOV e assim por diante); Comrie (1991, p. 440) aduz que se antes a maioria dos estudantes de linguística aprenderia a classificação morfológica das línguas em isolantes, aglutinantes e fusionais, “hoje, é provável que tenham uma ideia similarmente vaga de que línguas são classificadas em SOV, SVO e VSO”. No entanto, termos como *língua isolante*, *língua aglutinante*, *língua fusional* e *língua polissintética* enraizaram-se firmemente no imaginário coletivo dos que estudam a linguagem, ainda sendo encontrados com frequência em publicações e estudos de diversas áreas.

---

3 Na publicação de Greenberg, têm-se os termos *eskimo* e *annamite* para se referir, respectivamente, ao inuit e ao vietnamita.

## O problema da palavra

Muito foi dito quanto à natureza das línguas, das construções ou dos formativos com respeito a seu comportamento morfossintático, sempre tendo como baliza teórica a palavra. No entanto, esta é uma definição que está longe de ser unanimemente aceita – e aplicável na linguística – (DIXON; AIKHENVALD, 2003), o que põe em xeque grande parte da utilidade e poder descritivo das classificações aqui abordadas, como já afirmava Tauli em 1945.

Idiossincrasias do grego e latim clássicos acabaram por moldar profundamente o pensamento crítico acerca da língua, e a herança mais óbvia é o conceito de palavra. Nestas línguas, bem como em outras da família indoeuropeia, a palavra parece ser uma unidade de análise facilmente delimitável. A presença de formativos cumulativos<sup>4</sup> acompanhando as raízes, fenômeno desproporcionalmente concentrado em línguas desta família (DRYER; HASPELMATH, 2013), desencorajou a consciência e/ou proposição de quaisquer itens portadores de significado menores que a palavra, como o morfema. Em vez disso, os gramáticos clássicos listavam as diferentes formas que suas palavras poderiam assumir em longas listas de declinação ou conjugação, associadas a diferentes paradigmas. Veja-se:

Tabela 1. Declinação da palavra latina para nuvem – 3ª declinação, genitivo em *-ium*

Caso	Singular	Plural
nominativo	nūbēs	nūbēs
genitivo	nūbis	nūbium
dativo	nūbī	nūbibus
acusativo	nūbem	nūbēs
ablativo	nūbe	nūbibus

A segmentação interna de seu item de análise, a palavra, passou despercebida pelos gramáticos clássicos por não ser facilmente visualizável – em franco contraste com determinadas línguas tidas por “aglutinantes” como o turco ou o quéchua, cujos itens delimitáveis de análise costumam ser compostos por morfemas de conteúdo gramatical único linearmente associados.

Perceba-se que no parágrafo anterior, para explicitar os graus de adequação da palavra a diferentes línguas, aludimos ao conceito de “línguas aglutinantes”. Não só a classificação em tipos morfossintáticos depende da noção de palavra, como a própria noção de palavra – e muito de sua problematização nas últimas décadas – acabou se apoiando na classificação em tipos morfossintáticos, num espetacular exemplo de definição circular que perpassa toda uma disciplina. Para ilustrar, vejamos o que dizem Dixon e Aikhenvald ao levantar as inadequações do conceito de palavra:

4 Morfemas veiculadores de mais de um conteúdo gramatical; ver seção “O século XXI e a análise por formativos de Bickel e Nichols”.

A ideia de “palavra” como uma unidade da língua foi desenvolvida para as línguas familiares da Europa, as quais em geral possuem uma estrutura *sintética* [...] alguns dos critérios para “palavra” são apenas aplicáveis totalmente a línguas *deste tipo*. E quanto às línguas dos outros extremos do *continuum* tipológico – as de perfil *analítico* ou *polissintético*? (DIXON; AIKHENVALD, 2003, p. 3) [tradução e grifos nossos]

Portanto, começa a se revelar uma linguística apoiada em circularidades que, para além de enfraquecer o poder explicativo das teorias acerca da linguagem, também preserva laivos de etnocentrismo. É um quadro não tão distante do espírito carregado de nacionalismo e ideologias de pureza racial do século XIX quanto gostaríamos de imaginar.

### A palavra e o trabalho de missionários

Foi no século XIX que muitas línguas das possessões coloniais decorrentes de invasão da África e da Ásia foram descritas – e, caso ágrafas, registradas no alfabeto latino. As tentativas de grafar as línguas africanas, por exemplo, viram diferentes maneiras de escrita sendo aplicadas a uma mesma língua, de modo que aparentemente, “dois escritores não dividiam a mesma palavra de um mesmo jeito” (LOUWRENS, 1991). Ao se falar na grafia de línguas da África Subsaariana, duas estratégias ortográficas se destacavam: a disjuntiva, que tendia a grafar os morfemas como palavras separadas, e a conjuntiva, que os reunia em longas palavras com sequências morfêmicas lineares (LOUWRENS, 1991; MALEPE, 2006).

Bastante figurativo é o caso do sepedi (ou sotho do norte), língua bantu falada na África do Sul cujo registro gráfico se iniciou com missionários europeus há dois séculos. Juntamente com a escrita, veio o conceito de palavra – que, nesta língua, assumiu proporções curiosas: logo se estabeleceu extenso debate quanto ao sepedi ser uma “língua isolante” ou “aglutinante”, mormente por conta de indefinições de sua proposta gráfica. A título de ilustração, temos a sentença sepedi para “ele não me vê”, que pode ser grafada de maneira disjuntiva – *ga a mponne* – ou conjuntiva – *gaamponne* (MALEPE, 2006). Lombard, sobre o sepedi, adverte os leitores a não se enganarem pela ortografia prática desta língua, pois “‘palavras’ são indicadas de maneira relativamente arbitrária para se representar na escrita” e “muitas partes de palavras são escritas separadamente [...] como se tivessem *status* de palavra” (LOMBARD, 1985, p. 15). O disjuntivismo acabou sendo a estratégia mais adotada para esta língua (TALJARD; BOSCH, 2006); percebe-se, porém, que o construto *palavra* acomodou-se de forma discutível ao sepedi, tão distante das línguas para as quais foi originalmente formulado.

A adoção de diferentes estratégias de grafia, além de afetar estudos que envolvam marcação de classes lexicais e uso de *corpora*, oculta similaridades morfológicas entre línguas estreitamente aparentadas (TALJARD; BOSCH, 2006). Como o sepedi, o zulu, pertencente ao mesmo subgrupo de línguas bantu meridionais, também teve a ortografia concebida por missionários no século XIX, embora grafado de maneira conjuntiva. Vejam-se duas sentenças,

uma em sepedi e uma em zulu, compostas por morfemas correspondentes – mas grafadas respectivamente como quatro palavras (*ke a ba rata*) e uma única (*ngiyabathanda*):

(6) Sepedi (nígero-congolesa: bantu)  
 ke a ba rata  
 SUJ.1SG PRES OBJ.CL2 gostar  
 “Eu gosto dele/dela.”

(7) Zulu (nígero-congolesa: bantu)  
 ngi-ya-ba-thanda  
 SUJ.1SG-PRES-OBJ.CL2-gostar  
 “Eu gosto dele/dela.”

No entanto, as dúvidas sobre a validade da palavra também são expressivas entre representantes da própria família indoeuropeia. Schwegler (1990) aventa, num capítulo dedicado à evolução do francês, a possibilidade de se analisarem diversas construções desta língua de forma a considerar os pronomes de sujeito e objeto não como palavras, como preconiza a análise tradicional, mas sim como afixos. Veja-se a sentença tradicionalmente grafada *il la lui a donné* sob a luz desta perspectiva<sup>5</sup>:

(8) il-la-lui-a-donn-é  
 3SG.M.SUJ-3SG.F.OBJ-3SG.OI-haver.3SG-dar-PART  
 “Ele deu-lha”

Tal análise, segundo o autor, teria implicações diretas no cálculo da sinteticidade do francês, língua historicamente posicionada no espectro mais analítico inclusive pelo proponente original da classificação, von Schlegel –, mas que então apresentaria um sintagma verbal tão morfológicamente complexo quanto os de línguas como o basco, o georgiano e o quíchua (SCHWEGLER, 1990). Tendo os linguistas do século passado ora situado o francês como mais analítico, ora como mais sintético, este exemplo ilustra simultaneamente a inadequação do item **palavra** e o impacto direto de sua delimitação problemática na classificação em tipos morfológicos. As falhas do arcabouço teórico apresentado são de tal magnitude que, durante o século XX, não houve consenso quanto ao perfil tipológico do francês, língua europeia de fartíssima documentação disponível.

### A palavra e a China

Tendo em vista que uma vasta porção do pensamento científico acerca da linguagem elaborou-se na Europa, carregando vieses linguísticos – a palavra – e extralinguísticos – ideologias como o *Volksgeist*, por exemplo –, vale investigar as demais tradições gramaticais, filológicas ou quaisquer outros campos do conhecimento que tenham se desenvolvido de

<sup>5</sup> Glosa nossa.

modo independente, ocupando-se de línguas de um perfil outro que não o das indoeuropeias. Amplamente utilizado como exemplo do que seriam línguas isolantes, o chinês<sup>6</sup> possui uma comunidade expressiva de falantes e uma história escrita que remonta a milênios; é tão diferente das línguas indoeuropeias quanto possível, e conta com um extenso corpo de escritos historiográficos, técnicos e artísticos. Isto em consideração, analisemos como se deu o estudo da gramática na China, bem como possíveis proposições conceituais advindas deste legado.

É curioso que apesar da grande tradição chinesa na compilação de dicionários elaborados, e estudos de léxico e fonologia, com foco em rimas silábicas aliado à análise dos caracteres, a produção de gramáticas na China – sejam normativas ou descritivas – é muito mais recente. As primeiras gramáticas de línguas siníticas sob o formato que conhecemos foram escritas somente a partir do século XVII por ocidentais, encaixando-se firmemente, portanto, no arcabouço teórico da Europa (MAIR, 1997; PEYRAUBE, 1999; CHAPPEL, 2006). Os eruditos chineses tradicionalmente focaram suas atenções no chinês clássico, não se demorando em análises dos vernáculos contemporâneos. Ironicamente, seria com as gramáticas missionárias dos ocidentais que os diferentes vernáculos siníticos seriam alvo de estudos. Mair (1997) contrasta a situação chinesa com a indiana, na qual, antes do contato com os europeus, havia extensa tradição gramatical – e a gramática era vista como um dos cinco *vidya*, tipos básicos de conhecimento. O autor especula que o sânscrito, como uma língua “extremamente flexional”, teria suscitado diversos estudos, contrariamente ao chinês clássico, “talvez a língua menos flexional da Terra”<sup>7</sup>. Procedente ou não, fato é que a gramática e o estudo da língua no geral desfrutavam de menor prestígio na China pré-moderna, tendo a maioria dos esforços se concentrado em seu intrincado sistema de escrita.

### **Gramáticas ocidentais para línguas orientais**

Uma destas primeiras gramáticas concebidas e escritas por ocidentais é a *Arte de la lengua Chiõ Chiu*, obra do espanhol Melchior de Mançano, no século XVII. Nela, ao tratar de verbos – inequivocamente invariáveis em todas as pessoas do discurso nas línguas siníticas –, o autor traz uma tabela de conjugação verbal nos moldes das conjugações indoeuropeias; na tabela 2, observa-se a conjugação do verbo *lāy*, “vir”<sup>8</sup>, em chiõ chiu, língua min nan falada no sul da China.

---

6 Por chinês, referimo-nos às línguas siníticas, um conjunto de línguas aparentadas, embora nem sempre mutuamente inteligíveis, e que compartilham a escrita logográfica e um panorama político-cultural.

7 Pellin (2008) faz uma afirmação semelhante, atribuindo a escassez de estudos morfológicos na China pré-moderna à “falta de flexão” do chinês, bem como à “indiferença cultural” dos chineses para com as línguas dos demais povos.

8 Os itens *guà*, *lu*, *ý*, *guàn*, *lǔn* e *ín* são os pronomes pessoais de 1<sup>a</sup>, 2<sup>a</sup> e 3<sup>a</sup> pessoas do singular e do plural, nesta ordem.

Tabela 2. Conjugação do verbo *lāy* “vir” proposta em *Arte de la lengua Chiō Chiu*

Romanização	Tradução
guà lāy	eu venho
lu lāy	tu vens
ý lāy	ele/a vem
guàn lāy	nós vimos
lǔn lāy	vós vindes
ín lāy	eles vêm

É um flagrante exemplo da inadequação do aparato linguístico europeu, construído sob medida para línguas tidas como flexionais, para descrever línguas tão tipologicamente apartadas das línguas indoeuropeias quanto as siníticas. Similarmente, a *Arte de la lengua mandarina*, escrita no mesmo século por um missionário também espanhol, Francisco Varo, traz em suas páginas descrições de casos sintáticos e declinações para o chinês – fenômenos que não encontram correspondência na realidade (MAIR, 1997). Por outro lado, diversos fenômenos característicos das línguas siníticas, como verbos seriais e classificadores nominais, foram cobertos de forma muito breve nos estudos conduzidos por estes gramáticos europeus, a despeito da centralidade de tais estruturas para a gramática das línguas da China.

Somente em 1898 surgiria a primeira gramática de uma língua sinítica feita nativamente, o *Mǎ Shì Wéntōng* (马氏文通) “Tratado de Gramática de Mǎ”, escrito por Mǎ Jiànzhōng. O autor, no entanto, se limita ao mandarim literário; além disso, é assumidamente afiliado ao modelo greco-latino de descrição gramatical. Insere-se num contexto cultural de esforços de modernização da China da dinastia Qing, imbuídos de urgência para defender a cultura e ordem social tradicionais chinesas da ameaça dos poderes ocidentais (PELLIN, 2008). Ao seguir os moldes das gramáticas europeias, o autor acreditava contribuir para a prosperidade de sua nação – chegando a atribuir o sucesso da riqueza do ocidente a suas gramáticas<sup>9</sup> (MAIR, 1997).

### A linguística chinesa a partir do século XX

Ainda que hoje em dia a linguística não mais busque decalcar os conceitos de caso, declinação e conjugação para as línguas siníticas, outros construtos provenientes da tradição ocidental têm se provado mais resistentes; é o caso da *palavra*. Sua definição e aplicabilidade são dificultadas por fatores intrínsecos à língua, mas também gráficos: as línguas siníticas são grafadas utilizando *hànzì* (汉字), um sistema logográfico de base silábica que é significativamente diferente dos modos de escrita alfabética do ocidente. Por conseguinte, a unidade básica mais visível é o *zì* (字), termo que descreve o símbolo gráfico veiculador de uma ideia, mais próximo em correspondência com o morfema que com a palavra. Escrito em sequência e sem o uso de

<sup>9</sup> Apesar de superficialmente contraditório, incorporar aspectos da cultura ocidental percebidos como chaves para o desenvolvimento da indústria, ciência e economia foi a estratégia adotada pelo Japão desde a Restauração Meiji (séc. XIX), e em seguida, pela China, com vasta elaboração/importação de vocabulário técnico-científico (MAIR, 1997).

espaços<sup>10</sup>, foi sobre o *zì* que recaíram as discussões acadêmicas durante os muitos séculos da produção intelectual chinesa. Num primeiro momento, como na gramática de Mǎ Jiànzhōng, tentou-se equacionar *palavra* com a unidade de análise tradicional e óbvia do chinês, o *zì*; por isso, no *Mǎ Shì Wéntōng*, o autor dá os nomes de *dòngzì* (动字, “zì de ação”) para verbos, *míngzì* (名字, “zì de nome”) para substantivos, *liánzì* (连字, “zì de conexão”) para conjunções. Não tardou para se apontarem os problemas de igualar *zì* e *palavra*: sendo o mandarim uma língua na qual a maioria das palavras é dissilábica, elas se comporiam de dois *zì* cada (DUANMU, 2017).

Atualmente, a linguística chinesa tem se preocupado bastante com o conceito de *cí* (词), que seria a palavra como tentativamente entende a linguística ocidental<sup>11</sup>. É consenso entre sinólogos e acadêmicos nativos que o termo **palavra** nunca constou nem do vocabulário tradicional erudito, nem do vocabulário de falantes leigos de chinês antes do contato com o ocidente. Esta acepção do termo dataria do começo do século XX, mas a discussão sobre a palavra só se intensificaria a partir de 1950, com o projeto expresso da República Popular da China de introduzir um sistema oficial de romanização do mandarim – o *Hànyǔ Pīnyīn* –, para o qual a delimitação de palavras era assunto urgente (DUANMU, 1998; DUANMU, 2017; MYERS, 2020).

Uma série de desafios à aplicação da palavra ao chinês, porém, persiste; não poucos foram os linguistas chineses que se opuseram ao uso deste conceito, ou demonstraram que os critérios de definição da palavra eram conflitantes quando transplantados ao chinês (DUANMU, 1998). Em sua Gramática do chinês falado, Yuen Ren Chao afirma:

Nem toda língua tem um tipo de unidade que se comporte na maioria [...] dos respeitos como se comporta a unidade “palavra” quando falamos ou escrevemos *em* inglês sobre as subunidades *do* inglês [...] Como veremos nos casos em si, encontraremos vários tipos de unidades similares à palavra que podem ser chamadas de palavra, sobrepondo-se a ela em grande extensão, mas que não têm necessariamente o mesmo escopo. (CHAO, 1968, p. 159) [tradução nossa, grifos do autor]

Separar claramente o que seria uma palavra do que seria um sintagma nominal no chinês não é uma tarefa simples. Figuram diversos casos ilustrativos desta dificuldade na literatura, como as unidades dissilábicas chamadas de *líhécí* (离合词 – palavras separáveis). Um exemplo é *yǒuqián* (有钱) “rico”; o termo é modificável por *hěny* (很) “muito”, um advérbio, em *hěnyǒuqián* (很有钱) “muito rico”, no que se comporta como um composto. No entanto, pode acabar sendo separado, como na expressão *yǒu hěny duō qián* (有很多钱) “muito rico/ter muito dinheiro” (DUANMU, 2017). Outro exemplo são as construções interrogativas no formato A+não+AB, nas quais AB é uma unidade dissilábica; veja-se o trazido em Myers (2020):

10 Perceba-se a ausência de espaço gráfico em todos os exemplos em mandarim do presente artigo, que vêm acompanhados de sua grafia nativa em caracteres chineses; a sentença glosada em (1) na introdução escreve-se 龙在云里飞翔, por exemplo.

11 Para a nomenclatura gramatical chinesa da atualidade, verbos são *dòngcí* (动词, “palavras de ação”); substantivos, *míngcí* (名词, “palavras de nome”); conjunções, *liáncí* (连词, “palavras de conexão”), e assim por diante (YANG, 2007).

- (6) 你高不高兴  
 nǐ gāo-bù-gāoxìng  
 2SG feliz-NEG-feliz  
 “Você está feliz?”

Similarmente, as palavras ditas *elásticas* (DUANMU, 1998) ou *teléscópicas* (HUANG *et al.*, 2017) do mandarim são outro fenômeno curioso; são unidades dissilábicas, a exemplo de *xuéxí* (学习) “estudar”, *jìshù* (技术) “habilidade” e *lǎohǔ* (老虎) “tigre”, que podem se realizar em alguns contextos como *xué*, *jì* e *hǔ*, respectivamente, apresentando o mesmo significado das formas plenas.

As tentativas de conciliar a noção de palavra com a estrutura do chinês frequentemente envolvem a distinção entre compostos dissilábicos e sintagmas nominais qualificados (PAUL, 2005; SCHÄFER, 2009; XU, 2018), ou até a proposição de uma natureza multidimensional para a palavra nestas línguas, envolvendo sua expressão nos planos morfológico, sintático e prosódico (SHENGLI, 2001), ou postulando seu surgimento “por meio da interação de forças linguísticas distintas”. Myers, defendendo a existência da palavra no chinês, afirma:

Ter interface entre domínios tão fundamentalmente diferentes, porém, requer um acordo dinâmico. Palavras são algo como tornados, torcidas em pleno ar por forças concorrentes poderosas, e continuando a se torcer para longe de nossa compreensão mesmo após formadas, mas ainda assim inegavelmente reais. (MYERS, 2020, p. 17) [tradução nossa]

Parece-nos que **palavra** segue um conceito não-universal, tentativamente atingido a duras penas; a extensa discussão necessária para compreendê-lo, juntamente à falta de consenso quanto a sua delimitação ou até mesmo validade põem em dúvida seu caráter de unidade básica. Os esforços parecem partir de uma vontade preestabelecida de subsumir a realidade ao construto teórico, e não de postular um construto que reflita o que ocorre na linguagem.

### O século XXI e a análise por formativos de Bickel e Nichols

Em 2007, em um capítulo do livro *Language typology and syntactic description*, Balthasar Bickel e Johanna Nichols oferecem uma importante contribuição à tipologia morfológica, propondo o conceito de **formativo** como item de análise. Para os autores, formativos são marcadores de informação flexional. Diferem de palavras por não poderem reger ou ser regidos por outras palavras, não poderem requerer ou sofrer concordância, e não poderem ser núcleos de sintagma: “*formativos são entidades morfológicas, palavras, sintáticas*” (BICKEL; NICHOLS, 2007, p. 172). [tradução e grifos nossos]

Nas línguas da Europa Ocidental, aduzem os autores, formativos tipicamente se realizam por meio de morfologia presa; não obstante, ao se observarem as línguas do mundo, vê-se que nem todos os formativos são formas presas como são os morfemas flexionais das línguas

indoeuropeias. Para ilustrar, trazem o lai chin, uma língua tibeto-birmanesa falada em Myanmar. Nesta e em outras línguas do sudeste asiático, formativos ocorreriam *sob a forma de palavras livres*; observe-se o morfema veiculador de caso ergativo no exemplo:

- (7) TSEW MÁD NÍ? ?a-ka-ṭHOʔD  
Tsew Mang ERG 3SG-A-1SG.P-bater  
“Tsew Mang me bateu”

O formativo seria um item de análise que se apresentaria das mais diversas formas, e é sob ele que recai todo o aparato classificatório legado dos séculos de discussão quanto a analiticidade/sinteticidade e demais categorias morfológicas. Para Bickel e Nichols, não são as línguas por inteiro que receberiam a classificação; é salientado que as escalas dos parâmetros são aplicadas a formativos individuais ou a grupos de formativos. Os autores justificam apontando que formativos isolantes, por exemplo, encontram-se na quase totalidade das línguas, ao menos como algumas poucas partículas fonologicamente livres. Desencorajam que se usem termos tais quais “línguas isolantes”, muito embora as línguas difiram no grau em que empregam os diferentes tipos de formativo – isto é, é possível afirmar se uma determinada língua apresenta mais ou menos formativos isolantes em relação a uma outra.

A classificação dos autores, porém, não é um *continuum* entre formas analíticas ou sintéticas, mas sim uma série de parâmetros ortogonais uns aos outros, oferecendo possibilidades combinatórias para se agruparem os formativos das diferentes línguas. Central para o estabelecimento de seus parâmetros, encontra-se a distinção entre palavra **fonológica** e palavra **gramatical**. Para discutirmos como cada destes dois conceitos é trazido pelos critérios de classificação, passemos a definir os parâmetros propostos, co-ocorrentes em cada formativo: o **grau de fusão**, a **flexividade** e a **densidade semântica**.

### Grau de fusão

Refere-se à fusão fonológica dos formativos às raízes, havendo um gradiente com respeito a quão firmemente estes estão unidos. Os autores apresentam uma escala *ISOLANTE* → *CONCATENATIVO* → *NÃO-CONCATENATIVO*.

Formativos isolantes são palavras fonológicas plenas, como no exemplo do lai chin em (7), ocorrendo livres em sequência às raízes. Já os concatenativos são fonologicamente presos e necessitam se ligar a alguma palavra para ocorrerem; incluem desinências flexionais e clíticos, e os autores salientam que a marca mais perceptível da concatenação é que os formativos são prontamente segmentáveis, como em diversas línguas bantu, quíchua ou túrquicas. Em turco, por exemplo, ad-lar “nome-pl”, ad-in “nome-gen” e ad-lar-in “nome-pl-gen” permitem visualizar claramente os formativos de caso genitivo e número plural como formando uma sequência linear com a raiz *homem*. A similaridade de tais sequências concatenativas com as sequências de formativos isolantes, apontada desde a lista de Tauli, não passa despercebida. Os autores justificam

a manutenção de duas categorias distintas sob a alegação de que formativos concatenativos disparam ajustes fonológicos e morfofonológicos “na *palavra* que eles constroem em conjunto com o hospedeiro” [grifos nossos], apoiando-se no conceito de palavra para diferenciar seus conceitos de isolamento e concatenação. São elencados como processos que afetariam unicamente os formativos concatenativos – e por isso, resultariam em palavras – assimilação, dissimilação, elisão e restrições prosódicas. Por fim, formativos não-concatenativos se realizariam por modificação direta da base; englobam mudanças prosódicas ou de tom na raiz, subtração de morfemas, ou o processo amplamente conhecido das línguas semíticas e suas raízes triliterais acompanhadas de diferentes vocalismos para expressar significados lexicais ou flexionais diferentes (como no árabe *k-t-b* “escrever”, de onde *katab* “ele escreve/escreveu” e *aktub* “ele está/estava escrevendo”).

### Flexividade

Os autores, para evitar confusão com “flexão”, propõem o termo “flexividade” para um de seus parâmetros de variação tipológica dos formativos. Diz respeito à alomorfa lexical, motivada não por contextos morfofonológicos, mas pelo alinhamento da raiz a determinadas classes flexionais presentes na língua. Boa parte dos casos de flexividade vêm de línguas indoeuropeias – principalmente as mais conservadoras. Nelas, os casos sintáticos, por exemplo, seriam expressos por formativos contidos em conjuntos de alomorfes selecionáveis com base no pertencimento da base lexical a determinado paradigma de declinação. Ilustra o genitivo latino, veiculado por um formativo variável a depender da vogal temática: esta, em latim, pode reunir os nomes numa classe de tema em -a como *nauta*, marinheiro; -e, como *dies*, dia; -o, como *lupus*, lobo; -u, como *cornus*, chifre; e um quinto paradigma reunindo atemáticos e palavras com tema em i- como *nubes*, nuvem. O morfema de genitivo, portanto, varia entre {-i} para as três primeiras classes e {-s} para as demais.

### Densidade semântica

Densidade semântica é o parâmetro proposto pelos autores para abordar o acúmulo ou não de diferentes significados gramaticais numa mesma unidade de análise – o próprio formativo, no caso da **exponência**, e a palavra – aqui, não mais a fonológica, mas sim definida como **palavra gramatical** –, no caso da **síntese**.

### Exponência

Refere-se ao grau no que diferentes conteúdos gramaticais são veiculados pelo mesmo formativo, que a despeito de portar mais de um significado gramatical, permanece uno e indivisível. Os autores distinguem dois protótipos, formativos **separativos** – como muitos dos formativos lineares das línguas tradicionalmente encaradas como *aglutinantes* a exemplo do turco em (2) ou do quéchua em (5); e formativos **cumulativos**, comuns em línguas indoeuropeias, nas quais um mesmo morfema carrega mais de um significado – como no exemplo do francês em (3), em que o artigo definido *le* veicula as noções de masculino e singular.

## Síntese

Aqui, Bickel e Nichols se apoiam fortemente no conceito de palavra e na lógica circular que lhe é subjacente. **Palavras analíticas** seriam compostas de um número muito limitado de formativos, frequentemente apenas um, ou de uma raiz lexical pura; ao se combinarem sintaticamente para exprimir uma categoria flexional, formariam perífrases. Os autores lançam mão de exemplos do inglês para demonstrar como o futuro nesta língua envolve duas palavras gramaticais distintas: ocupam posições variáveis (*Your friend will go* em contraste a *Will your friend go?*) e podem ter interposição de outros termos (*He will definitely go*). Já as **palavras sintéticas** teriam constituição interna mais elaborada, apresentando-se como uma sequência de formativos presos juntos à raiz. Os autores, no entanto, afirmam a possibilidade de sequências de formativos isolantes se combinarem numa só “palavra sintática complexa” e de ordem interna rígida<sup>12</sup>. Por fim, também recuperam a noção de polissíntese; uma palavra seria polissintética, como em muitas línguas indígenas da América do Norte, caso composta por uma série não só de formativos, mas também de raízes incorporadas e afixos lexicais.

## O formativo *sensu* Bickel e Nichols e a palavra

Como exposto na subseção anterior, os autores distinguem o conceito fonológico do conceito gramatical da palavra, trazendo este construto teórico ao cerne da proposta; para se definir a natureza de um formativo quanto ao grau de fusão, por exemplo, o conceito de palavra fonológica é o eixo orientador. Em relação aos demais parâmetros, principalmente à síntese, é a palavra gramatical que norteia como o formativo será enxergado. Na seção **Síntese**, os autores afirmam que a noção de palavra por eles usada para situar formativos em sua escala de densidade semântica quanto à palavra é a “gramatical, não a fonológica. A palavra gramatical é definida como *menor unidade da sintaxe*, tecnicamente o nodo terminal ou projeção mínima (X<sub>0</sub>) na estrutura do sintagma” (BICKEL; NICHOLS, 2007, p. 190). O uso de dois conceitos diferentes de palavra torna a análise complexa, com sequências de palavras independentes segundo um critério comportando-se como uma só palavra segundo o outro. Se por um lado precisar de quais critérios lançamos mão ao falarmos em “palavra” oferece maior detalhamento ao modelo explicativo, este quadro não deixa de representar mais uma tentativa de acomodação – a duras penas – da realidade ao construto “palavra”.

Na sua definição de polissíntese, os autores lastreiam sua proposta na circunscrição dos processos envolvidos à morfologia ou à sintaxe para classificar as sequências de formativos polissintéticas como sendo palavras – “a sequência envolve dependências puramente *morfológicas, não-sintáticas*” (BICKEL; NICHOLS, 2007, p. 194) [grifos nossos]. Fica patente

---

12 O exemplo usado pelos autores é da língua lai chin, da qual trouxemos uma sentença em (7). Para eles, o *status* de “palavra complexa” se justificaria com base nos mesmos testes aplicados aos exemplos do inglês; diferentemente da sequência inglesa *will go*, uma sequência lai chin exprimindo futuro não pode ter seus formativos trocados de posição nem interpostos por outro elemento.

que os autores trazem em seu cabedal teórico a dicotomia morfologia/sintaxe, lançando mão dela para delimitar o que é palavra, e, por conseguinte, estabelecer a maior parte da organização de seu sistema classificatório quanto à síntese. Esta operação, longe de ser algo incomum, norteia boa parte da linguística teórica; usemos de exemplo a colocação de San Duanmu quanto à delimitação da morfologia e a natureza de determinados processos fonológicos:

A distinção entre palavras e sintagmas, porém, é de vital importância para a morfologia e a fonologia. *Sem saber o que é uma palavra, não se pode significativamente falar em morfologia.* Similarmente, algumas regras fonológicas, como atribuição de tonicidade e determinação de domínios tonais, aplicam-se diferentemente no nível de palavra e no nível sintagmático [...] Sem uma distinção entre palavras e sintagmas, tais regras pareceriam ser *ad hoc*. (DUANMU, 1998, p. 135) [tradução e grifos nossos]

Dito de forma clara, a distinção entre morfologia e sintaxe é definida com base na palavra – e em diversos trabalhos, inclusive o de Bickel e Nichols, a palavra vem definida com base na distinção entre morfologia e sintaxe. A despeito da importante contribuição para o debate, e da grande utilidade da proposta de análise por formativos, esta é mais uma instância de definição circular subjacente a conceitos-chave para a teoria linguística.

### Outras perspectivas

Autores como John McWhorter e Walter Bisang, discutindo crioulos e demais línguas tidas por escassas em morfologia, oferecem uma fonte de novas percepções ao debate, a despeito de também utilizarem as noções de língua analítica e língua sintética. Os autores propõem que determinadas línguas, como muitas do Sudeste Asiático ou da África Ocidental, apresentariam uma gramática marcadamente “telegráfica”<sup>13</sup>, sendo por isso **radicalmente analíticas**. Diferentemente de línguas que seriam mais ou menos analíticas a depender dos parâmetros adotados – que eles tratam por analíticas apenas relativamente, como o inglês –, as línguas radicalmente analíticas se caracterizariam pela ausência quase completa de morfologia flexional **contextual**. Apresentariam, quando muito, morfologia flexional **inerente**, não importando para esta distinção se os morfemas envolvidos na flexão são presos ou livres (MCWHORTER, 2016; MCWHORTER, 2019; BISANG, 2020).

Proposta por Booij (1993), a distinção entre morfologia flexional inerente e contextual diz respeito às diferentes informações gramaticais veiculadas pelos morfemas. A morfologia flexional inerente não seria requerida pelo contexto sintático, a exemplo de número em substantivos, grau em adjetivos, e tempo e aspecto em verbos. As línguas apontadas por Bisang e McWhorter como radicalmente analíticas – casos emblemáticos usados pelos autores são o iorubá, o chinês e línguas austronésias da ilha de Flores, além de diversos crioulos –, apresentariam somente

13 A expressão é usada em diversas publicações de McWhorter para se referir a línguas com escassa morfologia flexional.

este tipo de morfologia flexional. Em contraposição, a morfologia flexional contextual, como caso e concordância, seria ditada pelo contexto sintático. Esta é mais uma ocorrência em que a distinção morfologia/sintaxe – e, em seu bojo, a palavra – é estruturante para a linguística. Porém, embora não ocupem centralidade em sua proposição, Booij oferece conceituações da dicotomia inerente/contextual que não aludem à palavra: flexão inerente teria **significado**, importância semântica, sendo por isso adquirida mais precocemente por crianças; a contextual desempenharia **funções**, daí sua aquisição mais tardia (BOOIJ, 1993). É dizer, de certa forma, que a morfologia flexional inerente teria um caráter dêitico, ancorado a elementos da realidade extralinguística, enquanto a morfologia flexional contextual veicularia relações e propriedades do sistema interno à língua.

McWhorter utiliza esta distinção, além de hipóteses históricas e geográficas, para propor que a analiticidade radical seria um estado tão marcadamente singular entre as línguas humanas que todas as instâncias de sua ocorrência se explicariam por aquisição imperfeita por parte de adultos (MCWHORTER, 2016; MCWHORTER, 2019). Independentemente da validade ou não de suas alegações para a linguística histórica, para a crioulística e o estudo do contato entre línguas, o autor suscita questionamentos estimulantes munido dos conceitos de analiticidade radical e morfologia flexional contextual/inerente – esta distinção, caso apoiada na semântica, aponta para possibilidades de análise apartadas do uso da palavra.

McWhorther também faz alusão à oposição proposta por William Thurston (1987) entre gramáticas **esotéricas**, de línguas de uso restrito a sua comunidade imediata de falantes e raramente aprendidas por adultos, e gramáticas **exotéricas**, de línguas de uso generalizado como língua franca/segunda língua<sup>14</sup>. Em sua obra, Thurston afirma que “é de grande ajuda ao aprendiz [adulto] se cada morfema puder ser facilmente analisado e identificado” (THURSTON *et al.*, 1987, p.58), de onde a morfologia mais simplificada das gramáticas exotéricas em contraste com a das esotéricas; estas últimas, aprendidas pelas crianças em transmissão regular, priorizariam compacidade, favorecendo elisão de vogais, redução fonêmica dentro dos morfemas, e fusão destes entre si e às raízes. É ao perfil exotérico que McWhorter subsume as línguas radicalmente analíticas, alvos de aquisição adulta por conta de movimentos populacionais na história dos povos envolvidos. A discussão, embora envolvendo estudos exteriores à linguística, tem seu mérito por provocar novos enfoques, mormente por utilizar dados da comunidade de falantes e uma possível análise semântica da morfologia flexional das línguas.

Ademais, há propostas que desconsiderem totalmente a palavra como item de trabalho, focando unicamente nos morfemas – uma vez que mais facilmente delimitáveis e analisáveis –, seguindo o caminho anunciado por Zellig Harris (1946). Em Lazzarini-Cyrino e Mattos (2020),

---

14 Wray e Grace (2007) elaboram este conceito, referindo-se não a *língua* ou *gramática*, mas a *comunicação* esotérica – entre membros de uma mesma comunidade, que compartilham conhecimento; e *comunicação* exotérica – entre membros de diferentes comunidades, ou em sociedades amplas, estratificadas e especializadas.

utilizamos sentenças provenientes de gramáticas descritivas, segmentadas em morfemas e devidamente acompanhadas de suas glosas, como fonte de dados para algoritmos de aprendizado de máquina não supervisionado. Investigamos, num primeiro momento, quatro línguas sem relação geográfica ou genética, araweté (tupi-guarani: Brasil), yakkha (sino-tibetana: Nepal), saami de Pite (fino-úgrica: Suécia e Noruega) e khwarshi (nakh-daguestaniana: Rússia), reunindo os morfemas de cada língua em grupos com o auxílio dos algoritmos. Os agrupamentos gerados foram analisados, bem como associações entre bigramas<sup>15</sup> dos morfemas e suas possibilidades combinatórias. Obtivemos métricas indicadoras de diferenças quantitativas entre as línguas quanto à maneira em que combinam morfemas de grupos distintos – seja de forma mais livre, seja de forma mais restritiva, o que poderia representar indícios de delimitação ou não de unidades supramorfêmicas. Tal uso de técnicas de processamento de linguagem natural nos estudos em tipologia, a despeito de inicial, aponta para uma direção potencialmente proveitosa, oferecendo métodos de comparação operacionais e aplicáveis às mais diversas línguas, além de baseados na observação dos dados – e não puramente na tradição.

## Conclusão

Acompanhamos a história da discussão analiticidade/sinteticidade e dos empreendimentos classificatórios na linguística, partindo de sua concepção comprometida por ideais acientíficos até as contribuições mais recentes ao debate, ainda marcado de forma profunda pelo eurocentrismo. Vimos também como o conceito **palavra** é um construto europeu forçadamente imposto a toda a disciplina, por conta do qual são feitos esforços tremendos a fim de adequar a realidade das diferentes línguas do globo a uma entidade teórica de validade questionável. Frente a isto, cabe a reflexão quanto a alternativas viáveis para servir de baliza aos estudos subsequentes que se debrucem sobre a linguagem, e um fazer científico que não utilize conceitos e itens de análise fundados em tradição etnocêntrica.

## REFERÊNCIAS

- ALIGHIERI, D. *De vulgari eloquentia*. Cambridge: Cambridge University Press, 1996.
- BICKEL, B.; NICHOLS, J.: Inflectional morphology. In.: SHOPEN, Timothy (ed.). *Language typology and syntactic description*. V. 3. Cambridge: Cambridge University Press, n. 2, 2007, p. 169-240.
- BISANG, W. Radical analyticity and radical pro-drop scenarios of diachronic change in east and mainland Southeast Asia, West Africa and pidgins and creoles. *Asian Languages and Linguistics*, v. 1, n. 1, p. 34–70, 2020.

---

15 Bigramas, no caso, são sequências de dois morfemas sucessivos, unidades binárias cujos possíveis componentes apresentam graus variados de afinidade entre si.

BOOIJ, G. Against split morphology. In.: BOOIJ, G.; VAN MARLE, J. (eds.). *Yearbook of morphology*, Berlim: Springer, 1993, p. 27-49.

CHAO, Yuen Ren. *A grammar of spoken Chinese*. Berkeley: University of California Press, 2017.

CHAPPELL, H. From Eurocentrism to Sinocentrism: the case of disposal constructions in Sinitic languages. In.: AMEKA, F. K.; DENCH, A. C.; EVANS, N. (eds.). *Catching language: the standing challenge of grammar writing*, Berlim: Walter de Gruyter, 2006, p. 441-486.

COMRIE, B. Holistic versus partial typologies. In: BAHNER, W.; SCHILDT, J.; VIEWEGER, D. (eds.). *Proceedings of the Fourteenth International Congress of Linguists*. Berlin, 139-148, 1991.

DIXON, R. M. W.; AIKHENVALD, A. Y. (eds.). *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press, 2003.

DRYER, M. S.; HASPELMATH, M. (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Disponível em: <<https://wals.info/>>. Acesso em: 17 nov. 2020.

DUANMU, S. Wordhood in Chinese. In.: PACKARD, J. L. (ed.). *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese*, Berlim: Walter de Gruyter, 1998, p. 135-196.

DUANMU, S. Word and wordhood, modern. In.: SYBESMA, R. P. E. *et al.* (ed.). *Encyclopedia of Chinese language and linguistics*. V. 4. Leiden: Brill, 2017, 2017, p. 543-549.

GREENBERG, J. H. A quantitative approach to the morphological typology of language. *International journal of American linguistics*, v. 26, n. 3, p. 178-194, 1960.

HARRIS, Z. S. From morpheme to utterance. In: *Papers on Syntax*. Springer, Dordrecht, 1946. p. 45-70.

HINTZ, D. J. *Aspect and aspectual interfaces in South Conchucos Quechua: The emergence of grammatical systems*. Tese (Doutorado em Linguística) da University of California, Santa Barbara, 2007.

HUANG, C. R.; HSIEH, S. K.; CHEN, K. J. *Mandarin Chinese words and parts of speech: A corpus-based study*. Oxford: Taylor & Francis, 2017.

LAZZARINI-CYRINO, J. P.; MATTOS, E. B. Um estudo exploratório sobre a classificação de morfemas por agrupamento hierárquico para comparação tipológica. *Revista do GELNE*, Natal, v. 22, n. 2: Dossiê Morfologia, p. 395-407, 2020.

LOMBARD, Daniel P. *Introduction to the grammar of Northern Sotho*. Pretória: JL van Schaik, 1985.

- LOUWRENS, L. J. *Aspects of Northern Sotho Grammar*. Cidade do Cabo: Via Afrika, 1991.
- MAIR, V. H. Ma Jianzhong and the invention of Chinese grammar. *Journal of Chinese Linguistics Monograph Series*, n. 10, p. 5-26, 1997.
- MALEPE, M. K. (2006). *Elements of Northern Sotho orthography: a linguistic approach*. Dissertação (Mestrado em Línguas Africanas) da Escola de Linguagens e Estudo em Comunicação, Faculdade de Humanidades, Universidade do Limpopo, Mankweng, 2005.
- MAZZOCCO, A. *Linguistic theories in Dante and the Humanists: Studies of language and intellectual history in Late Medieval and Early Renaissance Italy*. Leiden: Brill, 1993.
- MCWHORTER, J. Is radical analyticity normal? In.: VAN GELDEREN, E. (ed.). *Cyclical change continued*, Amsterdam: John Benjamins Publishing Company, 2016, p. 227-249.
- MCWHORTER, J. The radically isolating languages of Flores: A challenge to diachronic theory. *Journal of Historical Linguistics*, v. 9, n. 2, p. 177-207, 2019.
- MYERS, J. Wordhood and disyllabicity in Chinese, 2020. Disponível em: < <http://personal.ccu.edu.tw/~lngmyers/>>. Acesso em: 17 de novembro de 2020
- PARRY, D. Umberto Eco and the echoes of Adamic language. *Zagadnienia Rodzajów Literackich*, v. 58, n.2, p. 13-28, 2015.
- PAUL, W. Adjectival modification in Mandarin Chinese and related issues. *Linguistics*, v. 43, n. 4, p. 757-793, 2005.
- PELLIN, T. Inventing a modern lexicon for grammar in Chinese: the experience of Wang Fengzao, Ma Jianzhong and Yan Fu. *Language Sciences*, v. 30, n. 5, p. 529-545, 2008.
- PEYRAUBE, A. Sur les sources du Ma Shi Wen Tong. *Histoire épistémologie langage*, v. 21, n. 2, p. 65-78, 1999.
- SAPIR, E. *Language: An introduction to the study of speech*. Nova York: Harcourt, Brace and Company, 1921.
- SCHÄFER, M. A+N constructions in Mandarin and the ‘compound vs. phrase’ debate. *Word Structure*, v. 2, n. 2, p. 272-293, 2009.
- SCHWEGLER, A. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlim: Walter de Gruyter, 1990.
- SHENGLI, F. Lun hanyu ci de duwei xing [on the multi-principles of Chinese words]. *Dangdai yuyanxue*, v. 3, p. 161-174, 2001.
- STOCKING, G. W. *Volksgeist as method and ethic: Essays on Boasian ethnography and the German anthropological tradition*. Madison: Univ of Wisconsin Press, 1996.

- TALJARD, E.; BOSCH, S. E. A comparison of approaches to word class tagging: Disjunctively vs. conjunctively written Bantu languages. *Nordic journal of African studies*, v. 15, n. 4, 2006.
- TAULI, V. Morphological analysis and synthesis. *Acta Linguistica*, v. 5, n. 1, p. 80-85, 1945.
- THURSTON, W. R. *Processes of change in the languages of north-western New Britain*. Canberra: The Australian National University, 1987.
- VARO, F. *Francisco Varo's Grammar of the Mandarin Language, 1703: An English Translation of 'Arte de la Lengua Mandarina'*. COBLIN, W. S; LEVI, J. A. (eds.). Amsterdam: John Benjamins Publishing, 2000.
- VICK, B. The origins of the German Volk: Cultural purity and national identity in nineteenth-century Germany. *German Studies Review*, v. 26, n. 2, p. 241-256, 2003.
- VON SCHLEGEL, A. W. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande, 1818.
- WRAY, A., GRACE, G. W. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, v. 117, n. 3, 543-578, 2007.
- XU, Z. The word status of Chinese adjective-noun combinations. *Linguistics*, v. 56, n. 1, p. 207-256, 2018.
- YANG, J. *Hanyu Jiaocheng: Language Skills Chinese courses*. Beijing: Beijing Language & Culture University Press, 2007.