Correlações entre a expressão de categorias gramaticais-funcionais e parâmetros intrínsecos a morfemas em sentenças codificando ações de diversas línguas

Eudes Barletta Mattos[1]; João Paulo Lazzarini Cyrino[2]

eudesbarlettam@gmail.com; jpcyrino@gmail.com

RESUMO: A classificação morfológica em tipos holísticos, surgida no século XIX, tradicionalmente subsumiu línguas inteiras a um de geralmente três perfis, isolante, aglutinante e fusional. As línguas fusionais, como muitas línguas indoeuropeias, costumam reunir como características uma série de parâmetros que, na verdade, são ortogonais uns em relação aos outros, podendo *a priori* ocorrer quaisquer combinações entre eles. São parâmetros de destaque dentre os característicos das línguas tidas por fusionais a preponderância de alomorfia e poliexponência entre os morfemas funcionais, bem como seu reduzido tamanho ou fusão entre eles, além de significativa expressão de categorias morfológicas obrigatórias, chegando a apresentar morfologia sintaticamente redundante sob a forma de regência ou concordância. Crescentemente, porém, a linguística compreende que numa mesma língua, vários perfis morfológicos coexistem, podendo variar de acordo com o tipo de construção. O presente estudo, portanto, utiliza sentenças codificando ações em 50 línguas divididas em 10 grupos em sua maioria genéticos, triando e quantificando-lhes os morfemas funcionais, bem como calculando dois grupos de parâmetros, i. os desvios da relação de biunivocidade forma-significado e tamanho do morfema, e ii. os conceitos funcionais, categorias morfológicas adquiridas e redundância sintagmática. Estabeleceram-se correlações de Pearson entre os parâmetros dos dois grupos, demonstrando, sem se apoiar em noções mal-definidas de simplicidade ou complexidade, nem na tipologia holística, que os morfemas funcionais de sentenças codificando ações têm a tendência de serem menores e sofrerem mais acomodações morfofonológicas, como alomorfia e poliexponência, de acordo com níveis mais expressivos de morfologia funcional expressa na sentença.
PALAVRAS-CHAVE: complexidade morfológica; tipologia linguística; exponência; sentenças de ações.

ABSTRACT: The morphological classification into holistic types, which emerged in the 19th century, traditionally subsumed entire languages into one of generally three profiles: isolating, agglutinative, and fusional. Fusional languages, like many Indo-European languages, tend to share a series of characteristics that actually are orthogonal to each other, and any combination between them can occur *a priori*. Key parameters characteristic of languages considered fusional include high levels of allomorphy and polyexponence among functional morphemes, as well as their reduced size or combined state, as well as the significant expression of obligatory morphological categories, even exhibiting syntactically redundant morphology in the form of government or agreement. Increasingly, however, linguistic studies understand that several morphological profiles coexist within the same language, varying according to the type of construction. This study, therefore, uses action sentences in 50 languages divided into 10 mostly genetic groups, sorting and quantifying their functional morphemes, as well as calculating two groups of parameters: i. deviations from the form-meaning one-to-one relationship and morpheme size, and ii. functional concepts, acquired morphological categories, and syntagmatic redundancy. Pearson correlations were established between the parameters of the two groups, demonstrating, without relying on ill-defined notions of simplicity or complexity, nor on holistic typology, that the functional morphemes of sentences encoding actions

[1] Doutorando; Universidade Federal da Bahia.
[2] Professor Adjunto; Universidade Federal da Bahia.

tend to be smaller and undergo more morphophonological accommodations, such as allomorphy and polyexponence, according to more expressive levels of functional morphology expressed in the sentence.
KEYWORDS: morphological complexity; linguistic typology; exponence; action sentences.

Introdução

O longo debate acerca da complexidade morfológica estabeleceu a tradicional associação entre uma expressão mais rica de categorias funcionais e a presença de complexificações nas próprias unidades portadoras de tais categorias. Desde seu início, a classificação morfológica holística – isto é, subsumir línguas em sua inteireza a um dado perfil tipológico com base em sua morfologia – atribuiu línguas com expressiva presença desses dois grupos de parâmetros, demonstrando ricos paradigmas flexionais e extensa alomorfia, ao tipo **fusional**[3]**,** embora sem um método padronizado e quantificável. Cada vez mais, entende-se que classificar morfologicamente as línguas de maneira holística é contraproducente; após um extenso percurso, desde as contribuições de Edward Sapir (1921) e Valter Tauli (1950), passando por Östen Dahl (2004) e Bickel e Nichols (2007), compreende-se hoje que no seio de uma língua convivem diferentes estratégias para realização de construções, cada uma apresentando diferentes níveis de complexidade.

O próprio conceito de complexidade, igualmente, tem sido tensionado, sendo muitas vezes decomponível e analisável numa série de parâmetros, como propõem a taxonomia de John McWhorter (2007) e a análise por complexificações de Peter Trudgill (2011). As línguas ditas **fusionais**, portanto, representariam a confluência de um feixe de características que historicamente foram encaradas como mais complexas, como a expressão explícita de categorias gramaticais e morfologia sintaticamente redundante, e grande presença de morfemas portando mais de uma categoria gramatical cada. Ao menos numa primeira análise, esses dois grupos de parâmetros (i.e., expressão explícita de categorias gramaticais e morfologia

[3] Proposta inicialmente por August von Schlegel (1808), em sua obra **Observations sur la langue et la littérature provençales**, a classificação era tripartite, consistindo em i. línguas isolantes, como as línguas do leste e sudeste da Ásia, que, para ele, seriam desprovidas de estrutura gramatical; ii. línguas aglutinantes, como as línguas bantu e túrquicas, que codificariam as relações gramaticais por meio de justaposição de elementos prontamente separáveis uns dos outros; e iii. línguas fusivas ou fusionais, como as línguas da Europa Ocidental, que realizariam as relações gramaticais por meio de modificações nas raízes, além de afixação, resultando em elementos flexionados que mais facilmente se organizam em paradigmas do que se decompõem em subunidades claras.

sintaticamente redundante) que juntos pareciam caracterizar o tipo fusional revelam-se na verdade ortogonais um em relação ao outro (Mattos e Lazzarini-Cyrino: 2021; Mattos: 2023; Dahl: 2004; Bickel e Nichols: 2007). O surgimento desta combinação de parâmetros poderia indicar o resultado de processos de acomodação de morfemas funcionais se fundindo entre si ou interagindo uns com os outros de maneira mais integrada – como Stéphane Robert (2018) afirma:

> [L]ínguas flexionais (por exemplo, latim, russo) representam o grau extremo de integração entre os níveis sintático e morfológico. Nessas línguas [...] em geral os morfemas gramaticais que se combinam com os radicais do léxico **constituem amálgamas inquebráveis: um mesmo morfema exerce diversas funções** [...] Os diferentes componentes, originalmente justapostos, sofreram, portanto, durante a história da língua**, uma integração tal que se fundiram formalmente e não podem mais ser divididas em subunidades.** (pp. 19-20, tradução e grifos nossos)

Ora, o que Robert chama de **amalgamação** é justamente o acúmulo de uma série de processos que culminam em maiores medidas de complexidade intrínseca ao morfema em si[4]. Haja vista que tais parâmetros seriam, **a priori**, distintos um do outro, é pertinente indagar qual a relação entre uma maior presença de material morfológico funcional na sentença de uma dada língua e as modificações estruturais que os morfemas nesta sentença teriam sofrido. Investigam-se no presente trabalho tais fenômenos, organizados em parâmetros quantificáveis de complexidade e não se apoiando na terminologia holística, num estudo exploratório que busca lançar as bases para direcionamentos futuros e estabelecer correlações entre os parâmetros.

Para tanto, foram utilizadas 50 línguas, divididas em 10 grupos majoritariamente genéticos. Os grupos foram bastante diversificados genética e arealmente, havendo línguas indígenas da América do Sul, tupi-guarani ou não, amazônicas ou não; línguas da África Ocidental, região de alta diversidade de famílias, e da África Meridional, vastidão territorial dominada pela família bantu; línguas do arquipélago de Nusantara, entre Ásia e Austrália; línguas do Sudeste Asiático continental; línguas não-indoeuropeias do Subcontinente Indiano, línguas da China Interior, pertencentes ao subgrupo das siníticas, de uso urbano ou não; e

[4] A ressalva é de que a tipologia de complexidade da autora considera línguas de maneira holística, classificando-as em isolantes, aglutinantes e fusionais; em Mattos (2023) e Mattos e Lazzarini-Cyrino (2024), embora haja resultados que apontam que maiores escores de complexidade estão associados a menor proporção de morfologia gramatical livre, vemos que a subsunção de línguas inteiras a um perfil tipológico encobre uma diversidade de estratégias dentro da própria língua, bem como classifica no mesmo nível línguas que muitas vezes apresentam diferenças significativas de parâmetros de complexidade.

línguas do Oriente Médio. Tipologicamente, estão representadas línguas que a classificação tradicional atribuiu a cada um dos três tipos – **isolantes**, **aglutinantes** e **fusionais**.

Estabelecemos e quantificamos duas séries de parâmetros a partir da análise e segmentação morfológica de sentenças codificando ações concluídas. A primeira série, referente a parâmetros de verbosidade (isto é, de expressão de material funcional na sentença) consiste em três parâmetros, a **contagem de conceitos funcionais**, a **aquisição de categorias morfológicas** (isto é, a expressão explícita de valores de alguma categoria morfológica, como tempo, modo, aspecto ou definitude) e a **redundância sintagmática**. Já a segunda, referente a parâmetros intrínsecos aos morfemas funcionais, consiste em dois parâmetros, o **tamanho em segmentos fônicos** e os **desvios da relação de biunivocidade forma-significado**.

Buscou-se definir valores mensuráveis para cada parâmetro, bem como estabelecer correlações balizadas em estatística, de modo a avaliar a hipótese de que há correlação entre um maior número de material funcional expresso por proposição (representado por maiores parâmetros de verbosidade) e morfemas funcionais dotados de mais complexificações intrínsecas. A presença de tais complexificações em resposta a uma maior verbosidade representaria as consequências morfofonológicas da acomodação de material morfológico obrigatório no espaço limitado da sentença.

1. Avaliando a complexidade de construções

As línguas do mundo apresentam uma grande diversidade de estratégias para veicular proposições, e as possibilidades para as construções, ao se compararem línguas através do mundo, são inúmeras, e não raro apresentam uma grande variação de complexidade. Muitas vezes, porém, ao se falar em complexidade, o que se tem em mente é a exigência de expressar na morfologia uma série de categorias funcionais[5] para a construção de uma sentença válida numa dada língua, enquanto línguas tidas

[5] Não se confunda categoria com os diversos valores possíveis para uma dada categoria; GÊNERO é uma categoria, tendo masculino e feminino como os dois valores possíveis dessa categoria no português, e masculino, feminino e neutro como os três valores para a categoria no latim, por exemplo. "Categoria" no presente estudo equivale aos conceitos de "feature", para Zwicky (1985) e Kibort (2010) e "category" para Matthews (1972, 1991), Mel'čuk (1993) e Stump (2005), sendo as categorias aqui grafadas em versalete. "Valor" corresponde aos conceitos de "value" para Zwicky (1985) e Kibort (2010), "property" para Matthews (1972, 1991) e Stump (2005) ou "grammeme" para Mel'čuk (1993), e os diferentes valores vêm grafados em texto normal.

como **mais simples** são aquelas que exigem menor expressão de material para além do conteúdo lexical. Compare-se uma mesma proposição numa língua que exige maior expressão de categorias morfológicas para construir sentenças válidas, o pastó, e uma língua que as exige em muito menor grau, o mandarim:

1. Pastó (Afeganistão e Paquistão: indoeuropeia, irânica)
   زه ډېر خوندور توتان خورم.

za	**ḍer**	**-Ø**	**xwandawər**	**-Ø**	**tut**
1SG.STR.DIR	muito	-PL.M.DIR	saboroso	-PL.M.DIR	amora

-ān	**Ø**	**-xor**	**-əm**
-PL.M.DIR	CONT	-comer.PRS	-1SG

   “Estou comendo muitas amoras saborosas.” (David: 2014)

2. Chinês mandarim (China: sino-tibetana, sinítica)
   我正在吃很多好吃的桑葚

wǒ	**zài**	**chī**	**hěnduō**	**hǎochī**	**de**	**sāngrèn**
1	CONT	comer	muitos	saboroso	ATTR	amora

   “Estou comendo muitas amoras saborosas.”, conhecimento pessoal.

Frise-se, no entanto, que não só o pastó expressa mais categorias em sua sentença, como os diferentes valores das categorias muitas vezes aparecem associados num mesmo morfema, enquanto os morfemas do mandarim expressam apenas um valor cada. Igualmente, no mandarim há morfemas que carregam valores de apenas uma categoria gramatical cada, como primeira pessoa (**wǒ**), ou aspecto contínuo (**zài**); já no pastó, o morfema que veicula as categorias de NÚMERO, GÊNERO e CASO em seus valores plural, masculino e direto não só é poliexponente[6], veiculando três categorias em uma só forma, como também tem diversas realizações possíveis, {-ān} (**tut-ān**, “amoras”); {-Ø} (**gdən-Ø**, “painços”); {-una} (**čanel-una**, “canais”); {-i} (**čārwāk-i**, “oficiais”); {-ina} (**šay-ina**, “coisas”); e {-ə} (**sāṛ-ə**, “frios”). Os

[6] A poliexponência, isto é, veiculação de mais de uma categoria gramatical pelo mesmo morfema, é por vezes trazida na literatura como “exponência cumulativa”, como em Stump (2001), por exemplo. Buscamos trazer o termo utilizado na acepção de Bickel e Nichols por ser também a nomenclatura adotada pelo *World Atlas of Language Structures* (WALS, Dryer; Haspelmath, 2013), marco de estudos em tipologia e atlas prontamente acessível pela internet.

fenômenos da poliexponência e extensa alomorfia (especialmente a alomorfia condicionada à existência de classes de declinação/conjugação) são, **grosso modo**, as características mais imediatamente perceptíveis das línguas do perfil tipológico das **fusionais**, cujas palavras, mais que se decompor em subunidades prontamente destacáveis mapeando forma-significado de maneira clara, organizam-se em paradigmas.

1.1. Os parâmetros de complexidade

Em 1945, a publicação do artigo **Morphological analysis and synthesis** por Valter Tauli, apesar de não ter despertado muita atenção nos círculos da linguística ocidental (Schwegler: 1991), representou um momento fulcral para o debate da complexidade. Nela, o linguista estoniano elenca 11 critérios para se determinar se uma construção – e não uma língua inteira – é mais ou menos **analítica**. Vejam-se:

> I. linearidade, sendo uma forma linear mais analítica que uma forma não-linear;
> II. presença ou não de alomorfia em bases, sendo formas invariáveis mais analíticas que formas alternativas;
> III. presença ou não de alomorfia em morfemas, sendo formas invariáveis mais analíticas que formas alternativas;
> IV. ocorrência independente das bases, sendo as independentes mais analíticas;
> V. ocorrência independente dos morfemas, sendo os independentes mais analíticos;
> VI. força fonética do morfema, sendo um morfema foneticamente forte – silábico, por exemplo – mais analítico que um não-silábico;
> VII. se os elementos de uma forma são separáveis – sendo assim mais analíticos – ou não;
> VIII. se a ordem desses elementos pode ser trocada – sendo assim mais analíticos – ou não;
> IX. se no caso de coordenação, o morfema aparece em apenas uma palavra, sendo mais analítico – ou em todas, mais sintético;
> X. se no caso de subordinação, há concordância ou não, sendo expressões sem concordância mais analíticas;
> XI. ordem dos elementos, sendo ordem progressiva mais analítica que a regressiva.

Percebe-se, de um rápido exame dos critérios elencados por Tauli, que muitos dos itens se referem a medidas de complexidade, como mais tarde seriam encaradas no começo do século XXI. Em especial, os critérios I-III e VI se referem à **opacização morfológica**, que Peter Trudgill (2011) à frente trará como uma de

suas complexificações[7], e os critérios IV e VII podem ser associados à expressão explícita de categorias morfológicas.

Em sua obra **The growth and maintenance of linguistic complexity**, Östen Dahl (2004) propõe que a complexidade tem a ver com as regras que mapeiam conteúdo e forma, enunciando-a sob a forma de uma questão-chave: dado um conteúdo X, quantas regras são necessárias para exprimir X em dada língua? Isto é, ao analisarmos construções de diferentes línguas, vemos a medida que tais construções se afastam de um hipotético estado de mera **concatenação irrestrita**[8] de seus elementos básicos (lexicais), com a expressão de mais ou menos conceitos gramaticais e/ou funcionais variando de língua para língua. Numa construção, o grau de desvios desse estado hipotético de mera concatenação dos elementos lexicais é chamado pelo autor de **verbosidade**, ou dispensabilidade translinguística.

Perceba-se que, assim como na proposta de Tauli (1945), para Dahl, as diferentes construções é que são mais ou menos complexas sob essa acepção, e não necessariamente a língua como um todo. Em Mattos e Lazzarini-Cyrino (2021), vemos que línguas de um mesmo tipo holístico podem apresentar parâmetros de complexidade grandemente diferentes, e subsumi-las ao mesmo tipo poderia encobrir essa diversidade de estratégias de realização. Deste modo, evitar a tipologia holística, ou, no mínimo, associá-la a uma caracterização empírica de parâmetros mensuráveis, permite a compreensão não só da complexidade das línguas em sua diversidade de construções, como também o desenvolvimento de tal complexidade e os processos que lhe são subjacentes.

---

[7] O conceito de **complexificações**, elaborado por Peter Trudgill (2011), busca detalhar o ganho da complexidade morfossintática de maneira qualitativa, caracterizando-o como processos ao longo do tempo. São quatro as complexificações elencadas pelo autor, os processos por meio dos quais as estruturas morfossintáticas das línguas se tornam mais complexas: aquisição de categorias morfológicas, aumento na redundância sintagmática, opacização morfológica e irregularização. O enfoque dado ao se usarem tais conceitos, portanto, é mais dinâmico que simplesmente descritivo, permitindo investigar o surgimento e elaboração do aparato gramatical por processos diacrônicos, como a gramaticalização e a fusão de morfemas.

[8] Podemos compreender tal estado de concatenação irrestrita ao apelar para o exemplo da lista de compras em ímãs de geladeira, dado pelo próprio Dahl (2004); com o fito de informar os demais moradores de uma casa da necessidade de comprar duas bananas, isto é, transmitir o conteúdo proposicional <DOIS, BANANA>, é possível reunir um ímã com a imagem da fruta e um ímã com o número nas sequências **BANANA 2** ou **2 BANANA**. A linguagem natural humana, porém, raramente funciona dessa forma. Em línguas como o português e o inglês, há um primeiro desvio da concatenação irrestrita, uma restrição de ordem: o quantificador vem antes do quantificado. Ademais, nas duas línguas, há um segundo desvio: a obrigatoriedade de adicionar o morfema de plural (que é {-s} nos dois casos). Por fim, a expressão em português conta com mais uma regra, a atribuição de gênero: **banana** é um substantivo feminino, não sendo licenciada pela gramática dessa língua a expressão ***dois bananas**, e sim **duas bananas**. Pode-se dizer que a expressão em português **duas bananas** tem maior complexidade absoluta que sua respectiva em inglês, **two bananas**, devido à expressão de um conceito a mais, a categoria GÊNERO.

Além da necessidade de se explicitar morfologicamente uma série de categorias funcionais, uma construção de uma dada língua se torna mais **verbosa** por conta, por exemplo, de **redundância sintagmática**. A redundância sintagmática representa um passo a mais na aquisição de categorias; não só explicitamente codificadas, elas também são repetidas ao longo de um mesmo sintagma, sendo informação sobressalente de um ponto de vista estritamente comunicativo. O resultado desse processo é a existência de morfologia contextual (Booij: 1993) ou morfossintática (Kibort: 2008).

Outras vezes, por complexidade, pretende-se descrever uma série de processos que ocorrem no seio da morfologia gramatical-funcional de uma dada língua, como extensa alomorfia, expressão de mais de uma categoria gramatical pelo mesmo morfema (o que se chama de poliexponência segundo Bickel e Nichols: 2007), dentre outras características que não se referem à adição de mais material fonológico à sentença, mas sim a características intrínsecas aos morfemas daquela língua – sendo caracterizadas como mais simples as línguas cujos morfemas funcionais são mais facilmente decomponíveis e apresentem morfemas funcionais invariáveis e portadores de apenas uma categoria gramatical. Tais associações, iniciadas desde a caracterização das línguas fusionais sintéticas por Schlegel como mais elaboradas, passando por ao menos seis dos onze itens da lista de Tauli (1945)[9], atinge na proposta das complexificações de Trudgill (2011) um alto grau de estruturação. O autor reúne estes fenômenos sob o rótulo de **opacização morfológica**, que se refere ao surgimento de quaisquer desvios da relação biunívoca forma ⇔ significado. Estes desvios tornariam a interpretação do signo linguístico mais opaca e menos imediatamente decomponível, aumentando a complexidade do sistema.

Um último parâmetro, intrínseco aos morfemas, é seu tamanho. Desde o item VI da lista de Valter Tauli (1945), vemos um aceno, embora indireto, ao fato de que morfemas independentes (partículas) ou ao menos afixos silábicos são indício do que ele chamou de **analiticidade** e que seria associado a uma menor complexidade com o findar do século XX. Menores morfemas gramaticais-funcionais são um fenômeno trazido como indicativo de maior maturidade de uma dada estrutura linguística (Dahl: 2004); da mesma forma, em sua análise escalar, Hopper e Traugott (2023)

[9] Processo de cooptação de material inicialmente lexical em material gramatical, ocorrendo sob a forma de erosão morfológica e branqueamento semântico (**semantic bleaching**), como a conversão de formas verbais plenas do verbo **haver** em sufixos de tempo-modo-aspecto e número-pessoa no futuro sintético das línguas românicas. Remetemos à leitura de Hopper e Traugott (2023) e Croft (2000) para melhor compreensão do extenso debate acerca do tema.

elencam erosão morfológica e concomitante redução de tamanho como características de um intenso grau de gramaticalização[10]. Não raro, as línguas que a classificação holística convencionou tratar por **isolantes** apresentam morfemas funcionais de tamanho considerável, em geral independentes e plenamente silábicos (Mattos: 2023; Mattos e Lazzarini-Cyrino: 2024), e essas línguas são justamente tratadas como mais simples pela mesma literatura (Mattos e Lazzarini-Cyrino: 2021).

1.2. Definição da unidade de análise

Avaliar a complexidade de uma língua em sua inteireza necessariamente implicaria a medição da complexidade de todas as suas possíveis construções, seguida de algum cálculo para lhe tirar a média, de maneira a obter valores estimados para a língua de maneira global. A viabilidade de tal esforço já é duvidosa para uma única língua – revelando-se virtualmente impossível em curto prazo para estudos em tipologia linguística (Nichols: 2009), que não raro comparam dezenas de línguas. Avaliar a complexidade de um tipo de construção, porém, é algo mais imediatamente exequível. Orações principais, declarativas, ativas e codificando ações concluídas revelam-se um objeto de análise interessante, não apenas por sua alta frequência e importância no discurso espontâneo (Givón: 2008), como também pela possibilidade de, em se as analisando, avaliar e comparar a expressão de categorias gramaticais em diferentes línguas, a exemplo de TEMPO, MODO, ASPECTO, CASO, VOZ, NUMERO e PESSOA. Sobre a natureza de sentenças declarativas e ativas, por ele chamadas de **tema**, e sua saliência cognitiva e caráter menos marcado na comunicação oral espontânea, Talmy Givón afirma:

> No contexto da comunicação oral cotidiana, [a sentença] passiva é marcada. No contexto do discurso acadêmico escrito, o ativo. A marcação, portanto, diz respeito não apenas a categorias linguísticas, mas também aos contextos comunicativos nos quais elas estão inseridas. E é provavelmente razoável supor que a comunicação formal, escrita, acadêmica é o tipo de discurso marcado, e **a comunicação oral cotidiana face a face, a não-marcada** [...] **se o tema é, por uma enorme ordem de magnitude, o tipo de oração mais comum no discurso humano, então é provável que tenha um status cognitivo privilegiado** (2008, pp. 39-40, grifos nossos).

Aqui, portanto, o objeto a ter seus padrões estruturais mensurados e comparados são as construções codificando ações concluídas[11], corporificadas em sentenças principais, declarativas e ativas.

1.3. Metodologia

A amostragem consistiu em selecionar 50 línguas, divididas em 10 grupos com cinco línguas cada, feita a partir de algum grau de conveniência; todos os 10 grupos apresentaram um membro com perfil de língua franca, enquanto a macrofamília indoeuropeia esteve representada em dois grupos, embora distantes cultural, histórica, filogenética e geograficamente, as línguas irânicas e as línguas românicas. Não se ignorando a conveniência presente no delineamento amostral, buscou-se uma amostra o mais diversa possível, dada a natureza exploratória deste primeiro estudo e as limitações de tempo e recursos. Outrossim, a despeito de haver línguas participantes de diversos *Sprachbünde*, nenhuma área linguística específica enviesou os dados. Em geral, os grupos foram baseados em critérios filogenéticos, de modo a amostrar um grande número de diferentes famílias; um único grupo, o 8., das línguas do Alto Rio Negro, foi formado por uma combinação de critérios areais e genéticos: possui membros do *Sprachbund* do rio Vaupés (tucano, desano, hup e dâw), bem como membros da família nadahup (hup, dâw e nadëb, este último fora do referido *Sprachbund*), devido à escassez de material de línguas da família nadahup, e ao reduzido número de línguas da referida família.

A tabela 1 abaixo sintetiza as 50 línguas amostradas, suas afiliações genéticas, origens geográficas e a fonte dos dados utilizados:

Família	Língua	Origem	Fonte de dados
1. Irânicas	Persa	Irã	Mahootian e Gebhardt: 1997
	Talyshi	Azerbaijão	Schulze: 2000; Stilo: 2023
	Pastó	Afeganistão e leste do Paquistão	David: 2013
	Osseta iron	Ossétia	Erschler: 2020; Belyaev: 2014;

[11] Ou seja, sentenças cujas predicados principais são verbos agentivos télicos perfectivos – não necessariamente no passado.

			Belyaev e Vydrin: 2011
	Yaghnobi	Tajiquistão	Bird: 2007
2. Sinítica	Mandarim	China: planície do rio Amarelo	Her: 1996; Kuo: 2010; Smith e Erbaugh: 2005; Thompson: 1973
	Wutun	China: Qinghai	Sandman: 2016
	Shaowu	China: Fujian	Ngai: 2021
	Gan de Yichun	China: Jiangxi	Li: 2018
	Min de Hui'an	China: Fujian	Chen: 2020
3. Tupi-guarani	Nheengatu	Brasil: cabeça do cachorro	Cruz: 2011
	Wayampi	Guiana Francesa	Copin: 2012
	Achê	Paraguai	Rößler: 2008, 2015, 2018
	Kamaiurá	Brasil: Mato Grosso	Seki: 2000
	Araweté	Brasil: Pará	Solano: 2010
4. Românicas	Português brasileiro	Brasil	Santiago: 2019
	Normando Jersês	Reino Unido: Jérsei	Liddicoat: 2011
	Tuatschin	Suíça	Maurer-Cecchini: 2021
	Catalão	Espanha: Catalunha	Hualde: 1992
	Romeno	Romênia	Alboiu e Motapanyane: 2000
5. Austronésias	Malaio bazar	Singapura	Bao e Aye: 2012; Aye: 2013
	Tukangbesi	Indonésia: ilhas Célebes	Donohue: 2011
	Dhao	Indonésia: pequenas ilhas de Sonda	Balukh: 2020
	Kéo	Indonésia: ilha de	Baird: 2013

		Flores	
	Paiwan	Taiwan	Chang: 2006
6. Austroasiáticas	Khasi	Nordeste da Índia	Nagaraja: 2014
	Khmer	Camboja	Bisang: 2015
	Wa	Myanmar e Sudoeste da China	Ma: 2012
	Kharia	Nordeste da Índia	Peterson: 2010
	Santali	Nordeste Índia	Ghosh: 2015
7. Bantu	Suaíli	África oriental	Mursell: 2018; Seidl e Dimitriadis: 1997; Vitale: 2019
	Gikuyu	Quênia	Jeon: 2015
	Gyeli	Camarões	Grimm: 2021
	Cuwabo	Moçambique	Guérois: 2015
	Fwe	Namíbia	Gunnink: 2018
8. Alto R. Negro	Tucano ye'pâ-masa	Brasil e Colômbia	Ramirez: 1997
	Desano	Brasil e Colômbia	Lima Silva: 2012
	Hup	Brasil: cabeça do cachorro	Epps: 2008
	Dâw	Brasil: cabeça do cachorro	Carvalho: 2006
	Nadëb	Brasil: cabeça do cachorro	Epps e Obert: 2021; Weir: 1984
9. Mandê	Bambara	Mali	Creissels: 2007
	Hainyaxo	Mali	Heath: 2023
	Jalkunan	Burkina Faso	Heath: 2023
	Zialo	Guiné	Babaev: 2010
	Mano	Guiné e Libéria	Khachaturyan: 2014
10. Semíticas	Árabe de Juba	Sudão do Sul	Manfredi: 2013

	Faifi	Sudeste da Arábia Saudita e Iêmen	Alfaife: 2018
	Bathari	Omã	Gaspari: 2018
	Maltês	Malta	Borg e Azzopardi-Alexander: 1997
	Socotri	Socotra	Aloufi: 2016

Tabela 1: Total de línguas amostradas, agrupadas por família, com respectivas origens geográficas e fontes dos dados utilizados.

Em seguida, uma vez obtidos valores para os parâmetros globais das línguas, prosseguimos para a obtenção das sentenças. Para cada língua, dez sentenças já glosadas exprimindo ações concluídas foram obtidas a partir de material publicado, conforme exposto na tabela 1; foi dada preferência ao uso de gramáticas descritivas[12], em especial as que continham pequenas narrativas anotadas em suas seções finais ou anexas.

O critério para a seleção das dez sentenças foi o de codificarem ações concluídas. Destas dez, para cada língua, seis consistiram em ações concluídas com dois participantes, duas consistiram em transferências com três participantes, e outras duas sentenças, uma ação com apenas um participante.

As sentenças, então, foram anotadas num banco de dados, respeitando-se sua segmentação morfológica e glosas. Para cada uma delas, explicitaram-se seus conceitos básicos, de natureza lexical (**concretos radicais** seguindo o entendimento de Sapir (1921), ou **raízes** segundo Haspelmath (2024)), que denotam entidades, propriedades ou eventos; com base neles, o conteúdo proposicional de cada sentença foi anotado em maiúsculas, contendo os (geralmente dois) participantes e uma ação. Veja-se, por exemplo, a sentença persa **namero neveštæm** "eu escrevi a carta"; seu conteúdo proposicional anotado foi <EU, CARTA, ESCREVER>. Todos os conceitos não-básicos[13] presentes na glosa segmentada da sentença foram considerados **funcionais**.

[12] Apenas o mandarim, o português brasileiro e o suaíli tiveram sentenças coletadas a partir de material de outra natureza que não gramáticas descritivas.

[13] Isto é, que não tenham natureza lexical, não sendo um dos participantes (no caso, EU e CARTA) ou a ação em si (ESCREVER). A problemática acerca da divisão lexical/funcional é um extenso debate, que fugiria do escopo do presente trabalho; escolheu-se, operacionalmente, definir os participantes e a ação como lexicais, e os demais conceitos como funcionais. Para uma abordagem esclarecedora do

Cada morfema portador de conceitos funcionais foi avaliado de forma binária (0 ou 1) de acordo com os parâmetros (i) alomorfia; (ii) poliexponência; (iii) aquisição de categorias morfológicas; (iv) redundância sintagmática; e (v) número de segmentos fônicos.

Língua	**Glosa**				**CF**	**ALO**	**POLI**	**ACM**	**RED**	**NS**
	name	**-ro**	**nevešt**	**-æm**	-OBJ.DEF	1	1	1	1	2
			escrever		<>.PST	0	0	1	0	0
Persa	carta	-OBJ.DEF	.PST	-1sg	-1sg	0	1	0	1	2

Tabela 2: Segmentação e glosa da sentença persa. CF: conceitos funcionais; ALO: alomorfia; POLI: poliexponência; RED: redundância sintagmática; ACM: aquisição de categoria morfológica; NS: número de segmentos do morfema.

Para ilustrar a diferença entre os parâmetros CF, a contagem de conceitos funcionais, e ACM, a aquisição de categorias morfológicas, vejamos que dentre os morfemas da sentença araweté em (3), **he** 1SG recebeu 0 para ACM uma vez que representa um dos participantes da ação, informação indispensável (ou prontamente recuperável no discurso, nos casos em que a língua licencia elisão de algum participante na sentença).

3. Araweté (Brasil: tupi-guarani)

a	**-ʔu**	**ku**	**he**	**takunerẽ**
1SG.SBJ	-comer	FOC	1SG	tucunaré branco

"Eu comi o tucunaré branco." (Solano: 2010)

No entanto, o afixo **a** 1SG.SBJ veicula também a noção de sujeito, representando a aquisição de uma categoria e sua expressão por meios morfológicos, recebendo 1 para esse parâmetro. Caso veiculasse apenas número-pessoa, assim como **he**, receberia 0. Igualmente, vogais temáticas, temas, vogais finais, morfemas de pertencimento a classes paradigmáticas, aumentos vocálicos e demais morfemas de caráter funcional, mas que não expressam um valor de uma categoria gramatical, computados em CF, não foram considerados no presente estudo para cômputo de ACM.

tópico, com referência ao complexo processo de gramaticalização e à natureza gradual e dinâmica do surgimento do inventário de morfemas funcionais de uma língua, remetemos à leitura de Croft (2000).

Outrossim, os valores contabilizados morfema-a-morfema tiveram suas médias retiradas para cada sentença, conforme a tabela 3 abaixo.

Língua	Sentença	CF	ALO	POLI	ACM	RED	NS
Persa	**1. namero neveštæm**	3	1	1	2	1	1,33

Tabela 3: Anotação dos valores médios dos parâmetros da primeira das dez sentenças do persa.

Destas, subsequentes médias foram calculadas, também para a obtenção de valores para cada língua, conforme a tabela 4. O parâmetro **desvio da biunivocidade (BIUN)** foi adicionado à tabela nesta etapa, correspondendo à expressão de um mesmo significado por formas diferentes (alomorfia), ou de mais de um significado pela mesma forma (poliexponência), sendo, portanto, a média por língua dos parâmetros **ALO** e **POLI**:

Família	Língua	CF	ALO	POLI	BIUN	ACM	RED	NS
	Persa	3,7	0,7	0,5	0,6	1,4	0,7	1,475
	Talysh	4,6	1,7	1,1	1,4	1,5	0,9	1,483
	Pastó	9,1	3,7	2,3	3	0,44	0,9	1,17
	Osseta iron	5,6	0,9	1,3	1,1	2,6	1,1	0,993
IRA	Yaghnobi	5,2	0,2	0,8	0,5	1,7	0,6	1,55

Tabela 4: Médias dos valores de dez sentenças das línguas da família das irânicas. BIUN: desvio da relação de biunivocidade.

Por fim, foram processados os dados no **software** R (R CORE TEAM, 2024), utilizando-se os valores de complexidade obtidos por língua.

2. Correlações

Conforme exposto, os parâmetros escolhidos para o presente estudo se dividem em dois grupos; um primeiro, composto por **contagem de conceitos funcionais** (CF), **aquisição de categorias morfológicas** (ACM) e **redundância sintagmática** (RED) diz respeito a medidas de verbosidade, isto é, de uma complexidade *in praesentia*, sintagmática, associada a uma obrigatoriedade de exprimir mais material morfológico para construir uma sentença bem formada

numa dada língua. Doravante, serão chamados de **parâmetros de verbosidade**. O segundo grupo diz respeito a características que, embora apontem para uma maior complexidade (ou **maturidade** dos fenômenos linguísticos *sensu* Dahl: 2004), são intrínsecas ao morfema, e não dizem respeito necessariamente a uma maior expressão de categorias funcionais – aqui se computam, portanto, o **tamanho do morfema em segmentos fônicos** (NSEG) e o **desvio da relação de biunivocidade forma-significado** (BIUN). Serão doravante chamados de **parâmetros intrínsecos**.

Veremos a seguir, portanto, todas as correlações de Pearson possíveis entre parâmetros de verbosidade e parâmetros intrínsecos, para melhor avaliar a relação entre a expressão de material funcional e complexificações no inventário morfêmico expresso.

2.1. Contagem de conceitos funcionais

Houve forte correlação positiva (r = 0,828) entre BIUN e CF – é dizer, quanto mais conceitos funcionais a língua expressa por sentença, mais seus morfemas funcionais se afastam de uma relação biunívoca (de um para um) entre forma e significado, apresentando maiores taxas de alomorfia e poliexponência (figura 1).

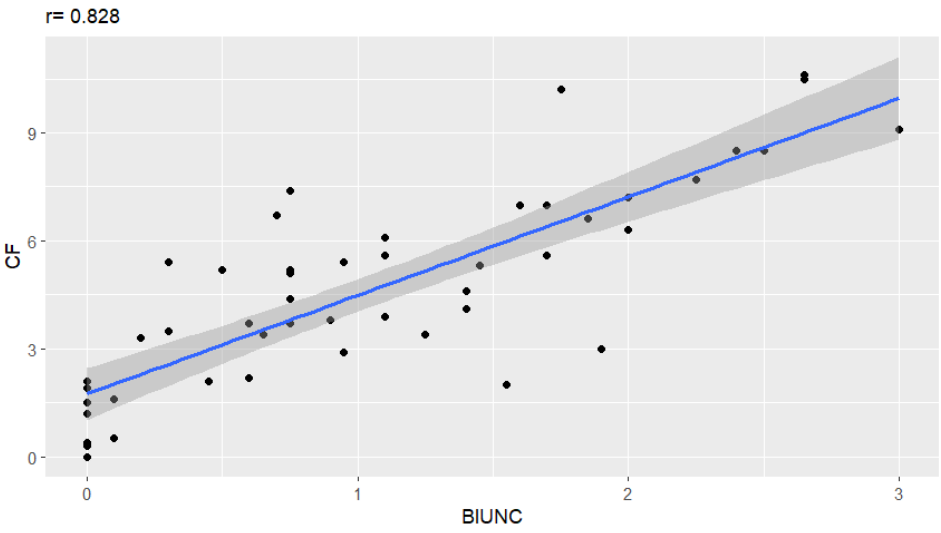


Figura 1: Correlação entre desvio da relação de biunivocidade (BIUN) e contagem de conceitos funcionais (CF) nas línguas.

Já com o tamanho do morfema em número de segmentos (NSEG), a correlação foi moderada e negativa (-0,606); isto é, quanto mais conceitos funcionais expressos, menor o tamanho do morfema funcional, conforme a figura 2 demonstra:

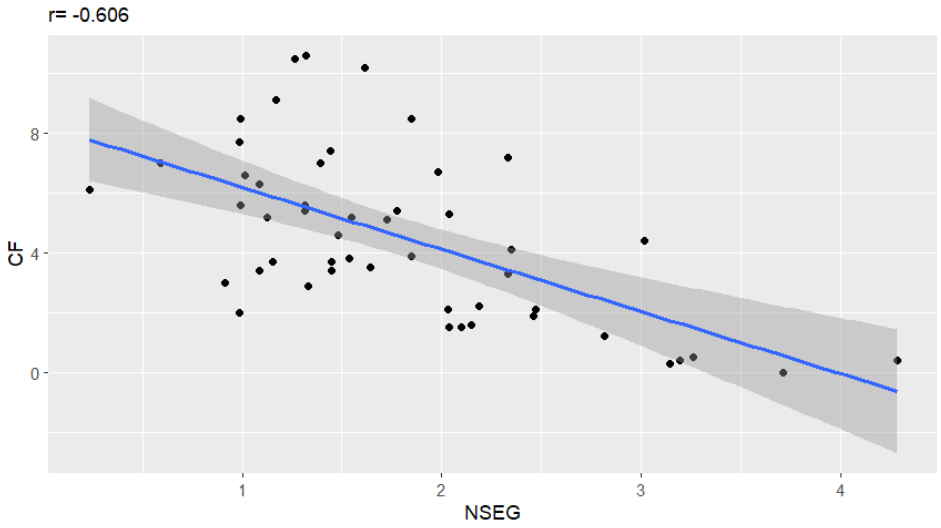


Figura 2: Correlação entre tamanho do morfema em segmentos (NSEG) e contagem de conceitos funcionais (CF) nas línguas.

2.2. Aquisição de categorias morfológicas

O parâmetro aquisição de categorias morfológicas apresentou, assim como conceitos funcionais, correlação positiva com desvios da biunivocidade (r = 0,688) embora moderada, e não forte (figura 3).

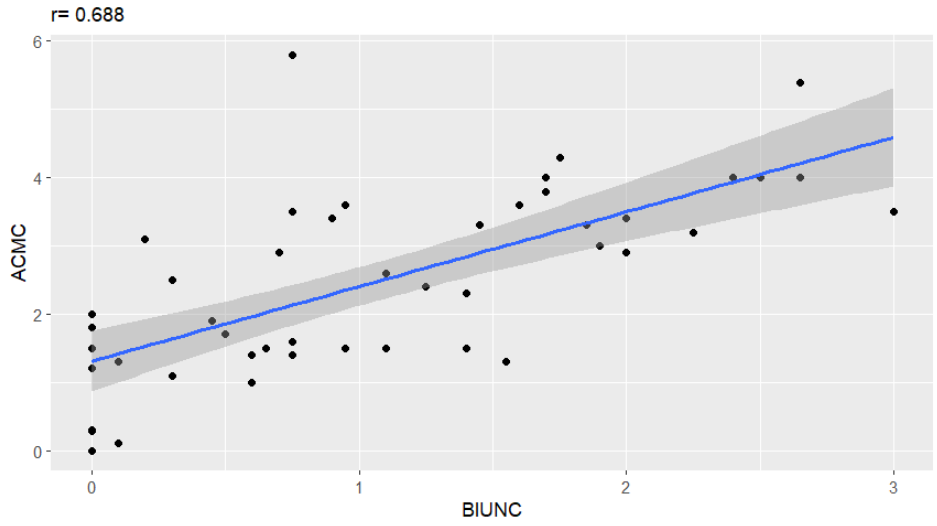


Figura 3: Correlação entre desvio da relação de biunivocidade (BIUN) e aquisição de categorias morfológicas (ACM) nas línguas.

Da mesma forma, a correlação entre o parâmetro aquisição de categorias morfológicas e tamanho do morfema também foi moderada, embora negativa (-0,596, figura 4).

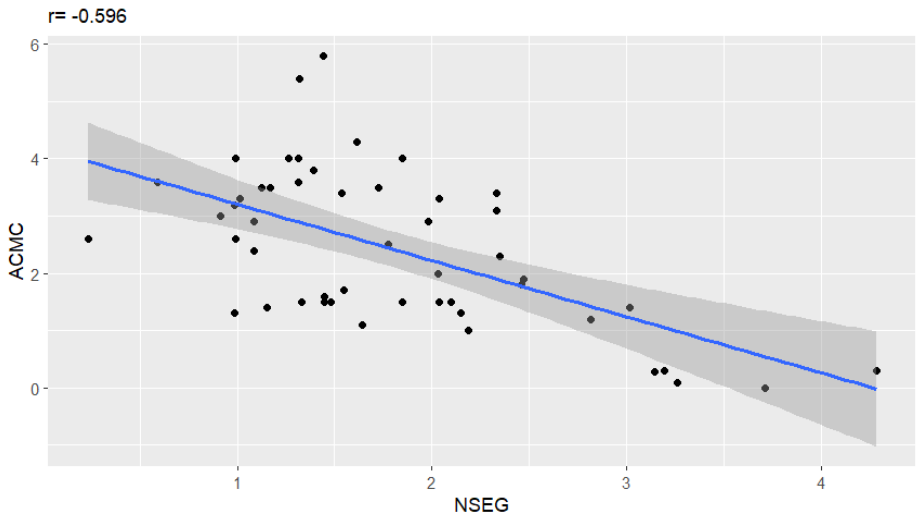


Figura 4: Correlação entre tamanho do morfema em segmentos (NSEG) e aquisição de categorias morfológicas (ACM) nas línguas.

Da mesma forma, a correlação entre o parâmetro aquisição de categorias morfológicas e tamanho do morfema também foi moderada, embora negativa (-0,596, figura 5).

2.3. Redundância sintagmática

Por fim, o parâmetro redundância sintagmática também apresentou correlação positiva e moderada (0,676) com desvios da biunivocidade, conforme a figura 5:

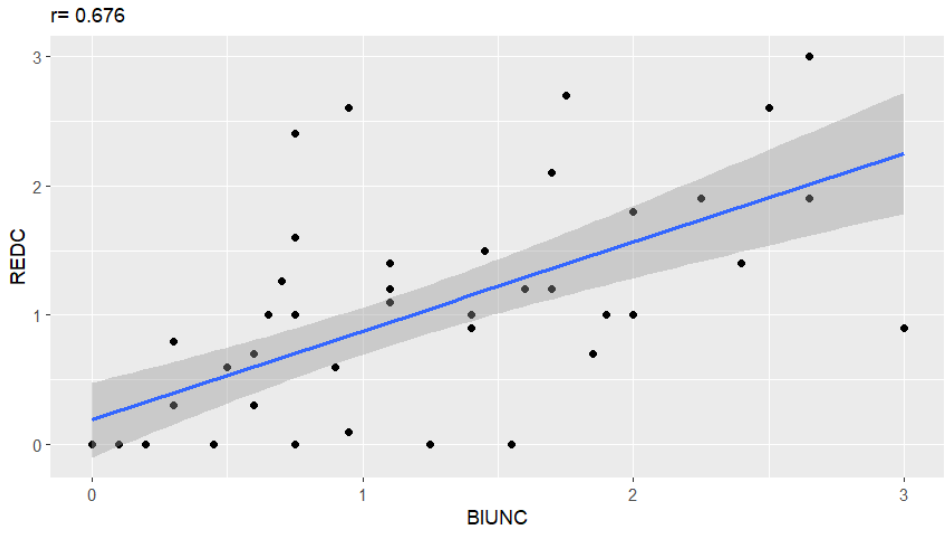


Figura 5: Correlação entre desvio da relação de biunivocidade (BIUN) e redundância sintagmática (RED) nas línguas.

Com o parâmetro tamanho do morfema, no entanto, a correlação se revelou fraca, porém negativa como com os demais parâmetros de verbosidade (-0,44):

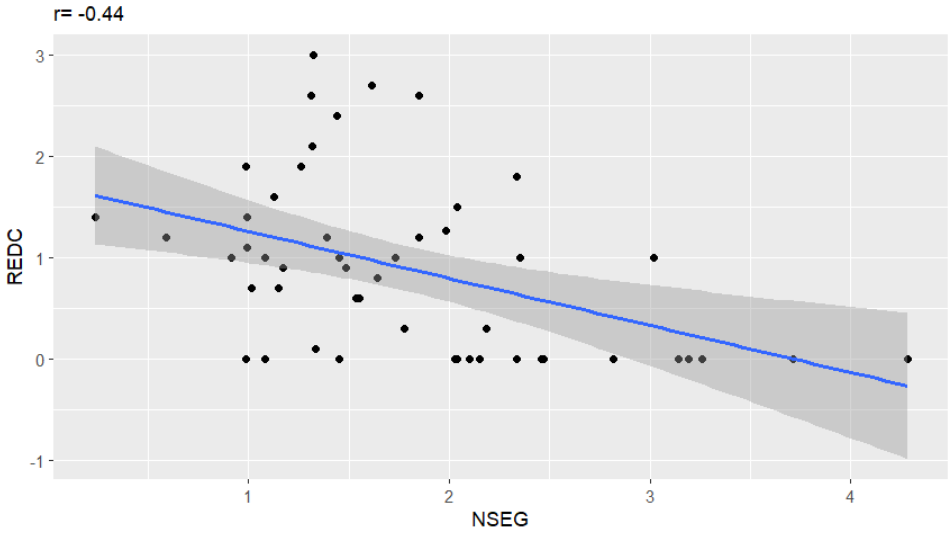


Figura 6: Correlação entre tamanho do morfema em segmentos (NSEG) e redundância sintagmática (RED) nas línguas.

De toda sorte, ao se plotarem os valores das 50 línguas para redundância sintagmática contra tamanho do morfema, colorindo os pontos por média de redundância sintagmática, podemos ver claramente que nenhuma língua com redundância sintagmática significativamente acima da média aparece entre as línguas com morfemas de tamanho grande (acima de 2,5 segmentos), como mostra a figura 7:

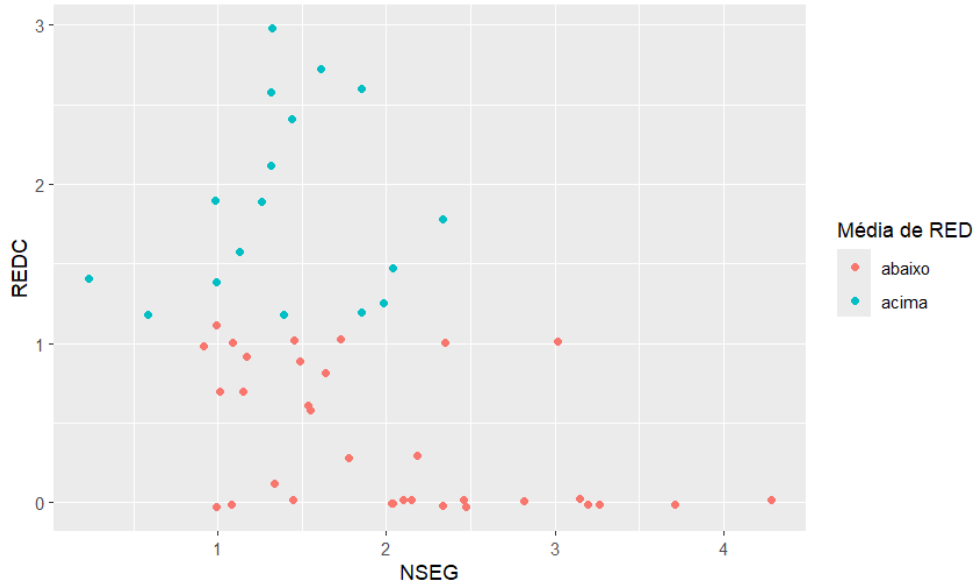


Figura 7: Valores de redundância sintagmática (RED) plotados contra valores de tamanho de morfemas (NSEG), coloridos por média de RED.

3. Implicações translinguísticas

A correlação significativa encontrada entre maior expressão *in praesentia*[14] de material funcional (aqui representada pelos parâmetros CF, ACM e RED) e medidas de complexidade intrínseca aos morfemas (o desvio da relação de biunivocidade e o tamanho em número de segmentos) pode ser compreendida como decorrente da necessidade de obrigatoriamente explicitar um maior aparato morfossintático numa proposição. Essa necessidade teria relação com redução de tamanho, permitindo a otimização da transmissão de um maior número de categorias, bem como com complexificações como a poliexponência e a alomorfia.

Ou seja, quanto maior a verbosidade numa sentença, menores os formativos em número de segmentos, de acordo com a hipótese aventada quanto à pressão exercida por uma maior complexidade de categorias gramaticais no espaço proposicional, que teria correlação com processos de acomodação morfofonológica, resultando em complexificações intrínsecas, bem como atuaria para favorecer formativos menores, por meio de erosão, fusão entre morfemas, e adoção de estratégias suprassegmentais caso a língua delas disponha. Não à toa, as línguas com

[14] *In praesentia* aqui faz referência a uma maior expressão na dimensão sintagmática de categorias morfológicas (contra um maior número de valores possíveis para cada categoria, isto é, paradigmática ou *in absentia*) e morfemas não-lexicais no geral. É dizer, maior verbosidade, presença de material funcional obrigatório nas sentenças.

maiores tamanhos de morfema (média de NSEG considerando as 50 línguas: 1,79) possuem em geral baixíssimos escores de conceitos funcionais, aquisição de categorias morfológicas e redundância sintagmática, com valores bem abaixo de suas médias (CF: 4,55; ACM: 2,43; RED: 0,89). A tabela 5 abaixo mostra as dez línguas com maiores morfemas:

Língua	**Grupo**	**NSEG**	**CF**	**ACM**	**RED**
Kéo	ASN	4,28	0,4	0,3	0
Khmer	AST	3,71	0	0	0
Wa	AST	3,26	0,5	0,1	0
Juba	SEM	3,19	0,4	0,3	0
Pasarmelayu	ASN	3,14	0,3	0,28	0
Tukano	RNG	3,01	4,4	1,4	1
Dâw	RNG	2,81	1,2	1,2	0
Hup	RNG	2,47	2,1	1,91	0
Gan de Yichun	SIN	2,45	1,9	1,8	0
Kharia	AUS	2,35	4,1	2,3	1

Tabela 5: As cinco línguas com maiores médias de número de segmentos por morfema funcional. ASN: austronésias; AST: austroasiáticas; SEM: semíticas; RNG: Rio Negro; SIN: siníticas.

Note-se também que, das línguas acima, a grande maioria não apresenta qualquer morfologia sintaticamente redundante em suas sentenças de ações. O kéo, língua austronésia, apresentou os maiores morfemas dentre todas as 50 línguas amostradas – morfemas que, em sua maioria, veiculam número-pessoa, ou são classificadores nominais. Apresentou valores de categorias funcionais e aquisição de categorias morfológicas muito próximos a zero, bem como nenhuma morfologia sintaticamente redundante. As sentenças em (4) ilustram as sentenças que codificam ações nessa língua, bem como seus morfemas funcionais de grande tamanho:

4. Kéo (Ilha de Flores: austronésia)

a. Morfema com cinco segmentos: [ŋgaʔe]

'ata	**ngga'e**	**dima**	**ka**
pessoa	CL	cinco	comer

"Cinco pessoas comeram." (Baird: 2013)

b. Morfema com oito segmentos: [ʔimu koʔo]

'imu ko'o	**rasa**	**ta**	**'aé**
3PL	provar	EMPH	água

"Eles provaram a água." (**id**.)

Entre as línguas que apresentaram os menores morfemas, estão línguas indoeuropeias, como o osseta iron, o jersês e o português brasileiro, e, especialmente, quatro das cinco línguas semíticas, o bathari, o faifi, o maltês e o socotri. Ambas as famílias possuem extensa poliexponência e fusão à base, processos que culminam num menor número de segmentos realizados para expressar categorias gramaticais, com destaque para as línguas semíticas, que frequentemente se valem de apofonia para realização de diversos valores gramaticais.

Língua	**Grupo**	**NSEG**	**CF**	**ACM**	**RED**
Bambara	MAN	1,08	3,4	2,4	0
Bathari	SEM	1,08	6,3	2,9	1
P. brasileiro	ROM	1,01	6,6	3,3	0,7
Osseta iron	IRA	0,993	5,6	2,6	1,1
Faifi	SEM	0,993	8,5	4	1,4
Jalkunan	MAN	0,986	2	1,3	0
Jersês	ROM	0,983	7,7	3,2	1,9
Hainyaxo	MAN	0,915	3	3	1
Maltês	SEM	0,588	7	3,6	1,2
Socotri	SEM	0,233	6,1	2,6	1,4

Tabela 6: Línguas com menores médias de número de segmentos por morfema funcional. IRA: irânicas; SEM: semíticas; MAN: mandê; ROM: românicas.

Igualmente figuram na tabela 6, com as línguas de menores médias de NSEG, o bambara, o jalkunan e o hainyaxo, línguas mandê que também apresentam poliexponência elevada e uso ubíquo de mecanismos suprassegmentais para veicular informação gramatical. Faz exemplo a categoria DEFINITUDE do bambara, que é veiculada por meio de um tom baixo flutuante[15]:

5. Bambara (Mali: mandê)

   a.

mùsò	**sén**	**-`**
mulher	perna	-ART

   "A perna de mulher." (Vydrin: 2016)

   b.

mùsó	**-`**	**sĕn**	**-`**
mulher	-ART	perna	-ART

   "A perna da mulher." (**id.**)

Pela natureza suprassegmental da expressão de muitos dos conceitos funcionais de línguas dessa família, não é de se espantar que tenham aparecido entre as línguas com menores morfemas funcionais, uma vez que tais formativos se computaram como tendo um valor 0 (zero) de NSEG.

Já o socotri, a língua com morfemas funcionais de menor tamanho médio dentre as 50, é ilustrado pela sentença em (6):

6. Socotri (Socotra: semítica)

toʔo	**-k**	**tɛfħah**
comer.PFV	-1SG.SBJ	maçã

   "Eu comi a maçã." (Aloufi: 2016)

Vê-se que o aspecto perfectivo é veiculado por meio de alteração na raiz verbal, e um formativo de um segmento apenas, **-k**, veicula as categorias de NUMERO-PESSOA e

[15] Tal formativo "manifesta-se como uma curva de entonação descendente antes de uma pausa ou em modificações tonais nos domínios adjacentes." (Vydrin: 2020, p. 272)

CASO; as sentenças do socotri possuem formativos que são caracteristicamente baixos em número de segmentos.

Quanto ao parâmetro BIUN (média: 1,027), vemos que, novamente, as línguas com menores valores possuem escores baixíssimos nos parâmetros de verbosidade. As 10 línguas com menor BIUN apresentaram CF (média: 4,55) e ACM (média: 2,43) em níveis muito abaixo da média, e nenhuma delas apresentou qualquer morfologia sintaticamente redundante (RED), conforme demonstra a tabela 7:

Língua	**Grupo**	**BIUN**	**CF**	**ACM**	**RED**
Dâw	RNG	0	1,2	1,2	0
Min de Hui'an	SIN	0	1,5	1,5	0
Árabe de Juba	SEM	0	0,4	0,3	0
Kéo	ASN	0	0,4	0,3	0
Khmer	AUS	0	0	0	0
Mandarim	SIN	0	1,5	1,5	0
Malaio bazar	ASN	0	0,3	0,28	0
Shaowu	SIN	0	2,1	2	0
Gan de Yichun	SIN	0	1,9	1,8	0
Achê	TUP	0,1	1,6	1,3	0

Tabela 7: Línguas com menores médias de desvio da relação de biunivocidade. RNG: Rio Negro; SIN: siníticas; SEM: semíticas; ASN: austronésias; AUS: austroasiáticas; TUP: tupianas.

As sentenças de nove das dez línguas da tabela 7 acima não apresentaram quaisquer morfemas poliexponentes ou com alomorfia (é dizer, nenhum que se desviasse da relação um-para-um entre forma e significado), de modo que várias dentre elas se prestam a exemplificar o fenômeno. Veja-se, portanto, o caso do dâw em (7):

7. Dâw (Brasil: nadahup)

wa'	**buy**	**dâr**	**mĩs**	**=ũ̃ũ̃j'**
urubu	jogar	PONT	jabuti	=OBJ

"O urubu jogou o jabuti." (Carvalho: 2016)

A sentença da língua não só apresenta poucos morfemas funcionais, como estes são invariáveis e monoexponentes (**dâr**, veiculando apenas aspecto pontual, e o clítico **ũũj'**, veiculando apenas o objeto direto). No outro extremo, dentre as dez línguas que apresentaram os maiores níveis de BIUN, nenhuma apresentou contagem de conceitos funcionais ou aquisição de categorias morfológicas abaixo da média:

Língua	**Grupo**	**BIUN**	**CF**	**ACM**	**RED**
Pastó	IRA	3	9,1	3,5	0,9
Romeno	ROM	2,65	10,5	4	1,9
Catalão	ROM	2,65	10,6	5,4	3
Santali	AUS	2,5	8,5	4	2,6
Faifi	SEM	2,4	8,5	4	1,4
Jersês	ROM	2,25	8,7	3,2	1,9
Bathari	SEM	2	6,3	2,9	1
Paiwan	ASN	2	8,2	3,4	1,8
Hainyaxo	MAN	1,9	4,6	3	1
P. brasileiro	ROM	1,85	7,1	3,3	0,7

Tabela 8: Línguas com maiores médias de desvio da relação de biunivocidade. IRA: irânicas; ROM: românicas; AUS: austrasiáticas; SEM: semíticas; ASN: austronésias; MAN: mandê.

Quanto à redundância sintagmática, apenas o português brasileiro apresentou um escore abaixo da média. As sentenças do português brasileiro foram amostradas de entrevistas presentes em Santiago (2019) com falantes de baixa escolaridade da zona rural do semiárido nordestino, o que representa uma variedade não-padrão. Encontraram-se instâncias de realização do plural nominal apenas no elemento mais à esquerda ("**papai comprou uns livrinho**"; "**essas carta eu fiz**"; "**Jade pegou os papel"),** o que contribui para reduzir a morfologia sintaticamente redundante de suas construções, em especial se comparadas às construções das demais línguas românicas. Uma delas, o romeno, foi a segunda língua com os maiores escores de BIUN, atrás apenas do pastó:

8. Romeno (Romênia: indoeuropeia, românica)

fat	**-a**	**ne**	**a**	**salutat**	**-t**
girl	-DEF.F.SG	1PL.ACC	3SG.PST.PRF	greet	-PTCP

"A garota nos saudou." (Alboiu e Montapanyane: 2000)

Percebam-se no exemplo acima os morfemas de tempo-modo-aspecto, poliexponentes como nas demais românicas, como também os morfemas de definitude-caso-número-gênero (como **-a** ou **ne**), que grandemente contribuíram para aumentar os escores do romeno, se comparado com as demais línguas de sua própria família.

4. Considerações finais

O presente artigo buscou medir parâmetros de verbosidade e de complexificações de morfemas funcionais em sentenças codificando ações de 50 línguas, bem como, munidos dos escores objetivos para cada parâmetro, estabelecer-lhes as correlações. Vimos que, acima de tudo, a presença de desvios da relação biunívoca 1:1 entre forma e significado está fortemente associada a maior expressão de material funcional na sentença. Ademais, o tamanho dos morfemas revelou-se, no presente estudo, inversamente proporcional à quantidade de material funcional expresso na sentença. Tais achados apontam para a existência de possíveis acomodações morfofonológicas do material funcional *in praesentia*, que apresentaria suas unidades constituintes se relacionando de maneira mais íntima, fundindo-se e exibindo variação alomórfica. Consistentemente, as línguas com morfemas de maior tamanho em número de segmentos e menos desvios da relação 1:1 forma-significado foram as que tradicionalmente foram tidas por mais **simples**, e que a classificação holística atribuiu ao tipo **isolante**. Nossos achados, porém, em vez de se balizarem em conceitos de difícil operacionalidade, como simplicidade **versus** complexidade, ou na terminologia holística, basearam-se em dados mensuráveis e estatisticamente processados, de maneira a demonstrar a existência de tais correlações de forma empírica e reprodutível.

Importante salientar que as relações aqui encontradas se circunscrevem aos morfemas funcionais de sentenças codificando ações, haja vista que a medição e caracterização de todos os tipos de construções de uma língua dificilmente seria uma

tarefa exequível, ainda mais de uma amostra com dezenas de línguas. Um possível rumo para investigação futura é inventariar, mensurar e testar as correlações entre parâmetros de morfemas em sentenças de outros perfis, como as que codifiquem estados, ou incluir mais línguas na amostra, em especial as que historicamente foram classificadas como **polissintéticas**, a exemplo de muitas línguas indígenas da América do Norte – o que também contribuiria para cobrir esta última área, que não foi incluída na amostra.

Palavras para Ana Scher

A publicação desta edição temática dedicada à Morfologia convida-nos a rememorar o legado intelectual e a postura científica de Ana Paula Scher, cuja atuação é de grande contribuição para os estudos morfológicos contemporâneos no Brasil. Em seu papel pioneiro e desbravador na introdução e expansão da Morfologia Distribuída no cenário nacional, a Profa. Ana Scher sempre aliou o desenvolvimento e avaliação dos modelos formais a um grande respeito pelos dados linguísticos.

Essa herança metodológica, fundada em olhar para as línguas a partir de suas evidências materiais, é uma das chaves que nos moveu em direção à tipologia linguística e ao enfoque quantitativo-indutivo que ampara o presente artigo. Destaca-se que Ana Scher transmitiu uma orientação de pesquisa baseada na mais estrita liberdade intelectual. Sua generosidade permitia que seus orientandos explorassem caminhos autônomos, o que acreditamos que tenha sido um importante fator contribuinte na recente expansão do interesse pelos estudos morfológicos no Brasil, sobretudo em perspectivas formalistas.

Investigar as correlações morfológicas entre cinquenta línguas sem o aprisionamento de tipologias holísticas é, em última análise, um desdobramento da provocação científica de Ana Scher: a de que a morfologia deve ser explicada pela riqueza de seus dados reais. A ela, nossa mais profunda e respeitosa gratidão.

Lista de abreviações

1	primeira pessoa	ATTR	atributivo
3	terceira pessoa	CL	classificador nominal
ACC	caso acusativo	CONT	contínuo
ART	artigo	DEF	definido

DIR	caso direto	PST	passado
EMPH	enfático	PRF	perfeito
F	feminino	PRS	presente
FOC	foco	PTCP	particípio
M	masculino	SBJ	sujeito
OBJ	objeto direto	SG	singular
PL	plural	STR	série pronominal forte
PONT	pontual		

Referências Bibliográficas

ALBOIU, Gabriela; MOTAPANYANE, Virginia. The generative approach to Romanian grammar: An overview. *Comparative studies in Romanian syntax*, p. 1-48, 2000.

ALFAIFE, Saleem Mohammed. *A grammar of Faifi*. California State University, Long Beach, 2018.

ALOUFI, Amani. *A Grammatical Sketch of Soqotri: With Special Consideration of Negative Polarity*. Northeastern Illinois University, 2016.

AYE, Daw Khin Khin. *Bazaar Malay: History, grammar and contact*. Universidade Nacional de Singapura. Tese de Doutorado, 2006.

BABAEV, Kirill. *Zialo: The newly-discovered Mande language of Guinea*. Munique: Lincom Europa, 2010.

BALUKH, Jermy. A grammar of Dhao: An endangered Austronesian language in Eastern Indonesia, Wacana. *Journal of the Humanities of Indonesia*, v. 22, n. 2, p. 522-533, 2020.

BAIRD, Louise. *A grammar of Keo: an Austronesian language of East Nusantara*. 2002.

BAO, Zhiming; AYE, Khin Khin. Bazaar Malay topics. In: ANSALDO, U. *Pidgins and creoles in Asia*. Amsterdã: John Benjamins Publishing Company, 2012. p. 151-167.

BELYAEV, Oleg; VYDRIN, Arseniy. Participle-converbs in Iron Ossetic: Syntactic and semantic properties. *Topics in Iranian Linguistics*. Wiesbaden: Dr. Ludwig Reichert Verlag, 2011.

BELYAEV, Oleg. Anaphora in Ossetic correlatives and the typology of clause combining. In: SUIHKONEN, Pirkko; WHALEY, Lindsay. *On Diversity and*

Complexity of Languages Spoken in Europe and North and Central Asia (Studies in Language Companion Series 164). Amsterdam: John Benjamins, 2014.

BICKEL, Balthasar; NICHOLS, Johanna: Inflectional morphology. In: SHOPEN, Timothy. *Language typology and syntactic description***.** Cambridge: Cambridge University Press v. 3, n. 2, 2007.

BIRD, Brian. *Aspects of Yaghnobi grammar*. Tese de Doutorado. University of Oregon. 2007.

BISANG, Walter. Modern Khmer. In: JENNY, Mathias; SIDWELL, Paul. *The handbook of Austroasiatic languages*. Leiden: Brill, 2015.

BOOIJ, Geert. Against split morphology. In: *Yearbook of morphology 1993*. Dordrecht: Springer Netherlands, 1993.

BORG, Albert J.; AZZOPARDI-ALEXANDER, Marie. *Maltese*. Abingdon: Psychology Press, 1997.

CARVALHO, Maurício. *Aspecto verbal na língua Dâw*. Tese de Doutorado. Universidade de São Paulo, 2016.

CHANG, Anna Hsiou-chuan. *A reference grammar of Paiwan*. 2006.

CHEN, Weirong. *A grammar of Southern Min: The Hui'an dialect*. Berlim: Walter de Gruyter GmbH & Co KG, 2020.

COPIN, François. *Grammaire wayampi*. Tese de Doutorado. Univeristé Paris 7 Denis Diderot, 2012.

CREISSELS, Denis. A sketch of Mandinka. In: LÜPKE, Friederike. *The Oxford Guide to the Atlantic Languages of West Africa*. Oxford: Oxford University Press, 2024.

CROFT, William. Lexical and grammatical meaning. In: BOOIJ, Gert; LEHMANN, Christian; MUGDAN, Joachim. *Morphology: A handbook on inflection and word formation*. Berlin/New York: Mouton de Gruyter, 2000.

CRUZ, Aline. *Fonologia e gramática do Nheengatú*. 2011.

DAHL, Östen. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins, 2004.

DAVID, Anne. *Descriptive grammar of Pashto and its dialects*. Berlim: Walter de Gruyter, 2013.

DONOHUE, Mark. *A grammar of Tukang Besi*. Berlim: Walter de Gruyter GmbH & Co KG, 2011.

DRYER, M. S.; HASPELMATH, M. (eds.). *The World Atlas of Language Structures* Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. 2013. (Available online at http://wals.info, Accessed on 2021-03-10.)

ERSCHLER, David. Iron Ossetic. In: POLINSKY, Maria. *The Oxford handbook of languages of the Caucasus*. Oxford: Oxford University Press, 2020.

EPPS, Patience. *A grammar of Hup*. Berlim: Walter de Gruyter GmbH & Co KG, 2008.

EPPS, Patience; OBERT, Karolin. Marcación de número en Nadëb. *Visitas al Patio*, v. 15, n. 2, p. 255-273, 2021.

GASPARINI, Fabio. *The Baṭḥari language of Oman: Towards a descriptive grammar*. Unpublished Ph. D. dissertation, Università degli Studi di Napoli 'L'Orientale, 2018.

GHOSH, Arun. Santali. In: ANDERSON, Gregory. The Munda languages. Abingdon: Routledge, 2015.

GIVÓN, Talmy. *Syntax: an introduction*. Berlim: De Gruyter. v 1. 2008.

GRIMM, Nadine. *A grammar of Gyeli* (Volume 2). Language Science Press, 2021.

GUÉROIS, Rozenn. *A grammar of Cuwabo (Mozambique, Bantu P34)*. Tese de Doutorado. Université Lyon 2, 2015.

GUNNINK, Hilde. *A grammar of Fwe: A Bantu language of Zambia and Namibia*. Tese de Doutorado. Ghent University, 2018.

HASPELMATH, Martin. Inflection and derivation as traditional comparative concepts. *Linguistics*, v. 62, n. 1, p. 43-77, 2024.

HEATH, Jeffrey. *A Grammar of Jalkunan (Mande, Burkina Faso)*. Language Description Heritage Library. 2023.

HEATH, Jeffrey. *A Grammar of Kelenga (Hainyaxo), Bozo language of Mali*. Language Description Heritage Library. 2023.

HER, One-Soon. Variation of the VO construction in Chinese: a synchronic account. *Linguistics*, 34(4). 1996.

HOPPER, Paul J.; TRAUGOTT, Elizabeth Closs. *Grammaticalization*. Cambridge: Cambridge University Press, 2003.

HUALDE, José Ignacio. *Catalan*. Abingdon: Routledge, 2013.

JEON, Lisa et al. A Basic Sketch Grammar of Gĩkũyũ. *A Special Issue of Rice Working Papers in Linguistics* Volume VI, 2015.

KHACHATURYAN, Maria. *Grammaire de la langue mano (mandé-sud) dans une perspective typologique*. Tese de Doutorado. Institut National des Langues et Civilisations Orientales-INALCO PARIS-LANGUES O', 2014.

KIBORT, Anna. A typology of grammatical features. Grammatical Features. 2008. Disponível em: <http://www.grammaticalfeatures.net/inventory.html>. Acesso em: 18 de agosto de 2022.

KIBORT, Anna. Towards a typology of grammatical features. In: KIBORT, Anna. e CORBETT, Greville. *Features: Perspectives on a Key Notion in Linguistics*. Oxford: Oxford University Press. 64-106. 2010.

KUO, Pei-Jung. Transitivity and the BA construction. *Taiwan Journal of Linguistics*, v. 8, n. 1, 2010.

LI, Xuping. *A Grammar of Gan Chinese: The Yichun Language*. Berlim: Walter de Gruyter GmbH & Co KG, 2018.

LIDDICOAT, Anthony. *A grammar of the Norman French of the Channel Islands: The dialects of Jersey and Sark*. Berlim: Walter de Gruyter GmbH & Co KG, 2011.

LIMA SILVA, Wilson. *A descriptive grammar of Desano*. Tese de Doutorado. The University of Utah, 2012.

MA, Seng. *A descriptive grammar of Wa*. Chiang Mai: Payap University thesis, 2012.

MAHOOTIAN, Shahrzad; GEBHARDT, Lewis. *Persian (descriptive grammars)*. London: Routledge. v. 10, 1997.

MANFREDI, Stefano; PETROLLINO, Sara. Juba Arabic. *The survey of pidgin and creole languages*, v. 3, p. 54-65, 2013.

MATTOS, Eudes Barletta. *Estudo quantitativo da complexidade gramatical absoluta em diferentes línguas e suas implicações tipológicas*. Dissertação de Mestrado, UFBA. 2023.

MATTOS, Eudes Barletta; LAZZARINI-CYRINO, João Paulo. Classificação morfológica das línguas: da tipologia holística às definições circulares. *Revista Diadorim*, v. 23, n. 2, p. 288-312, 2021.

MATTOS, Eudes Barletta; LAZZARINI-CYRINO, João Paulo. A morfologia gramatical contextual e sua relação com a proporção da morfologia livre. *Veredas-Revista de Estudos Linguísticos*, v. 28, n. 2, 2024.

MAURER-CECCHINI, Philippe. *A grammar of Tuatschin: A Sursilvan Romansh dialect, v. 3*. Language Science Press, 2021.

MCWHORTER, John. *Language interrupted: Signs of non-native acquisition in standard language grammars***.** Oxford University Press on Demand, 2007.

MEL'ČUK , I. A. *Cours de morphologie générale. Introduction et première partie: Le mot*. v. 1. Montréal: Les Presses de l'Université de Montréal; Paris: CNRS Éditions. 1993.

MURSELL, Johannes. Object marking in Swahili is topic agreement. *Jezikoslovlje*, v. 19, n. 3., p. 427-455, 2018.

NAGARAJA, Keralapura. Standard Khasi. In: JENNY, Mathias; SIDWELL, Paul. *The handbook of Austroasiatic languages*, v. 2, p. 1145-1185, 2014.

NGAI, Sing Sing. *A grammar of Shaowu: A Sinitic language of Northwestern Fujian*. Berlim: Walter de Gruyter GmbH & Co KG, 2021.

NICHOLS, Johanna. Linguistic complexity: A comprehensive definition and survey. In: SAMPSON, Geoffrey; GIL, David; TRUDGILL, Peter. *Language complexity as an evolving variable*. Oxford University Press. p. 110-125. 2009.

PETERSON, John. *A grammar of Kharia: A South Munda language*. Leiden: Brill, 2010.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RAMIREZ, Henri. *A fala tukano dos Ye'pâ-masa*. Inspetoria Salesiana Missionária da Amazônia, 1997.

ROBERT, Stéphane. L'architecture fonctionnelle du langage et le modèle de la complexité en mosaïque. *Revue philosophique de la France et de l'étranger*. 143. 11. 10.3917/rphi.181.0011. 2018.

RÖßLER, Eva Maria. *Aspectos da gramática Achê*. Dissertação de Mestrado. Universidade Estadual de Campinas, 2008.

RÖßLER, Eva Maria. Inflectional morphology restructuring in Ache - Discussing grammatical change and language contact in Tupí-Guaraní subgroup - 1. *Boletim do Museu Paraense Emilio Goeldi: Ciencias Humanas*, v. 10, n. 2, 2015.

RÖßLER, Eva Maria. *Syntactic effects of inflectional morphology restructuring in Aché: on language change and language contact in Tupí-Guaraní subgroup-1*. Tese de Doutorado. Universidade Estadual de Campinas, 2018.

SANDMAN, Erika. *A grammar of Wutun*. Helsinki: University of Helsinki, 2016.

SANTIAGO, Huda da Silva. *A escrita por mãos inábeis: uma proposta de caracterização*. Tese de Doutorado. Universidade Federal da Bahia, 2019.

SAPIR, Edward. *An introduction to the study of speech*. Nova York: Harcourt, Brace, 1921.
SCHLEGEL, August. W. *Observations sur la langue et la litterature provençales*. Paris: Librairie grecque-latine-allemande, 1818.
SCHULZE, Wolfgang. *Northern Talysh*, 2000.
SCHWEGLER, Armin. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlim: Walter de Gruyter, 1990.
SEIDL, Amanda; DIMITRIADIS, Alexis. The discourse function of object marking in Swahili. *cls*, v. 33, p. 17-19, 1997.
SEKI, Lucy. *Gramática do kamaiurá: Língua tupí-guaraní do alto Xingu*. 2000.
SMITH, Carlota; ERBAUGH, Mary. Temporal interpretation in Mandarin Chinese. In*: Text, Time, and Context: Selected Papers of Carlota S. Smith*. Dordrecht: Springer Netherlands, 2009.
SOLANO, Eliete Bararuá. *Descrição gramatical da língua Araweté***.** 2010.
STILO, Donald. A Personal Narrative in Northern Talyshi with Grammatical Commentaries and Notes on the Areal Distribution of Grammatical Features. *Tehlikedeki Diller Dergisi*, v. 13, n. 22, 2023.
STUMP, Gregory T. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press, 2001.
TAULI, Valter. Morphological analysis and synthesis. *Acta Linguistica*, v. 5, n. 1, p. 80-85, 1945.
THOMPSON, Sandra A. Transitivity and some problems with the BA construction in Mandarin Chinese. *Journal of Chinese Linguistics*, v. 1, n. 2, p. 208-221, 1973.
TRUDGILL, Peter. Social structure, language contact and language change. In: WODAK, Ruth; JOHNSTONE, Barbara; KERSWILL, Paul. *The SAGE handbook of sociolinguistics*. Nova York: Sage Publications, 2011.
VITALE, Anthony. *Swahili syntax*. Berlim: Walter de Gruyter GmbH & Co KG, 2019.
VYDRIN, Valentin. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gwɛɛtaa. In: PALANCAR, Enrique L.; LÉONARD, Jean Léo. *Tone and Inflection: New facts and new* perspectives. Berlim: Walter de Gruyter GmbH & Co KG. 2016.
VYDRIN, Valentin. Featural foot in Bambara. *Journal of African languages and linguistics*, v. 41, n. 2, p. 265-300, 2020.

WEIR, Evelyn Mary Helen. *A negação e outros tópicos da gramática Nadeb*. Tese de Doutorado, 1984.
ZWICKY, Arnold M. How to describe inflection. In: *Annual Meeting of the Berkeley Linguistics Society*. 1985. p. 372-386.