



**UNIVERSIDADE FEDERAL DA BAHIA**  
**INSTITUTO DE LETRAS**  
**PROGRAMA DE PÓS GRADUAÇÃO EM LÍNGUA E CULTURA**

**EUDES BARLETTA MATTOS**

**ESTUDO QUANTITATIVO DA COMPLEXIDADE  
GRAMATICAL ABSOLUTA EM DIFERENTES LÍNGUAS E  
SUAS IMPLICAÇÕES TIPOLÓGICAS**

Salvador  
2023

**EUDES BARLETTA MATTOS**

**ESTUDO TIPOLOGICO QUANTITATIVO DA  
COMPLEXIDADE GRAMATICAL ABSOLUTA EM  
DIFERENTES LÍNGUAS**

Dissertação apresentada ao Curso de Mestrado em Letras,  
do Programa de Pós-Graduação em Língua e Cultura da  
Universidade Federal da Bahia – UFBA, para a obtenção  
do título de mestre.

Orientador: Prof. Dr. João Paulo Lazzarini Cyrino

Salvador  
2023

Mattos, Eudes Barletta

Estudo quantitativo da complexidade gramatical absoluta em diferentes línguas e suas implicações tipológicas / Eudes Barletta Mattos. -- Salvador, 2023.  
119 f. : il

Orientador: João Paulo Lazzarini Cyrino.

Dissertação (Mestrado - Língua e Cultura) --  
Universidade Federal da Bahia, Instituto de Letras,  
2023.

1. Morfossintaxe. 2. Tipologia linguística. 3. Complexidade gramatical. I. Lazzarini Cyrino, João Paulo. II. Título.

BANCA EXAMINADORA

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Fernanda de Oliveira Cerqueira

---

Prof. Dr. Paulo Chagas de Souza

---

Prof. Dr. João Paulo Lazzarini Cyrino (orientador)

## Agradecimentos

Meus agradecimentos ficam para os encontros, avisados ou desavisados, que de alguma forma contribuíram para a feitura deste trabalho e para minha perseverança na vida acadêmica. A meu orientador, que tornou nossa pesquisa possível: *merci le poto!* Mãe, pai, irmãs, obrigado pelo amor. Amigos, conhecidos, orientadores prévios e professores, servidores e funcionários da UFBA, povo do Brasil, pesquisadores citados aqui ou não, comunidades e nações falantes de cada uma das línguas com que trabalhei... sou muito grato, de verdade, não só por protocolo.

## RESUMO

Medir a complexidade de uma língua é uma tarefa trabalhosa; inúmeros problemas, como o da comparabilidade (CROFT, 2003), comuns a estudos tipológicos, apresentam-se no caminho, juntamente com a questão axial de se definir complexidade – que é necessário antes de mais nada para poder mensurá-la (MIESTAMO, 2008). Ademais, estudos comparativos na linguística não raro lançam mão de um aparato terminológico proveniente da classificação holística do século XIX, subsumindo línguas inteiras em perfis *isolantes*, *aglutinantes*, *fusionais*, entre outros. Soma-se a estes desafios o fato de a maioria das análises balizar-se em uma unidade translinguisticamente pouco consistente, a palavra (MATTOS; LAZZARINI-CYRINO, 2020). Tendo esta problemática em vista, o presente estudo busca, utilizando o morfema como unidade de análise, inventariar os morfemas gramaticais mais frequentes de 19 línguas amostradas pelo mundo, caracterizando-os com base em critérios semânticos que evitem operar o conceito de palavra (CROFT, 2000), e quantificando-os de acordo com medidas de complexidade absoluta, isto é, não referente a sua dificuldade, mas sim ao número de partes do sistema. Para tanto, utilizamos a taxonomia de complexidade de McWhorter (2007), a noção de verbosidade de Dahl (2004) e as complexificações de Trudgill (2011), classificando morfema a morfema com base em cinco parâmetros – o primeiro dizendo respeito a se o morfema é glosado como preso ou livre, e os quatro demais propriamente referentes à complexidade: se veiculador de flexão contextual, seu grau de verbosidade, se irregular ou com elaboração estrutural, e se poliexponente. Frequências relativas para cada parâmetro foram obtidas, e os resultados, processados no *software R*, foram comparados. Descobrimos uma forte correlação entre maior proporção de morfologia gramatical livre e menores escores nos quatro parâmetros de complexidade, bem como a existência de línguas dramaticamente *simplificadas* em comparação com as demais seguindo nossa quantificação; hipotetizamos, de acordo com McWhorter (2016), que isso se dá por eventos de contato intenso em suas histórias. Além disso, percebemos similaridades no comportamento das línguas das Américas, e propomos abandonar a terminologia da classificação holística, bem como a caracterização negativa de línguas menos complexas.

**Palavras-chave:** Complexidade absoluta, Tipologia linguística, Morfossintaxe

## ABSTRACT

Measuring the complexity of a language is a strenuous task; one runs into several problems such as the issue of comparability (CROFT, 2003), commonplace in typological studies, as well as the central issue of defining complexity, obviously necessary before we measure it (MIESTAMO, 2008). Besides that, comparative studies in linguistics often use terminology from 19th-century holistic classification of languages, classifying them wholly as *isolating*, *agglutinative*, *fusional* and so on. Also, most analyses are based on a cross-linguistically inconsistent unit, the word (MATTO; LAZZARINI-CYRINO, 2020). With all this in mind, the present study, using morphemes as analysis units, lists the most frequent grammatical morphemes of 19 languages sampled worldwide, characterizing them by semantic criteria that don't use the notion of word (CROFT, 2000), and quantifying them according to absolute complexity measures, that is, not referring to the language's perceived difficulty, but to the number of parts of the system. In order to do so, we used McWhorter's taxonomy of complexity (2007), Dahl's notion of verbosity (2004) and Trudgill's complexifications (2011), classifying each morpheme based on five parameters – the first one, if a given morpheme is glossed as free or bound, and the four subsequent ones properly being about complexity: if a given morpheme realizes contextual inflection, its degree of verbosity, if it presents irregularities and/or structural elaboration, and if it is polyexponent. Relative frequencies for each parameter were taken, and the results were processed using the software R and compared. We found a strong correlation between a higher proportion of unbound grammatical morphology and smaller scores on the four used complexity parameters, as well as the existence of dramatically *simplified* languages compared to the rest; it's hypothesized, following McWhorter (2016), that this is due to intense contact events in their past histories. Moreover, we noticed behavior similarities between American languages, and we propose relinquishing holistic classification terminology, as well as not portraying less complex languages negatively.

**Keywords:** Absolute complexity, Linguistic typology, Morphosyntax

## LISTA DE ABREVIACOES

1	primeira pessoa	DU	dual
2	segunda pessoa	ERG	ergativo
3	terceira pessoa	EXCL	exclusivo
A	agente	F	feminino
ABL	ablativo	GEN	genitivo
ABS	absolutivo	INAN	inanimado
ACC	acusativo	IND	indicativo
AGR	concordância	INESS	inessivo
ANIM	animado	INF	infinitivo
ART	artigo	INSTR	instrumental
ASP	aspecto	IPD	impeditivo
AUX	auxiliar	IPFV	imperfectivo
CL	classificador	IRR	irrealis
CF	característica constante	LOGO	pronome logofórico
CFIN	final de constituinte	LOC	locativo
COM	comitativo	M	masculino
CONNEG	conegativo	MIN	forma mínima
COP	cópula	N	neutro
CORE	núcleo não-nominativo	NEG	negação
CTMP	cotemporal	NMLZ	nominalizador
DAT	dativo	NOM	nominativo
DECL	declarativo	NPST	não-passado
DEF	definido	NS	não-singular
DEM	demonstrativo	NSUJ	não-sujeito
DIR	direcional	O	objeto
DISC.CONJ	conjunção discursiva	OBL	oblíquo
DIST	distal	PFV	perfectivo
DJ	disjuntivo	PL	plural
DO	objeto direto	POSS	possessivo

PRED	predicativo	RECIP	recíproco
PROX	proximal	S	sujeito
PRS	presente	SBJ	subjuntivo
PST	passado	SFOC	foco de sujeito
PTCP	particípio	SG	singular
QUE	interrogativo	TEL	télico
R	realis	VENT	ventivo
RED	reduplicação	VP	prefixo vocálico
REL	relacional		
REP	relatado		
RES	resultativo		

## LISTA DE FIGURAS

Figura 1. Distribuição geográfica das línguas utilizadas na dissertação	10
Figura 2. A plataforma òcun, dados do sumério	11
Figura 3. Fragmento do documento de texto com sentenças segmentadas e glosadas	12
Figura 4. Valores utilizados na classificação em quatro dimensões	25
Figura 5. Proporção de morfemas gramaticais livres nas diferentes línguas	27
Figura 6. Relação entre proporção de morfemas livres e grau de elaboração estrutural	39
Figura 7. Relação entre proporção de morfemas livres e proporção de morfemas veiculando mais de uma categoria gramatical	47
Figura 8. Comparação entre flexão contextual <i>sensu stricto</i> e <i>sensu lato</i>	51
Figura 9. Proporção de morfemas veiculadores de informação gramatical determinada sintaticamente, sem considerar palavra como item de análise	52
Figura 10. Proporções de informação gramatical determinada sintaticamente e morfemas livres	54
Figura 11. Relação entre proporção de morfemas livres e proporção de morfemas veiculadores de flexão contextual	55
Figura 12. Proporção de morfemas presentes em construções dos três graus de complexidade fenogramatical-verbosidade nas línguas	64
Figura 13. Relação entre proporção de morfemas livres e média de complexidade fenogramatical-verbosidade nas línguas	65
Figura 14. Correlação entre complexidade fenogramatical-verbosidade e proporção de morfemas flexionais contextuais	67
Figura 15. Agrupamento em dendrograma por <i>clustering</i> hierárquico de acordo com a proporção de morfemas gramaticais livres	69
Figura 16. Agrupamento em dendrograma por <i>clustering</i> hierárquico de acordo com os quatro demais parâmetros	70
Figura 17. Agrupamento em dendrograma por <i>clustering</i> hierárquico de acordo com os quatro parâmetros de complexidade e a proporção de morfologia presa	71
Figura 18. Mapa da região ocupada pela família austronésia	80

## LISTA DE TABELAS

Tabela 1. Línguas utilizadas na dissertação	9
Tabela 2. Origem dos dados das línguas utilizadas	10
Tabela 3. Os sete primeiros morfemas do zapoteca	12
Tabela 4. Seis dos morfemas mais frequentes do araweté	26
Tabela 5. Os cinco morfemas com menores valores de frequência dentre os 20% mais frequentes do araweté	26
Tabela 6. Classificação de morfemas em livres <i>versus</i> presos e elaboração estrutural/irregularidade; exemplos do daai chin	38
Tabela 7. Classificação dos morfemas por poliexponência; exemplos do domari	45
Tabela 8. Classificação de morfemas por flexão contextual <i>sensu stricto</i> ou <i>sensu lato</i>	50
Tabela 9. Gradação de complexidade fenogramatical	57
Tabela 10. Análise por graus e categorias gramaticais em tabela; exemplos do cuwabo	63
Tabela 11. Morfemas aspectuais do dâw e suas origens em itens lexicais	79

## SUMÁRIO

<b>Introdução – A complexidade das línguas</b>	<b>1</b>
1.1. A classificação holística das línguas	3
1.2. O problema da palavra	6
1.3. A proposta desta dissertação	8
1.4. A estrutura desta dissertação	14
<b>CAPÍTULO 1. Da análise morfêmica</b>	<b>16</b>
2.1. Morfemas gramaticais <i>versus</i> morfemas lexicais	18
2.1.1. As quatro dimensões de Croft	21
2.2. A triagem de morfemas	25
2.3. Proporção de morfemas livres	27
<b>CAPÍTULO 2. Classificações de complexidade</b>	<b>29</b>
3.1. A taxonomia da complexidade de McWhorter	32
3.1.1. Sobrespecificação	33
3.1.2. Elaboração estrutural	35
3.1.3. Irregularidade	37
3.1.4. Medição da elaboração estrutural/irregularidade	37
3.2. A verbosidade de Dahl	40
3.2.1. Tectogramática <i>versus</i> fenogramática	41
3.2.2. Complexidade fenogramatical e verbosidade	43
3.3. As complexificações de Trudgill	45
<b>CAPÍTULO 3. Medidas de sobrespecificação</b>	<b>48</b>
4.1. Flexão contextual	48
4.2. Quantificação da morfologia flexional contextual	50
4.3. Complexidade fenogramatical	57
4.3.1. Grau 1	58
4.3.2. Grau 2	59
4.3.3. Grau 3	61
4.3.4. A análise	63
<b>CAPÍTULO 4. Implicações</b>	<b>68</b>
5.1. Dendrogramas	68
5.1.1. Agrupamento por morfologia livre	68
5.1.2. Agrupamento pelos parâmetros de complexidade	69
5.1.3. Agrupamento pela união de todos os parâmetros	70
5.2. Relação entre morfologia presa e complexidade	72
5.3. Mudança por contato	74
5.4. Crioulos e perda da flexão contextual	76
5.5. As línguas GML	77
5.5.1. Dâw	77
5.5.2. Kéo	80
5.5.3. Mandarim	83

5.5.4. Rapa nui	86
5.5.5. Daai chin e koyra chiini	89
5.6. Línguas das Américas	91
5.6.1. Velho mundo <i>versus</i> áreas de colonização	92
5.6.2. O yupik	93
<b>Considerações finais</b>	<b>95</b>
<b>Referências</b>	<b>97</b>

三十輻，共一轂，當其無，有車之用。

埴埴以為器，當其無，有器之用。

鑿戶牖以為室，當其無，有室之用。

故有之以為利，無之以為用。

trinta raios perfazem o meão  
no imanifesto o uso do carro

barro moldado faz o jarro  
no imanifesto o uso do jarro

talham-se portas e janelas para a casa  
no imanifesto o uso da casa

portanto  
utilizando-se o manifesto o útil é o imanifesto

Dao De Jing - XI  
(trad. Mário Sproviero)

## Introdução – A complexidade das línguas

*Simples e complexo* são conceitos operados diuturnamente, sem muita reflexão, sobre uma infinidade de assuntos – inclusive as línguas. O falante leigo, mas não desprovido de intuição, não raro faz afirmações acerca da simplicidade ou complexidade de algo na língua, ou de uma língua em específico. A academia o faz igualmente – e nem sempre de forma tão esclarecida quanto gostaríamos de acreditar. Mas o que é a complexidade? O que torna uma língua mais ou menos complexa, e em relação ao quê? A linguística por muito tempo respondeu esses questionamentos sob pontos de vista repreensíveis, confusos ou questionáveis.

### 1. Persa antigo† (Irã: indoeuropeia)

taya	manā	kərtam	idā	utā
isso.que.N.SG	1SG.GEN/DAT	fazer.PTCP.N.SG	aqui	e

taya-mai	apataram	kərtam
isso.que.N.SG-1SG.GEN/DAT	longe	fazer.PTCP.N.SG

awa	wisam	wašna	auramazdāhā	akunav-am
isso	tudo	vontade.INSTR	Ormasde.GEN	fazer.IPFV-1SG

“O que fiz aqui, e o que fiz lá longe, tudo isso fiz pela vontade de Ormasde.” (HAIG, 2015)

### 2. Farsi (Irã: indoeuropeia)

ne-mi-dun-am	kodum	aks-o	in	bače	pāk
NEG-IPFV-saber-1SG	qual	imagem-o	esse	garoto	limpar

kard-e  
fazer-PST.3SG

“Não sei qual imagem esse garoto deletou.” (JASBI, 2020)

Ao se observarem as sentenças em 1. e 2., não é difícil perceber o maior grau de complexidade do aparato gramatical trazido na sentença do persa antigo. O que traria a perda de complexidade entre essa língua e seu descendente direto, exemplificado logo em seguida?

O sistema de casos gramaticais, três gêneros e concordância de número do persa antigo desapareceram do farsi, no qual o plural é vastamente inexistente ou opcional, não há marcação de caso – exceto um marcador de objeto que é também opcional na maioria das instâncias –, e o sistema de gênero inexistente (MCWHORTER, 2007; JASBI, 2020). Comparações do latim clássico com suas descendentes neolatinas, e descrições de outras línguas de similar antiguidade e pertencentes à mesma família, como o sânscrito e o avéstico, não tardaram a suscitar teorias de *decadência* e *degradação* de um passado ideal, com reflexos na complexidade das línguas (MATTOS; LAZZARINI-CYRINO, 2020).

Logo, porém, a existência de línguas de perfil diferente do das indoeuropeias deu azo a caracterizações racistas da inteligência de seus falantes, e associadas a *primitivismo*. Dominada pela produção científica europeia, herdeira de uma tradição gramatical construída tendo como objeto línguas clássicas da família indoeuropeia como o grego antigo e o latim, infundida com interpretações equivocadas do darwinismo e ideias de supremacia racial que culminaram em presunções acientíficas com ares de teoria da conspiração (SCHWEGLER, 1990; MATTOS; LAZZARINI-CYRINO, 2020), a linguística em seus primórdios acabou por eleger uma explicação *evolucionista*<sup>1</sup> para dar conta da complexidade das diferentes línguas do mundo. Relíquias desse período racista persistem até hoje no fazer científico acerca da linguagem, e demonstram sua persistência toda vez que se usam os termos *língua isolante*, *aglutinante*, *fusional*, *analítica*, *sintética*, provenientes da classificação de línguas inteiras em tipos morfológicos – e hierarquicamente estabelecidos (*ibid*). Tão tardiamente quanto este século, ainda se veem em publicações das mais diversas áreas da linguística (DIXON, 1997; BAIRD, 2002; LUO, 2004; MCWHORTER, 2007, 2011; GIL, 2008; MIESTAMO, 2008; FENK-OCZLON; FENK, 2008; RIDDLE, 2008; SYBESMA, 2008; ENFIELD, 2009; SO-HARTMANN, 2009; JAGERSMA, 2010; MIYAOKA, 2012; DELANCEY, 2013; ARCODIA, 2015; CARVALHO, 2016; DELMER, 2016; KIEVIET, 2017; ANSALDO *et al.*, 2018; MASSAM, 2020; KORNAL, 2020; OBERT, 2019, 2020)<sup>2</sup>.

O estudo da complexidade na linguística comparativista, portanto, nasce atrelado à classificação em tipos morfológicos, tendo sofrido com entraves a seu desenvolvimento oriundos das crenças dos linguistas europeus do século XIX. Contudo, e como reação compreensível a tais crenças advindas de uma ideologia eurocêntrica, racista e xenofóbica, surgiu a *hipótese da compensação* – que atualmente representa o consenso na linguística

---

<sup>1</sup> Aqui, na acepção menos feliz e mais equivocada do termo, a de gradação entre formas mais evoluídas e formas mais atrasadas.

<sup>2</sup> E mais uma infinidade de trabalhos, ao ponto que seria mais fácil citar os que explicitamente evitam tal nomenclatura.

(MCWHORTER, 2007; GIL, 2008). Essa hipótese representa o ponto de vista chamado de *igualitarismo linguístico*, ou da *equi-complexidade*, parte dos esforços de se desmontar o consenso prévio de que haveria línguas e culturas mais “avançadas” que outras; tal empreitada contrária à noção de evolução como superioridade de certas culturas e línguas foi iniciada por Franz Boas, e inevitavelmente apresentaria reflexos na linguística (MCWHORTER, 2007).

A hipótese da compensação sugere que qualquer perda de complexidade em um dado nível de uma língua é prontamente compensada por um aumento de complexidade em outro nível. Mais frequentemente, é utilizada para se fazerem afirmações acerca da perda de complexidade morfológica numa língua; se a morfologia é mais “empobrecida”, o corolário lógico é de que algum outro domínio deve ter se tornado mais complexo, como a sintaxe ou a semântica. Casos em que se invoca a hipótese da compensação são a evolução das línguas românicas a partir do latim (e a subsequente perda de complexidade morfológica representada pela erosão do sistema de casos e sua substituição pelo uso de preposições), e comparações de línguas ditas “mais sintéticas”, como as línguas clássicas da Europa, com línguas “mais analíticas”, como as do Sudeste Asiático ou África Ocidental.

A hipótese da compensação é bem sintetizada por Hockett (1958):

[uma m]edição objetiva é difícil, mas impressionisticamente parece que a complexidade gramatical total de qualquer língua, contando-se tanto morfologia quanto sintaxe, é aproximadamente a mesma que a de qualquer outra. Não é de se surpreender, uma vez que todas as línguas têm trabalhos mais ou menos igualmente complexos a fazer, e o que não é feito morfológicamente deve ser feito sintaticamente. Fox, com uma morfologia mais complexa que o inglês, então deve ter uma sintaxe algo mais simples; e é esse o caso. Então uma escala para a comparação dos sistemas gramaticais de diferentes línguas é a do grau médio de complexidade morfológica – carregando consigo uma implicação inversa em relação ao grau de complexidade sintática. (pp. 180-181, tradução nossa)

Por mais louvável que tenha sido seu surgimento, e sem esquecer a necessidade de descartar conceitos acientíficos e oriundos de uma episteme declaradamente supremacista racial, a hipótese da compensação permanece sem demonstrações inequívocas e quantitativas.

### 1.1. A classificação holística das línguas

Os primeiros esforços de comparação entre as línguas validados como fazer científico foram lançados por Adam Smith (1759), mas se firmariam no contexto do século XIX e as classificações das línguas em *tipos morfológicos* por August von Schlegel (1808) – cujos

nomes são bastante conhecidos e utilizados em linguística até o presente momento: *isolantes*, *aglutinantes* e *fusionais*, estas últimas subdivididas em *analíticas* e *sintéticas*, com forte ênfase no componente flexional da morfologia. Para o autor alemão, uma morfologia flexional robusta seria sinal de um suposto desenvolvimento não só da língua, mas de sua comunidade de falantes (SCHWEGLER, 1990). As línguas *isolantes* são descritas por ele como *desprovidas de gramática*, exemplificadas por línguas como o chinês; segundo o autor, seriam *estéreis* e *primitivas*. As línguas aglutinantes e fusionais, que realizariam categorias gramaticais por meio de afixação e/ou mudanças nas raízes, seriam mais *avançadas*, as últimas sendo o ápice do desenvolvimento intelectual humano. Schlegel ainda aduz que as línguas fusionais, caso disponham de literatura que lhes sirva de modelo, bem como de um sistema de instrução formal, tenderiam a permanecer sob o mesmo perfil morfológico; no entanto, sem este apoio, inevitavelmente passariam a um perfil mais analítico, degenerado, por submetidas à instabilidade inerente às coisas humanas (SCHLEGEL, 1818).

É importante delinear o contexto de proposição das ideias de August von Schlegel, a Alemanha do século XIX. Por conta de uma acentuada fase de romantismo e unificação nacional, surgem nesse território uma série de ideias comprometidas por eurocentrismo e supremacia racial, tendo como eixo epistemológico o conceito absurdamente essencialista do *Volksgeist*, o espírito de um povo. O espírito – ou gênio – de um povo seria concomitantemente a motriz de sua história, bem como a manifestação de seu caráter em todas as suas produções e atividades, costumes e língua. O estudo comparativista das línguas, portanto, grandemente fomentado pela divulgação do sânscrito nos círculos acadêmicos europeus, não passaria incólume por esta época, imiscuindo-se com teorias pseudocientíficas de pureza e hierarquia raciais; apresentam influências da concepção de *Volksgeist* os estudos e monografias de pensadores como Wilhelm von Humboldt<sup>3</sup>, Jakob e Wilhelm Grimm, Franz Bopp, e o próprio August e seu irmão Friedrich von Schlegel (MATTOS; LAZZARINI-CYRINO, 2021).

Tem-se na proposta de Schlegel, então, uma classificação de cunho hierarquizante e fundada nas presunções de que a organização de uma língua refletiria diretamente a inteligência de seus falantes; de que línguas inteiras se subsumiriam aos perfis descritos; de que análise e síntese são mutuamente excludentes, acontecendo ou uma ou outra num só

---

<sup>3</sup> Note-se que Humboldt, mesmo contrário à classificação de Schlegel – para o primeiro, classificações não-genealógicas das línguas seriam impossíveis ou não-operacionais –, ainda assim associa a estrutura de uma língua ao *Volksgeist*, ao espírito do povo que a fala. Para um debate minucioso sobre a evolução dos conceitos de analiticidade/sinteticidade, remeto à leitura de Schwegler (1990).

momento de mudança de uma dada língua; e que o movimento análise → síntese é não-cíclico, envolvendo *degradação* de um *estado ideal* (SCHWEGLER, 1990).

A despeito de modificações nesta argumentação teórica inicial, de cunho flagrantemente acientífico, a classificação de Schlegel manteve-se relevante para a linguística ao longo do século XX, mormente por ter sido elaborada por nomes como Edward Sapir (1921) e Joseph Greenberg (1960). Deste percurso que surge a disciplina da tipologia linguística, e o estudo comparativista da morfologia das línguas sempre ocupou papel de destaque em suas discussões. Conforme dito, tais estudos tradicionalmente são acompanhados do uso da velha terminologia da tipologia holística – embora frequentemente sob um enfoque cíclico, tentativamente abandonadas as ideias de *Volksgeist* e associação de inteligência dos falantes com organização da língua:

À medida que as línguas mudam com o tempo, elas tendem – *grosso modo* – a se mover ao redor de um círculo tipológico: isolante a aglutinante a fusional, de volta a isolante, e assim por diante. Se colocarmos o tipo isolante na posição de quatro horas, aglutinante em oito horas, e fusional em doze horas, ao redor de um relógio, é possível descrever movimentos recentes em diversas famílias de línguas. (DIXON, 1997, pp. 41-42, tradução nossa)

No entanto, não é apenas por seus vícios decorrentes de eurocentrismo e racismo que a abordagem holística é condenável. Numa mesma língua, diferentes estratégias convivem, de modo que certas construções são expressas de maneira *isolante* – como a posse em português, com o uso da preposição *de* entre possuído e possuidor – e não-isolante, como os tempos do indicativo nessa mesma língua. A proporção de construções de um ou outro tipo em cada língua é um dado que não é aferível com o uso da terminologia da tipologia holística.

Palavras, definidas fonologicamente como unidades mínimas entre pausas em potencial, ocupam uma posição interessante entre morfemas e sintagmas, em parte porque algumas línguas as fazem muito pequenas, enquanto outras as fazem muito grandes. Um bom exemplo do primeiro tipo (conhecido como línguas isolantes ou analíticas) é o vietnamita, no qual palavras são tipicamente compostas por um único morfema, e sintagmas são tipicamente compostos por muitas palavras. No outro extremo, sintético ou polissintético, encontramos línguas como finlandês e náuatle, as quais têm muitos morfemas por palavra e conseqüentemente apenas poucas (em casos limitantes, apenas uma) palavras por sintagma, ou até mesmo por sentença. (KORNAI, 2020, p. 130, tradução nossa)

Afirmações como a acima usam termos vagos, a exemplo de *tipicamente compostos*, para caracterizar as línguas, sem dizer em quais construções e/ou contextos, e tampouco sua proporção. Além disso, a terminologia da tipologia holística é indissociável de um outro

conceito de pobre validade translinguística e extensiva problemática em sua delimitação e aplicabilidade, embora à primeira vista pareça natural: a palavra.

## 1.2. O problema da palavra

Uma das unidades cuja validade tem sido fortemente questionada é justamente a *palavra*, mais afeita às línguas indoeuropeias ocidentais – e ainda assim, com inúmeras ressalvas; não raro, o conceito se revela problemático e de delimitação imprecisa à maioria das línguas (DIXON; AIKHENVALD, 2003; HASPELMATH, 2011; MATTOS; LAZZARINI-CYRINO, 2020). Como item de comparação translinguística, a palavra e todas as noções que dela derivem ou nela se baseiem acabam se revelando problemáticas e inconsistentes. No entanto, seu uso é tão naturalizado entre nós que propostas em outra direção não raro são vistas com espanto:

Uma versão da análise estruturalista (Harris, 1946) concebe sentenças como sendo diretamente compostas por morfemas, mas isso é algo como uma aberração, em que tanto a gramática clássica (iniciada com Pānini por volta de 500 antes da era comum) e a gramática gerativa contemporânea [...] admitem um nível intermediário de palavras compostas por morfemas e servindo de blocos de construção sentencial. (KORNAI, 2020, p. 130, tradução nossa)

Para se estudar a tipologia de fenômenos envolvendo a flexão morfológica, porém, é proveitoso utilizar outra unidade básica de análise que não a palavra, de maneira a se realizarem comparações válidas translinguisticamente. Como consequência, definições que utilizem a distinção entre afixos e formas livres, por baseadas na palavra, não permanecem operacionais. No estudo de processos de flexão, por exemplo, há acepções que considerem *flexão* como algo além da afixação; sob este posicionamento, Spencer (2000) afirmou:

É importante entender que a morfologia flexional é apenas uma maneira com a qual uma língua realiza suas características f(uncionais). Eu assumirei que cada língua tem um vocabulário de características-f, as quais têm dois papéis principais: (i) regular a estrutura sintática (ex.: concordância, regência) e (ii) expressar certos significados “funcionais”, como tempo, definitude, plural, etc. No contexto da flexão, estes dois papéis correspondem respectivamente a flexão contextual e flexão inerente<sup>4</sup> conforme Booij (...) **Características-f podem ser expressas numa grande variedade de maneiras além da flexão do núcleo, incluindo clíticos, palavras auxiliares, ordem das palavras e entonação, ou uma combinação destes.** (p. 337, tradução e grifos nossos)

---

<sup>4</sup> Spencer alude a uma divisão da morfologia flexional, proposta por Geert Booij em 1993 e subsequentemente elaborada pelo próprio e por linguistas de vários campos, que será apresentada no capítulo 5, uma vez que usada nesta dissertação como medida de complexidade.

Tendo isto em vista, há de se considerar a contribuição de Zellig Harris, que já em 1946 propôs o morfema como item de análise; muito embora a preocupação do autor fossem as análises morfossintáticas, de fato o morfema, o “mais simples [item] observável”, de fácil visualização e identificação, é consistente translinguisticamente<sup>5</sup> (HARRIS, 1946). As gramáticas descritivas – e a vastíssima maioria das publicações em linguística – trazem seus dados e exemplos segmentados em morfemas, devidamente glosados e traduzidos. Apesar de remontar à primeira metade do século passado, tal posicionamento permanece em consonância com problematizações bastante atuais em teoria e tipologia linguística acerca da validade da palavra; já é bem abordado o fato de não haver uma definição incontroversa de palavra que possa servir de item de análise através das diferentes línguas do globo sem incorrer em vieses e soluções *ad hoc* (DIXON; AIKHENVALD, 2003; HASPELMATH, 2011; MATTOS; LAZZARINI-CYRINO, 2020). Já em 1990, por exemplo, Armin Schwegler afirma, analisando o caminho percorrido pela terminologia legada da classificação holística:

Ficará claro que de maneira a se manter a significância dos parâmetros *sintético* e *analítico*:

- (1) seu uso deve se limitar a unidades de fala em vez de línguas inteiras,
- (2) a “palavra” não deve ser encarada como essencial aos conceitos, e
- (3) eles devem ser entendidos não como absolutos quantificáveis, mas como a medida grosseira da interdependência morfêmica geral das unidades de fala.

Sugerir-se-á que esta medida é melhor alcançada **levando-se em conta critérios semânticos e morfossintáticos, assim como fonológicos**. (1990, pp. 15-16, grifos e tradução nossos)

Não obstante o curso trilhado pela tipologia linguística, de seus desdobramentos mais recentes e da necessidade de se discutir a classificação das línguas – que não raro ecoa a já defunta tipologia holística em esforços de enquadrar toda uma língua num dado perfil (SCHWEGLER, 1990; BICKEL, 2007) –, noções como *língua isolante*, *aglutinante*, *analítica* ou *sintética* ainda não foram totalmente superadas na linguística, sendo sempre relevante realizar pesquisas que não se valham de tais abordagens. Além disto, a despeito de importantes contribuições na definição do que é *complexidade* (DAHL, 2004; MCWHORTER, 2007; TRUDGILL, 2011), estudos que a avaliem utilizando ferramentas estatísticas e sem recorrer às noções de *palavra* ou *afixação*, permanecem sem ser feitos – em especial ao se abordarem as línguas não com o intuito de classificá-las em tipos, mas de quantificar fenômenos ou características nelas de forma estatisticamente significativa. Tais

---

<sup>5</sup> O conceito de morfema também apresenta suas controvérsias; para uma breve discussão, ver o capítulo 2.

esforços têm o potencial de fornecer resultados e generalizações menos comprometidos com unidades controversas advindas da tradição gramatical.

### 1.3. A proposta desta dissertação

Antes de mais nada, a presente dissertação tenciona ser um estudo que não use nem a terminologia da classificação holística, nem a palavra como item básico de análise, com o fito de construir conhecimento acerca da linguagem de maneira empirista e aplicável ao maior número possível de línguas. Nossa proposta é de investigar a associação entre a proporção de morfemas gramaticais livres de uma língua e sua complexidade absoluta – esta última medida como as frequências de morfemas gramaticais veiculadores de flexão inerente *sensu* Booij (1993, 1996), de morfemas com algum grau de irregularidade e elaboração estrutural *sensu* McWhorter (2007), de morfemas veiculando mais de uma informação gramatical (poliexponentes), e de morfemas com alto grau de complexidade fenogramatical-verbosidade *sensu* Dahl (2004). Tal estudo se propõe a oferecer um ponto de vista frequentista e balizado estatisticamente em confronto à hipótese da compensação, que é o consenso atualmente vigente na linguística acerca da complexidade das línguas, sob a perspectiva da tipologia linguística.

Nesta dissertação, usamos gramáticas descritivas como fonte primária dos dados<sup>6</sup>, permitindo o acesso a sentenças em línguas minoritárias por toda a extensão do globo. Além disso, gramáticas descritivas também se prestam muito bem ao papel de fornecer dados segmentados morfológicamente – e devidamente glosados –, por conta da própria natureza da apresentação das sentenças em tais obras: exemplos ao longo do livro e histórias anexas muito frequentemente são mostrados em três linhas, como é corrente em trabalhos de linguística, a primeira com o dado original da língua segmentado em morfemas; a segunda, com a glosa morfema-a-morfema; e a terceira com a tradução da sentença, permitindo uma avaliação do significado geral dos elementos ali trazidos, individualmente e em combinação.

Desse modo, foram utilizadas gramáticas descritivas de 19 línguas, amostradas de maneira a compreender o máximo de diversidade de famílias e origem geográfica possível. Na tabela 1, veem-se nome, código ISO 639-3, localidade e afiliação genética das línguas utilizadas; a numeração apresentada serve como referência para o mapa na figura 1, enquanto

---

<sup>6</sup> À exceção dos dados do mandarim, obtidos de exemplos de sentenças reais num dicionário *online* e glosados por mim mesmo.

o código ISO 639-3 servirá de rótulo para cada uma das línguas nos gráficos encontrados ao longo da dissertação.

<b>Número</b>	<b>Língua</b>	<b>ISO 639-3</b>	<b>Local: afiliação</b>	<b>Número aproximado de falantes</b>
1	Araweté	awt	Brasil: tupiana	200
2	Cuwabo	chw	Moçambique: bantu	800.000
3	Daai chin	dao	Myanmar: sino-tibetana	45.000
4	Dâw	kwa	Brasil: nadahup	menos de 100
5	Domari	rmt	Oriente Médio: indoeuropeia	281.670
6	Kanoê	kxo	Brasil: isolada	3
7	Khwarshi	khv	Cáucaso: nakh-daguestaniana	1.700
8	Koyra chiini	khq	Mali: nilo-saariana	200.000
9	Kéo	xxk	Ilha de Flores: austronésia	40.000
10	Limilngan†	lmc	Austrália: línguas da região de Darwin	extinta
11	Mandarim	cmn	China: sino-tibetana	921.5 milhões
12	Mapudungun	arn	Chile: araucana	260.000
13	Moloko	mlw	Camarões: afroasiática	8.500
14	Pite Saami	sje	Noruega/Suécia: urálica	20
15	Rapa Nui	rap	Ilha de Páscoa: austronésia	1.000
16	Sumério†	sux	Mesopotâmia: isolada	extinta
17	Yakkha	ybh	Nepal: sino-tibetana	20.000
18	Yupik do Alasca central	esu	Alasca: esquimó-aleuta	19.750
19	Zapoteca de San Bartolomé Z.	zpq	México: oto-mangueana	1.400

Tabela 1. Línguas utilizadas na dissertação.



Figura 1. Distribuição geográfica das línguas utilizadas na dissertação.

Igualmente, a tabela 2 traz o número de frases utilizadas como amostra para cada uma das línguas, bem como o número de morfemas obtidos a partir dessas frases. Por fim, a tabela 2 mostra uma última cifra, o percentual dos 20% mais frequentes entre os morfemas, usados para análise.

<b>língua</b>	<b>fonte</b>	<b>número de frases retiradas</b>	<b>número de morfemas</b>	<b>20%</b>
Araweté	SOLANO, E. J. B., 2009	563	420	84
Cuwabo	GUÉROIS, R., 2015	200	595	119
Daai chin	SO-HARTMANN, H., 2009	846	1094	218,8
Dâw	CARVALHO, M. O. P., 2016	262	253	50,6
Domari	MATRAS, Y., 2012	201	437	87,4
Kanoê	BACELAR, L. N., 2004	934	543	108,6
Khwarshi	BAIRD, L., 2002	1149	1468	293,6
Koyra chiini	KHALILOVA, Z., 2009	200	337	67,4
Kéo	HEATH, J., 1999	211	397	79,4
Limilngan	HARVEY, M., 2001	201	295	59

Mandarim	LINE Dict Chinese-English	205	720	144
Mapudungun	SMEETS, I., 2008	200	430	86
Moloko	FRIESEN, D. <i>et al.</i> , 2017	200	299	59,8
Pite Saami	WILBUR, J., 2014	241	568	113,6
Rapa Nui	KIEVIET, P., 2017	221	488	97,6
Sumério	JAGERSMA, A. H., 2010	379	668	133,6
Yakkha	SCHACKOW, D., 2015	669	1027	205,4
Yupik do Alasca central	MIYAOKA, O., 2012	201	543	108,6
Zapoteca de San Bartolomé Z.	SONNENSCHNEIN, A. H., 2004	387	435	87

Tabela 2. Origem dos dados das línguas utilizadas.

Os dados provenientes das gramáticas foram inseridos manualmente na plataforma *online* òcun, desenvolvida pelo Laboratório de Tipologia (LATIP-UFBA), lotado no Instituto de Letras da Universidade Federal da Bahia, seguindo o método desenvolvido e exposto em Lazzarini-Cyrino e Mattos (2020). A plataforma preserva a estrutura em três linhas das gramáticas descritivas, com segmentação morfológica, glosas e traduções, conforme pode ser visto na figura 2:

## Língua Sumério

**Informações da Fonte**  
**Nome:** A Descriptive Grammar of Sumerian  
**Autor:** Abraham Hendrik Jagersma  
**Ano:** 2010  
[Abrir modo texto...](#)

**Frases da língua**  
Número de frases: **379**

en lord	en lord	ene PL	še TERM
------------	------------	-----------	------------

"for the lords"

dumu child	ane her
---------------	------------

"her child"

Figura 2. A plataforma òcun, dados do sumério.

A plataforma também disponibiliza os dados em modo texto, bastando copiá-los, colá-los e salvá-los como arquivo .txt para posterior processamento. A figura 3 mostra que o modo texto continua preservando a estrutura em três linhas dos dados:

```

en-en-enē-še
lord-lord-PL-TERM
for the lords

dumu-ane
child-her
her child

má-da tuš-Ø-'a
boat-COM sit-NFIN-NOM
who remains with the boat

```

Figura 3. Fragmento do documento de texto com sentenças segmentadas e glosadas; exemplo do sumério.

Uma vez salvos, os arquivos .txt foram processados por um programa feito especificamente para separar e listar as glosas dos morfemas na segunda linha dos trios, e ordená-los por frequência absoluta. Para cada uma das 19 línguas, foram geradas listas ordenadas por frequência, com os morfemas mais frequentes no topo e os menos frequentes no fim. Destas listas de morfemas, calculamos os valores correspondentes a 20% dos tipos de morfema para cada língua, e amostramos tal valor a partir dos morfemas do topo da lista. A tabela 3 ilustra o resultado do procedimento:

<b>morfe</b>	<b>freq</b>	<b>língua</b>
comp	143	zapotec
1sg	141	zapotec
cont	92	zapotec
3inf	67	zapotec
pot	61	zapotec
3f	56	zapotec
of	36	zapotec

Tabela 3. Os sete primeiros morfemas, ordenados por frequência, dentre os 87 morfemas que representam 20% do total de morfemas obtidos do zapoteca.

Tendo em vista que o propósito desta dissertação é de analisar a complexidade gramatical das línguas, interessam-nos apenas os morfemas (majoritariamente) gramaticais.

As listas geradas, porém, contêm uma mistura de morfemas cujos significados são gramaticais, lexicais ou intermediários, necessitando-se ser feita uma análise e triagem do material. Seguimos a proposta de William Croft (2002) para separar morfemas de acordo com o significado, face às imprecisões e problemática na delimitação de palavras, afixos, clíticos e outros conceitos baseados em *palavra*. Croft sugere classificar os significados dos morfemas de acordo com um *continuum* de quatro dimensões, no qual um extremo é mais lexical e o outro, mais gramatical. Vejam-se suas palavras:

Podemos resumir sugerindo que a distinção entre significado lexical e gramatical é melhor descrita como um *continuum*. Significados gramaticais são definidos sobre domínios que atravessam grandes áreas da experiência humana, são extremamente gerais em sua extensão, são grandemente definidos em termos de uma conceptualização da experiência humana, e requerem acesso a uma menor quantidade de conhecimento enciclopédico que significados lexicais (CROFT, 2002, p. 260, tradução nossa)

A classificação, portanto, foi feita de forma individual para cada morfema dentre os 20% de cada uma das 19 línguas, por meio do uso de formulários Google. Planilhas foram subsequentemente geradas, contendo morfemas que receberam escores mais gramaticais, mais lexicais, e intermediários. Foram descartadas todas as entradas referentes aos morfemas que não tivessem recebido escores mais gramaticais, e as tabelas resultantes foram conseqüentemente usadas para os próximos passos, das medidas de morfemas gramaticais livres e de complexidade absoluta.

Cada morfema foi classificado entre livre e preso com base na descrição da gramática de sua língua e a maneira com que tivesse sido glosado, recebendo valor positivo (S) caso preso, e negativo (N) caso livre; clíticos foram considerados presos. Já para as medidas de complexidade, cada morfema gramatical da nossa amostra foi individualmente analisado, tendo sido utilizados como parâmetros: 1. se o morfema é veiculador de flexão contextual *sensu* Booij e *sensu lato*, uma forma de sobrespecificação<sup>7</sup>; 2. o grau de complexidade fenogramatical/verbosidade<sup>8</sup> do morfema, também uma forma de sobrespecificação; 3. se o morfema possui elaboração estrutural e/ou irregularidades sob a forma de alomorfia ou supleção; e 4. se o morfema carregava informação de mais de uma categoria gramatical, isto é, se era poliexponente. Um quinto parâmetro, *proporção de morfemas livres*, foi comparado com os demais quatro.

---

<sup>7</sup> *Sensu* McWhorter; a ser apresentada no capítulo 3.

<sup>8</sup> *Sensu* Dahl, também a ser apresentada no capítulo 3.

O parâmetro 2 foi contabilizado atribuindo-se valores (de 1 a 3) para níveis de complexidade fenogramatical de cada morfema, somando-os e calculando-se a média para cada língua. Quanto aos parâmetros 1 e 3-5, utilizamos a frequência relativa: para cada língua, foi contabilizado seu total de morfemas gramaticais; em seguida, calculamos a frequência (absoluta) dos morfemas gramaticais que a) foram veiculadores de flexão contextual; b) possuíam elaboração estrutural e/ou irregularidades; c) carregavam significado de mais de uma categoria gramatical; e d) eram presos ou livres. Para cada um destes, as métricas obtidas foram utilizadas para calcular a frequência relativa  $F_r$  em relação aos morfemas gramaticais totais. O processo é resumido segundo a fórmula:

$$F_r = \frac{F}{N}$$

A expressão acima tem como numerador  $F$ , que é a frequência absoluta de morfemas gramaticais veiculadores de significados que se encaixem nos parâmetros a-d expostos acima – isto é, o número de morfemas que apresentem resultado positivo para um dado parâmetro numa dada língua –, e  $N$  como denominador, que é o número de morfemas gramaticais totais nesta mesma língua. Deste quociente, foi obtido o valor decimal correspondente à frequência relativa de morfemas apresentando cada um dos parâmetros por língua.

Os gráficos presentes nesta dissertação foram confeccionados com o pacote *ggplot* da linguagem R, assim como os testes estatísticos foram feitos na mesma linguagem.

#### 1.4. A estrutura desta dissertação

Este capítulo introdutório apresenta de forma geral um histórico do estudo da complexidade das línguas, iniciando-se com a tipologia holística e culminando no atual consenso, em muito uma reação aos primórdios da disciplina, que é a hipótese da compensação. Também se apresentam as línguas usadas na pesquisa e suas fontes, e o método seguido nas análises, a obtenção e processamento dos dados. Nos capítulos seguintes, não só os conceitos utilizados são tratados em maior detalhamento, como as subsequentes etapas do método são explicadas mais cuidadosamente.

O segundo capítulo aborda a divisão entre morfemas gramaticais e lexicais, a confusão que se encontra entre os termos *afixo*, *raiz*, *morfema gramatical* e *morfema lexical* e os planos *semântico* e *formal*, assim como o método aqui utilizado para discriminar por significado os morfemas das línguas do estudo. O terceiro capítulo traz uma discussão detalhada do conceito de *complexidade*, diferenciando as duas abordagens – a *absoluta* e a

*relativa* –, e também demonstrando a taxonomia da complexidade de John McWhorter, a distinção fenogramática *versus* tectogramática de Haskell Curry (1961) e a complexidade fenogramatical-verbosidade de Östen Dahl, e uma das medidas de complexificação de Peter Trudgill (2011). No quarto capítulo, temos uma descrição detalhada do método utilizado nesta dissertação para quantificar a flexão contextual, a complexidade fenogramatical e a poliexponência nas línguas usadas. O quinto capítulo traz algumas reflexões acerca do papel do contato linguístico na redução da complexidade gramatical, e analisa os casos das línguas menos complexas de acordo com os parâmetros desta dissertação, bem como o caso das línguas das Américas.

Já no sexto e último capítulo, tecemos algumas considerações finais acerca dos resultados aqui encontrados e suas implicações para o estudo do fenômeno da complexidade sob o enfoque da tipologia linguística, mormente da classificação tipológica das línguas, apontando também possíveis caminhos futuros de pesquisa na mesma linha.

## 1. DA ANÁLISE MORFÊMICA

Assim como a noção de *palavra* é controversa, a própria definição de *morfema* não é sem seus problemas de circunscrição. Fazemos direta alusão aqui à reflexão criteriosa tocada por Soledade e Lopes (2015), que encerra uma proposta de se conceber o morfema como uma categoria inserida num *continuum* de unidades não-discretas. Nas próprias palavras dos autores, sua concepção de *morfema* é:

[...] uma unidade recorrente que se encaixa em uma posição esperada na cadeia sintagmática interna do vocábulo, passível de ser apreendida como formativo através da análise mórfica via comutação e, geralmente (portanto, não sempre), apresentando uma carga semântica, que pode ser de matiz gramatical (funcional), categorial, lexical, contrastivo (distintivo) ou modificacional. [...] à categoria morfema corresponde um rol de elementos distintos (mas semelhantes, com um ou mais traços característicos em comum), sob o prisma do significante ou sob o prisma do significado, configurando uma escala gradativa de prototipicidade, que engloba desde os morfemas mais típicos, biplanos, dotados de um conteúdo (extra ou intralinguístico) e de uma realização acústica, até aqueles monoplanos, em que se encontra ausente (ou opaca) a face significante ou a face do significado. (*ibid.*, pp. 449-450)

Conforme veremos a seguir, na discussão sobre a distinção entre morfemas lexicais e morfemas gramaticais – também um *continuum* compreendendo dois extremos mais prototípicos e inúmeros elementos intermediários e não-modulares –, a linguística historicamente propôs categorias binárias e estanques na análise da língua, que escondem uma realidade muito mais complexa e fluida. Tais categorias têm sido alvo de intenso debate; a seção 2.1. apresenta uma problematização do binarismo gramatical *versus* lexical, bem como a atitude adotada no presente trabalho.

Quanto ao morfema, ficamos com a conceituação acima de Soledade e Lopes (2015), também operacionalmente aduzida pelos autores de uma maneira bastante afim com a natureza do trabalho aqui desenvolvido: o apreensível a partir de uma análise mórfica são os morfemas, por somente eles serem caracterizáveis como formativos morfológicos e genolexicais. Tendo em vista a fonte dos dados aqui utilizados ser a segmentação morfológica presente nas gramáticas descritivas das línguas abordadas, o uso de uma definição de morfema que traga em si referência à origem por meio da análise mórfica do material é compreensivelmente adequado<sup>9</sup>.

---

<sup>9</sup> Para um detalhamento da excelente argumentação por trás da conceituação, bem como um estudo das diferentes acepções de morfema ao longo do tempo e de diferentes autores, fazemos referência ao texto em si de Soledade e Lopes (2015).



[A] tradicional distinção entre palavras de conteúdo e palavras de função é uma em que palavras com o mesmo conteúdo geralmente têm formas livres não-marcadas através das línguas, enquanto palavras de função são frequentemente melhor traduzidas por morfemas presos em outras línguas. Também, diacronicamente, morfemas de conteúdo tendem a se manter livres, enquanto morfemas de função não raro migram entre formas presas (sufixais), semi-livres (clíticos) e livres. (KORNAI, 2020, p. 131. tradução nossa)

Como bem colocado por William Croft (2002), a divisão envolve quatro denominações muito utilizadas – *raiz*, *afixo*, *morfema gramatical ou de função* e *morfema lexical ou de conteúdo* –, e que não raro se confundem. Por exemplo:

Um morfema lexical é um morfema que pode ocorrer como uma palavra fonológica por si só, isso é, sem estar preso a um ou mais outros morfemas. **Morfemas gramaticais** não podem ocorrer como palavras fonológicas por si só, mas são **obrigatoriamente presos** a outros morfemas. (JAGERSMA, 2010, p. 77, tradução e grifos nossos)

No exemplo trazido acima, Abraham Jagersma utiliza um conceito formal – de morfema *preso* ou *livre* – para definir um conceito semântico – o de morfema *gramatical* ou *lexical*. Já Haspelmath e Sims colocam que “enquanto a distinção entre raízes por um lado e afixos por outro é em si bastante útil, esses conceitos acabam por ser mais complicados do que a imagem simples que vimos até então” (2010, p. 22, tradução nossa), e que “um afixo se prende a uma palavra ou à parte principal de uma palavra. Geralmente tem um significado abstrato” (*ibid.*, p. 19). É frequente assumir que formas presas têm significados mais “gramaticais” ou “abstratos”, e formas livres, significados mais lexicais, como vemos em Velupillai:

Em inglês, raízes e radicais tendem a ser livres, mas há muitas línguas em que são presos [...] A raiz e o radical carregam informação lexêmica, isto é, a informação semântica básica da palavra [...] **um afixo, por outro lado, é um morfema obrigatoriamente preso que não carrega informação lexêmica alguma.** (2012, p. 90, tradução e grifos nossos)

No capítulo 27, *Lexical and grammatical meaning* do livro *Morphology: A handbook on inflection and word formation* (2002), Croft expõe de maneira clara os problemas de se confundirem tais conceitos, mostrando que a caracterização dos morfemas de uma língua pode se dar em dois diferentes planos, o formal e o semântico. A distinção raiz *versus* afixo, portanto, pertenceria a uma caracterização no plano formal, e a morfema gramatical *versus* lexical, a uma caracterização no plano semântico.

Muitas vezes, há uma correlação direta entre os dois planos, e por isso é prontamente discernível se um morfema é lexical ou gramatical, se é uma raiz ou um afixo – além de haver correspondência entre raízes e morfemas lexicais, e afixos e morfemas gramaticais. Veja-se por exemplo o caso de uma expressão mapudungun:

4. Mapudungun (Chile e Argentina: araucana)

tripa-ke-rke-fu-y-ng-ün

sair-CF-REP-IPD-IND-3NS-PL

“(…) [eles] saíam (…)” (SMEETS, 2008)

É relativamente descomplicado, nesse exemplo, apontar que a raiz é *tripa* “sair”, e os morfemas *-ke* CARACTERÍSTICA FREQUENTE, *-rke* RELATADO, *-fu* IMPEDITIVO, *-y* INDICATIVO, *-ng* terceira pessoa, não singular e *-ün* plural, são afixos. Igualmente, pode-se afirmar sem maiores dificuldades que *tripa* é um morfema lexical, e os demais componentes, morfemas gramaticais.

No entanto, a situação nem sempre é tão clara:

5.

a. Djabugay (Austrália: pama-nyungana)

gulu-nda

ele-DAT

“Para ele” (CROFT, 2000)

b. Mam (México/Guatemala: maia)

t-ee

3sg-para

“Para ele” (*ibid.*)

No exemplo do djabugay, a raiz é o morfema *gulu* “ele”, à qual se sufixa *-nda* DATIVO. No entanto, a mesma construção em mam apresenta a situação inversa: a raiz é *ee*, “para”, à qual se prefixa *t-*, terceira pessoa do singular.

Uma vez que os dois planos são ortogonais, a caracterização de um dado morfema se daria, sob a acepção de Croft, com base nas duas dimensões, a formal e a semântica. Dessa forma, assim como existem morfemas gramaticais afixais – a exemplo dos morfemas presos a

*tripa* “sair” em 4. –, também existem morfemas gramaticais livres – frequentemente chamados na literatura de *partículas*, *palavras de função* ou *raízes vazias*, como o *le* no exemplo a seguir do mandarim:

6. Mandarim (China: sino-tibetana)

shìqǐng	yǐjīng	guòqù	le
assunto	já	passado	PFV

“(o assunto) são águas passadas.”

Nem todos os linguistas, porém, aceitam que partículas sejam tidas por raízes, por considerarem *raiz* como algo a que se aplicam flexões<sup>10</sup> (CROFT, 2000) – um reflexo de uma tipologia balizada no padrão encontrado na maioria das línguas europeias. Mas mesmo sob essa acepção há raízes de conteúdo gramatical; é esse o caso, por exemplo, dos auxiliares do kanoê, aos quais se prefixam vários morfemas também gramaticais, como desinências de pessoa e modo declarativo:

7. Kanoê (Brasil: isolada)

nukũ	aj	ry	õ-e-re
caba	1SG	ferroar	1-DECL-AUX

"A caba me ferrou." (BACELAR, 2004)

Já casos de morfemas lexicais livres e/ou com *status* de raiz, à semelhança de morfemas gramaticais afixais, são abundantes e intuitivos para boa parte dos ocidentais; é o exemplo das raízes em línguas indoeuropeias, ou as palavras lexicais (*palavras de conteúdo* ou *raízes cheias*) das línguas chinesas – no exemplo em 6., *shìqǐng* “assunto”. A última combinação, porém, de morfemas lexicais presos, sem *status* de raiz, mostra que a correspondência direta entre um extremo semântico e o formal não é regra. Veja-se:

8. Äiwoo (Ilhas Salomão: oceânica)

pelivano	sime	de-ku-luwa-kä-ji
crianças.3MIN	pessoa	coisa-IPFV-tomar.O-DIR:3-1+2MIN

“As crianças que adotamos” (NÆSS, 2017)

---

<sup>10</sup> O autor, ao se referir a *flexão*, entende-a como processos de afixação, e não como expressão de conteúdo gramatical independente da forma.

Em äiwoo, por exemplo, existem os ditos *substantivos genéricos presos*, que, como a denominação indica, não ocorrem de forma independente, e sim presos a outras raízes; essas formas têm correspondentes livres com o mesmo significado. Assim, para *de-* “coisa”, presente em 8., tem-se a forma independente *jelâ* “coisa”. É o mesmo caso de *gi-* “homem”, forma presa, e *sigiläi*, “homem”, forma livre; *nye-/nyi-* “lugar”, forma presa, e *numa* “lugar”, forma livre; e *si-* “mulher”, forma presa, e *singedâ*, “mulher”, forma livre.

Ademais, não só a correspondência entre os dois planos nem sempre é direta, como também a classificação de um dado morfema em um só plano já oferece desafios. Em se tratando de uma caracterização no plano semântico, a existência de processos de gramaticalização<sup>11</sup> dificulta qualquer classificação dicotômica entre gramatical ou lexical – uma vez que na prática essa divisão não possui contornos nítidos e abruptos. Em vez de uma distinção binária e absoluta, a proposta de Croft é de que cada morfema ocuparia uma posição num *continuum* entre puramente lexical e puramente gramatical, com muitos casos intermediários entre os dois pólos. Diacronicamente, os morfemas se moveriam entre esses extremos, de acordo com processos de relexificação ou gramaticalização, tornando a classificação algo dinâmico e não tão simples. O mesmo poderia ser dito de caracterizações no plano formal, com formas livres e presas nos extremos coexistindo com formas intermediárias, como os clíticos<sup>12</sup>.

### 2.1.1 As quatro dimensões de Croft

Vimos que para se classificarem semanticamente os morfemas de uma língua, a proposta mais básica é fazê-lo de acordo com suas posições num *continuum* entre um extremo lexical e outro gramatical, no qual não poucos acabam ocupando posições intermediárias (CROFT, 2000). Isso se dá pelo fato de a subsunção de cada um dos morfemas a duas classes claramente distintas ser impossível. Croft, portanto, sugere caracterizar cada morfema de acordo com quatro dimensões semânticas – natureza enciclopédica, natureza generalista, conceptualização do conteúdo da experiência e domínio semântico –, cada uma delas um *continuum*. Muito embora cada dimensão encarada sozinha possa apresentar

---

<sup>11</sup> O debate sobre gramaticalização, bastante extenso, não será apresentado nesta dissertação. Croft (2002), no entanto, realiza uma exposição preliminar do assunto, e também traz referências acerca do processo.

<sup>12</sup> Para os propósitos desta dissertação, clíticos serão encarados como formas presas, e o termo geral *morfema* será usado tanto para formas presas quanto formas livres.

exceções ou particularidades<sup>13</sup> – por exemplo, enquanto significados gramaticais tendem a ser mais gerais, também existem significados lexicais mais gerais, como “coisa”, “ir” e “ter” –, a união das quatro como baliza para a classificação semântica de um dado morfema torna-se mais segura (*ibid.*).

Tal classificação é independente de subsequentes classificações com base na forma, e por isso é a utilizada nesta dissertação. Isso se dá tendo em vista que não consideramos *palavra* como unidade de análise – logo, desconsideramos também distinções entre *afixos* e *clíticos*<sup>14</sup>, e até mesmo *raízes*; a proposta adotada é puramente semântica, e se baseia nas glosas dos morfemas. Cada uma das quatro dimensões será explicada a seguir.

### (I) Natureza enciclopédica

Segundo Croft, alguns semanticistas, como Ronald Langacker, propõem que significados lexicais possuem um caráter enciclopédico; isto é, em seu uso, acionam conhecimento de uma série de propriedades conhecidas acerca daquele item – sua cor, seu formato, seu comportamento, entre muitos outros.

Por óbvio, há características mais centrais que outras. Algumas são extremamente triviais e não compartilhadas por todos os falantes, como as que dizem respeito a itens individuais; ao falarmos em *banana*, não importa se o falante tem conhecimento de uma banana específica que sua irmã fatiou e comeu de certa forma em particular, e sim importam as propriedades centrais daquele item (LANGACKER, 1987), como por exemplo, seu formato, cor e comestibilidade. Outro aspecto é que algumas características são mais consistentes – não raro, o formato de algo é mais consistente que sua cor. Gatos podem ter as cores mais variadas, mas em geral, obedecem a um mesmo plano corporal (*ibid.*).

Croft, acerca da existência de propriedades mais centrais que outras, coloca:

Está claro que algumas propriedades semânticas são mais centrais para o significado de uma palavra que outras, particularmente aquelas que se aplicam a (quase) todas e apenas as instâncias do tipo, que são intrínsecas ao tipo, e que são de conhecimento convencional de (quase) toda a comunidade de falantes. (CROFT, 2002, p. 259)

---

<sup>13</sup> Para uma discussão das exceções e reverses no uso de cada uma das quatro dimensões, recomendo a leitura de Croft (2002).

<sup>14</sup> Tendo em vista que a diferenciação entre clíticos e afixos flexionais, bem como entre compostos e sintagmas, é de difícil operacionalidade e sem clareza (DIXON e AIKHENVALD, 2003; LIEBER e ŠTEKAUER, 2009; SPENCER e LUÍS, 2012; HASPELMATH, 2011, 2020), opta-se neste trabalho por utilizar o morfema como item mínimo de análise, sem discriminar serem formas presas ou livres.

O conhecimento enciclopédico, ao implicar o reconhecimento de propriedades de algum item, envolve diferentes aspectos de sua ocorrência no mundo real, compartilhados pelos falantes. Um falante, então, pode ter conhecimento enciclopédico do que é uma *árvore*, e acioná-lo para fazer uso da referência a árvore. No entanto, em geral, ele não tem conhecimento enciclopédico de “cardinalidades não-unitárias que influenciem [...] o uso do sufixo plural *-s*” (*ibid.*), um significado que é mais gramatical. É dizer, portanto, que enquanto significados lexicais possuem caráter enciclopédico, significados gramaticais não o possuem, e assim são estabelecidas as duas extremidades do primeiro *continuum* entre as quatro dimensões.

## (II) Natureza generalista

A segunda dimensão proposta consiste num *continuum* de generalidade do significado. Sucintamente, significados com valor mais gramatical são mais gerais que significados com valor mais lexical. Os pronomes ele/ela, por exemplo, são mais gerais que quaisquer substantivos aos quais façam referência. Croft (2002) apresenta também os exemplos do sufixo plural inglês *-s*, que é mais geral que quaisquer numerais, inclusive numerais vagos como *some* “alguns”, *a few* “uns poucos” e *many* “muitos”; e do tempo passado, que é mais geral que dêiticos temporais, como *yesterday* “ontem” e *long ago* “há muito tempo”.

## (III) Conceptualização do conteúdo da experiência

Em terceiro lugar, está a dimensão representada pelo *continuum* de conceptualização do conteúdo da experiência. Nas palavras de Croft, “a hipótese da conceptualização se baseia na observação de que significados não apenas simplesmente refletem a realidade objetiva, mas a enxergam ou até a moldam a partir de uma perspectiva particular” (2002, p. 259).

Vejamos os exemplos das sentenças dâw em 9:

### 9. Dâw (Brasil: nadahup)

a.

tír      yok      ãr

3SG    nadar   NEG

"Ele não nada/não está nadando." (CARVALHO, 2016)

b.

tir    yok    ãr    ãm  
3SG   nadar   NEG   TEL

"Ele já não nada/ele não nada mais." (*ibid.*)

A escolha do aspecto verbal diferencia as duas sentenças na medida em que ambas conceptualizam de forma diferente a frequência ou habilidades de natação do agente a quem o falante se refere. Em a), duas acepções são possíveis: o agente não sabe nadar ou não está nadando no momento do enunciado<sup>15</sup>. Já o uso do morfema *ãm* ASPECTO TÉLICO em b) exprime que o agente não nada mais, seja de forma permanente, seja por ter terminado a atividade de nadar naquele momento – isto é, conceptualiza a atividade de nadar de forma diferente da sentença em a).

Se compararmos a acepção de que o agente não sabe nadar em a) com a acepção de que o agente não tem mais essa habilidade em b), vemos como a diferença de aspecto, um significado gramatical, moldou a situação de modo a exprimir diferentemente uma propriedade do agente de acordo com o ponto de vista do falante, que pode ter escolhido dar destaque a uma ou a outra acepção – por exemplo, o agente não nada por nunca ter aprendido, ou não nada mais por conta de traumas decorrentes de um afogamento.

Utilizando exemplos do inglês, Croft comenta que ambas as sentenças trazidas por ele (e, por extensão, as nossas aqui) podem inclusive se referir à mesma situação, mas ditas por falantes diferentes. O autor então chega a afirmar que “significados gramaticais consistem em pouco mais que operações de conceptualização” (p. 260, tradução nossa).

#### (IV) Domínio semântico

Por fim, a última das quatro dimensões é o *continuum* referente à natureza do domínio semântico ao qual determinado significado pertence. O conceito é retirado de Langacker (1987), que expressamente define o domínio semântico como o contexto para a caracterização de uma dada unidade semântica, sendo os diferentes domínios entidades

---

<sup>15</sup> Perceba-se que o *dâw* é menos sobrespecificado que outras línguas ao codificar da mesma forma duas proposições diferentes – que em si sós já conceptualizam a experiência de forma diferente uma da outra. Um debate acerca da sobrespecificação encontra-se na seção 3.1.1.

cognitivas, uma vez que consistem em experiências mentais, espaços representacionais, conceitos ou complexos de conceitos. Nas palavras de Croft,

Uma inspeção nos significados gramaticais demonstra que certos domínios, como quantidade, tempo, espaço, causalidade, escalas, unidades e identidade, têm uma população desmesurada de significados gramaticais, enquanto outros domínios, como transações comerciais, beisebol, contratos de futuros e musicologia medieval não têm significado gramatical algum emoldurado a eles. (2002, p. 260, tradução nossa)

Croft equívale o conceito de domínio ao conceito de *frame* semântico proposto por Fillmore (1982), que é de “um sistema de conceitos relacionados de maneira que para compreender um deles, é necessário compreender toda a estrutura na qual ele se encaixa” (p. 111). Portanto, há uma tendência de os significados mais gramaticais pertencerem a domínios, ou *frames*, que lhes sirvam de *matriz* compartilhada com outros significados que em sua maioria também são mais gramaticais, e que em seu conjunto auxiliam na compreensão da função desempenhada por cada um de seus membros.

## 2.2. A triagem de morfemas

Tendo como base a proposta de Croft (2002) apresentada, os 20% mais frequentes dos morfemas das 19 línguas estudadas foram classificados um a um em cada uma das quatro dimensões de acordo com cinco valores: os extremos – L, mais lexical, e G, mais gramatical –, bem como um valor intermediário N, e gradações entre o valor intermediário e o extremo gramatical – NG – e entre o valor intermediário e o extremo lexical – NL (figura 4). Como visto nas tabelas 4 e 5, por exemplo, o morfema R1 “relacional 1” do araweté recebeu o valor G (mais gramatical) em cada uma das quatro dimensões. Já o morfema cujo significado é “jabuti” recebeu o valor L (mais lexical) também nas quatro dimensões.



Figura 4. Valores utilizados na classificação em quatro dimensões, de um extremo mais lexical ao mais gramatical.

Tabelas então foram geradas para processamento no R (tabelas 4 e 5)<sup>16</sup>. Na coluna *morfema*, tem-se a glosa do respectivo morfema da maneira encontrada na gramática descritiva, enquanto a coluna *freq* mostra a frequência de cada morfema na nossa amostra. *NE* refere-se à primeira dimensão, natureza enciclopédica; *NG* à segunda, natureza generalista; *CCE* à terceira, conceptualização do conteúdo da experiência; e, por fim, *DS* à quarta e última, domínio semântico.

morfema	freq	língua	NE	NG	CCE	DS
1	259	araweté	G	G	Ng	G
R1	258	araweté	G	G	G	G
FOC	243	araweté	G	G	G	G
2	148	araweté	G	G	Ng	G
23	103	araweté	G	G	Ng	G
13	101	araweté	G	G	Ng	G

Tabela 4. Seis dos morfemas mais frequentes do araweté e suas respectivas classificações nas quatro dimensões.

Em todas as línguas, os morfemas no topo da lista de frequência foram veiculadores de significados gramaticais (tabela 4), em contraste com os progressivamente menos frequentes, que *grosso modo* apresentaram valores mais lexicais; os últimos morfemas dentre os 20% mais frequentes do araweté na tabela 5 ilustram o fato. Tendo em vista que o propósito desta dissertação é analisar morfemas gramaticais, selecionamos apenas os morfemas que receberam G ou NG nas quatro dimensões, descartando os demais.

morfema	freq	língua	NE	NG	CCE	DS
quebrar	7	araweté	L	L	L	L
deitar	7	araweté	L	L	L	L
olhar	7	araweté	L	L	L	L
lavar	6	araweté	L	L	L	L
jabuti	6	araweté	L	L	L	L

Tabela 5. Os cinco morfemas com menores valores de frequência dentre os 20% mais frequentes do araweté, mostrando valores lexicais nas quatro dimensões, e, portanto, desconsiderados para a análise.

Uma vez descartados os morfemas mais lexicais, obtivemos listas de morfemas gramaticais para as 19 línguas, os quais foram submetidos às subsequentes análises de complexidade, abordadas nos capítulos 3 e 4.

<sup>16</sup> A lista completa com todos os morfemas analisados pode ser encontrada no anexo 1.

### 2.3. Proporção de morfemas livres

As 19 línguas utilizadas apresentaram proporções diferentes de morfemas gramaticais livres, conforme calculado por frequência relativa e demonstrado na forma gráfica na figura 5:

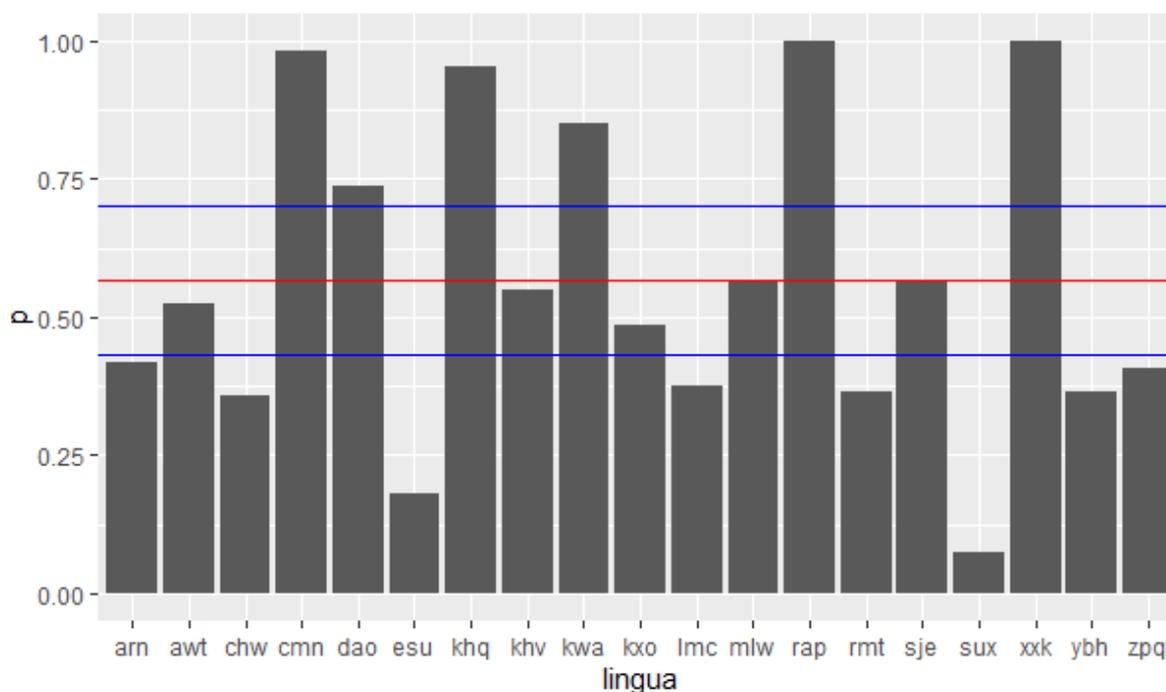


Figura 5. Proporção de morfemas gramaticais livres nas diferentes línguas, com média das proporções (em vermelho) e o intervalo de confiança<sup>17</sup> de 95% (em azul).

Seis línguas apresentaram proporção estatisticamente significativa (acima da linha vermelha) de morfemas gramaticais livres: mandarim (cmn, com 98,3%), daai chin (dao, com 73,7%), koyra chiini (khq, com 95,2%), dâw (kwa, com 85%), rapa nui (rap, com 100%) e kéo (xxk, com 100%). À exceção do koyra chiini, todas são descritas na literatura como sendo *isolantes* (SO-HARTMANN, 2009; DELMER, 2016; KIEVIET, 2017; MCWHORTER 2016, 2019; OBERT, 2020), o que, apesar da já exposta problemática em classificar holisticamente uma língua em tipos morfológicos, é consistente com suas altas taxas de morfemas livres. Este grupo de seis línguas, doravante referido como GML ([de morfemas] gramaticais majoritariamente livres), em geral se comporta da mesma forma em respeito aos parâmetros de complexidade, embora duas das línguas, o daai chin e o koyra chiini,

<sup>17</sup>O intervalo significa que 95% das amostras das línguas do mundo terão uma proporção média dentro desse intervalo.

apresentem algumas medidas que as aproximam das demais línguas não-GML, conforme veremos. Já línguas como o sumério (sum), o yupik do Alasca central (esu), o cuwabo (chw), o domari (rmt) e o yakkha (ybh) apresentaram as menores taxas de morfemas gramaticais livres, respectivamente com 7%, 18%, 35%, 36%, 36% e 37%.

## 2. CLASSIFICAÇÕES DE COMPLEXIDADE

Iniciamos a discussão acerca das diferentes classificações de complexidade surgidas na literatura, mormente a partir do final do século XX, apresentando a título de ilustração uma das instâncias em que o debate, inicialmente investido em classificar morfológicamente as línguas, começa a pender para uma avaliação da organização das peças de seus sistemas. É o caso da lista de Tauli, datada de 1945. Numa publicação no citado ano, o linguista estoniano Valter Tauli propõe uma lista com 11 critérios para se analisarem construções da língua quanto à *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 de uma expressão linguística, sendo a adição de subsequentes morfemas a ele numa ordem progressiva mais analítica que uma adição de elementos em ordem regressiva.

A lista de Tauli, embora pouco discutida pelos linguistas da época, tem como grande mérito a afirmação de que não a língua inteira, mas sim suas diferentes construções, é que poderiam ser alvo da análise em questão, uma conclusão à qual a linguística comparativa chegaria algumas décadas depois, malgrado ainda persistirem até o dia de hoje aplicações holísticas da terminologia classificatória morfológica. Uma observação, porém, merece ser feita: ao observarmos os critérios elencados em sua lista, percebemos que por *analiticidade*, Tauli parecia querer apontar, na verdade, formas menos *complexas* – menos *elaboradas* ou mais *diretas* – de realização de categorias gramaticais na morfologia. Vê-se aqui como as primeiras avaliações da complexidade das línguas fizeram-se inextricavelmente ligadas ao debate decorrente da classificação holística em tipos morfológicos – mas já indicavam emancipar-se dele.

Já no final do século XX e começo do século XXI, autores como David Gil (2008) e John McWhorter (1998, 2001, 2007, 2016, 2019 e possivelmente muitas outras) passam a argumentar contra a hipótese da compensação e o igualitarismo linguístico, embora sob um enfoque diferente do que inicialmente tinha dominado a discussão a respeito da complexidade das línguas. Não mais se estabelece a superioridade de uma língua em relação a outra – ao menos abertamente nos círculos acadêmicos –; o que autores como os citados argumentam é que nada garante que a diminuição da complexidade morfológica necessariamente implique aumento de complexidade em outros aspectos da língua, e que o fato de haver línguas menos complexas que outras é uma realidade inegável. Gil (2008) compara duas línguas “não-isolantes” (inglês e hebraico) com oito línguas tidas como “isolantes” – dois crioulos, papiamentu e bislama; duas línguas da África Ocidental, twi e iorubá; duas línguas do Sudeste da Ásia, cantonês e vietnamita; e duas línguas da Indonésia Ocidental, minangkabau e sundanês. Suas conclusões são contrárias ao que se esperaria segundo o igualitarismo linguístico; ao menos nos domínios morfologia, sintaxe e semântica composicional, as línguas “isolantes” aparentam ser consideravelmente menos complexas que as duas línguas “não-isolantes” utilizadas no estudo.

Da mesma forma, McWhorter, em seu livro *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars* (2007), traz uma série de argumentos contra a hipótese da compensação. A hipótese com a qual o autor trabalha ao longo de suas obras (1998, 2001, 2007, 2016, 2019) é de que a aquisição não-nativa por parte de um grande número de falantes tende a simplificar as línguas, sem necessariamente haver um aumento de complexidade em outro domínio como contrapartida. Ademais, McWhorter afirma que naturalmente, com sua transmissão não-interrompida, as línguas tendem a acumular complexidade “desnecessária”, não existindo casos de “deriva” que expliquem perda de complexidade como tão frequentemente assumido na literatura; a simplificação de, por exemplo, línguas românicas em relação ao latim nada tem de natural, e sim é o resultado da adoção do latim por aloglotas ao longo do território do antigo Império Romano. Como a aquisição por parte de adultos é imperfeita e tende a simplificar as filigranas e minúcias gramaticais da língua, a adoção por parte de uma massa de adultos de uma língua que não tenha uma norma escrita forte para coibir desvios resulta numa gramática significativamente “facilitada”. Não só essa interrupção na transmissão natural de uma língua resulta em gramáticas menos complexas, como também seria a *única* maneira de uma língua simplificar-se tão drasticamente, uma vez que o estado natural das línguas, como dito, é

acumular complexidades com o tempo<sup>18</sup>. Para McWhorter, o caso extremo dessa interrupção na transmissão natural, e consequente perda de complexidade, são as línguas crioulas – mas casos intermediários existem, nos quais a simplificação foi apenas parcial, e é nesses casos que o autor foca no livro.

Vale ressaltar que o autor usa com frequência a palavra inglesa *streamlined* para descrever línguas com menor complexidade, vocábulo este que tem conotações de otimização, agilidade e facilidade, diferentemente das conotações do século XIX de primitivismo ou debilidade do pensamento de seus falantes. Longe de um retorno aos conceitos do século XIX, o posicionamento de McWhorter é de que todas as línguas humanas são complexas – mas não necessariamente num mesmo grau, existindo uma complexidade de gradiente, sem demérito para as línguas que sejam menos complexas (mais *streamlined*). O autor avisa explicitamente aos leitores que o fato de que não há “línguas primitivas” já é consenso há muito confirmado.

No livro, McWhorter analisa cinco casos de línguas que, segundo ele, têm histórias de aquisição não-nativa em seu passado: o inglês, o persa, o mandarim, o malaio e os diversos árabes vernaculares<sup>19</sup>. A obra se inicia com uma definição em três subtipos do que é complexidade *sensu* o autor, e como medi-la<sup>20</sup>; uma vez resolvida a questão terminológico-metodológica, segue-se a maior parte do livro. Ao longo dos cinco capítulos nos quais o autor cobre essas línguas, são feitas comparações entre elas e línguas a elas relacionadas: compara-se o inglês a línguas germânicas como o alemão, o islandês e até o afrikaans; o persa às demais línguas irânicas, como o curdo, o pashto e o balochi; o mandarim às línguas siníticas mais ao sul, como o cantonês e o min; o malaio às línguas austronésias como as das Filipinas; e os árabes vernaculares sedentários ao árabe dos beduínos, o árabe clássico e outras línguas semíticas como o aramaico. Sua conclusão é que as cinco línguas, características de sociedades pós-neolíticas com história de expansão e aquisição não-nativa, são marcadamente mais simplificadas que suas línguas-irmãs em basicamente todos os parâmetros estabelecidos no livro para se medir a complexidade.

Outro dos entraves no estudo da complexidade das línguas é representado pela ampla e pervasiva confusão a respeito do próprio termo *complexidade*. Muitas vezes, entende-se complexidade como dificuldade de aquisição; autores como Kusters (2008) utilizam este

---

<sup>18</sup> No capítulo 5, questionamos a exclusividade da adoção por aloglotas na simplificação extrema da gramática de uma língua, usando dados do dáw.

<sup>19</sup> É digno de nota que o autor inclui o português popular brasileiro dentre as línguas de gramática simplificada por conta de um histórico de aquisição adulta por aloglotas, embora não se detenha muito nessa língua em específico.

<sup>20</sup> A definição do que é complexidade linguística e seus subtipos *sensu* McWhorter é apresentada na seção 3.1.

conceito ao se referir a complexidade; Kusters inclusive define complexidade tendo a relação com aprendizes adultos como base – e, como consequência, define como *complexas* as propriedades da língua que causam dificuldades aos aprendizes de L2.

No entanto, esta não é a única concepção possível, tampouco a utilizada nesta dissertação. Matti Miestamo (2008), abordando os problemas metodológicos do estudo da complexidade através das diferentes línguas, identifica dois diferentes enfoques: o absoluto e o relativo. O relativo considera a complexidade como *custo* ou *dificuldade*<sup>21</sup> aos usuários da língua, como colocado por Kusters; um fenômeno linguístico ser visto como simples ou complexo depende do ponto de vista a ser tomado – complexo para quem? Para o falante, o ouvinte, o adquirente L1, o aprendiz L2? É dizer, o caminho a ser tomado, bem como as definições, são função direta dos objetivos da pesquisa e do ponto de vista a ser levado em conta. Já o enfoque absoluto, utilizado por autores como McWhorter e Östen Dahl (2004), encara a complexidade como uma propriedade objetiva do sistema em estudo. É a essa acepção de *complexidade* que esta dissertação se afilia, considerando que quanto mais partes<sup>22</sup> um sistema tem, mais complexo ele é.

Por mais que se possam apresentar os problemas do enfoque relativo – como dito, *relativo a quem?* –, o enfoque absoluto da complexidade também não é sem seus desafios. O já conhecido problema da comparabilidade (CROFT, 2003) na tipologia se apresenta, obviamente, mas também a impossibilidade de se quantificar uma complexidade global (MIESTAMO, 2008). A solução para Miestamo, portanto, reside em focar em domínios específicos que sejam comparáveis através das línguas.

### 3.1. A taxonomia da complexidade de McWhorter

No segundo capítulo de seu livro *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars* (2007), John McWhorter delinea o que chama de uma *taxonomia da complexidade* – tendo em vista aqui que seu conceito, como o de Dahl, segue uma abordagem absoluta da complexidade. Para McWhorter, a complexidade pode ser

---

<sup>21</sup> Uma distinção entre complexidade, custo e dificuldade é apresentada na seção 3.2.

<sup>22</sup> Com a expressão *peças*, queremos dizer quaisquer e cada um dos processos que veiculem informação gramatical, seja por adição de material morfológico, seja por alterações segmentais ou suprasegmentais, seja por alterações na ordem dos constituintes; fazemos alusão aos possíveis mecanismos de veiculação de *conceitos de conteúdo* e *conceitos relacionais* segundo Edward Sapir (1921), e também aos diversos desvios do estado de concatenação irrestrita que consubstanciam a complexidade fenogramatical de Dahl (2004). Cada forma de expressar um significado gramatical – veiculando um *conceito* ou afastando a expressão de uma concatenação irrestrita – conta como uma peça do sistema.

medida de acordo com três aspectos: a sobrespecificação (*overspecification*, no original), a elaboração estrutural e a irregularidade.

### 3.1.1. Sobrespecificação

Sobrespecificação se refere ao grau de especificação para além do estritamente necessário para a comunicação que uma determinada língua use para expressar determinada categoria semântica. Nas palavras do próprio autor:

As línguas diferem no grau em que explícita e obrigatoriamente marcam distinções semânticas. Denominarei essa diferença como sendo de sobrespecificação [...] na sua marcação gramaticalizada de definitude e indefinitude de sintagmas nominais, o inglês é sobrespecificado em relação ao karok, que não possui esses traços. Nenhuma gramática é desprovida de sobrespecificação. Isto é, **todas as gramáticas marcam um número de distinções que não são necessárias à comunicação**. Por exemplo, à medida em que a maioria das gramáticas marca uma distinção clássica entre três pessoas no singular e no plural nos pronomes, flexão e daí por diante, elas são sobrespecificadas, uma vez que algumas línguas marcam apenas três pessoas, sem especificação de número. (2007, pp. 21-22. tradução e grifos nossos)

O argumento de McWhorter é que muito embora todas as línguas possuam gramáticas sobrespecificadas em alguma medida, elas em geral não são igualmente sobrespecificadas. As diferentes gramáticas difeririam no grau de sobrespecificação – e embora diferenças mínimas em grau sejam possíveis devido a uma possível deriva e ao acaso, diferenças extremas de sobrespecificação entre línguas se dariam pelo fato de a língua menos sobrespecificada ter passado por um período de aquisição não-nativa em sua história, logo, ter tido sua gramática simplificada (*streamlined*). É dizer, tais línguas teriam descartado características menos fundamentais à comunicação, especialmente em se tratando de características de difícil aquisição adulta.

O autor então elenca numa lista “representativa, mas não exaustiva” (p. 22) as características gramaticais que representam maior grau de sobrespecificação, e que servirão para sua própria análise das cinco línguas a que se propõe estudar no livro. As características trazidas por McWhorter são marcação de classe nominal, marcação numeropessoal, marcação obviativa, posse inalienável, marcação de definitude, marcação mais minuciosa de tempo-aspecto-modo, marcação evidencial, presença de cópulas ou maior número de diferentes cópulas, distintas formas de negação, gradação mais minuciosa de demonstrativos, ergatividade, ajuste de valência, presença de marcação de foco, presença de partículas pragmáticas e relativização distinta de subordinação.

Para ilustrar o conceito mais precisamente, utilizarei os exemplos de sobrespecificação de cópulas trazidos no livro. O primeiro caso, menos sobrespecificado, é o de línguas como o russo, em que a cópula é ausente em sentenças neutras (*Ivan ø moj brat'* “Ivan é meu irmão”), mostrando que só o fato de uma língua marcar morfologicamente a relação entre o sujeito e o predicado não-verbal já é uma sobrespecificação. Em relação ao russo, a sentença equivalente do inglês é mais sobrespecificada, por marcar a relação com a cópula *be* (*Ivan is my brother*).

Mais sobrespecificada ainda neste quesito é a gramática do fongbe<sup>23</sup>, que utiliza diferentes cópulas para o caso de o predicado ser equativo ou locativo. Veja-se:

#### 10. Fongbe (Togo e Benim: nígero-congolesa)

a.

ùn	nyí	Àfiávi
1sg	ser	Afiavi

“Eu sou Afiavi.” (LEFEBVRE; BROUSSEAU, 2001)

b.

wémâ	ó	dó	távò	jí
livro	DEF	estar em	mesa	sobre

“O livro está na mesa.” (*ibid.*)

As sentenças fongbe em 10., então, demonstram uma maior minúcia desta língua em codificar distinções semânticas que, na sentença neutra, não são cobertas de forma alguma em muitas línguas, como o russo, ou marcadas de forma menos específica em outras, como o inglês. Um caso de distinção ainda mais finamente marcada, porém, é trazido: o da língua panare, que apresenta cópulas diferentes para diferentes distâncias em relação ao falante, e também animacidade:

#### 11. Panare (Venezuela: caribe)

a.

maestro	kěj	e'ñapa
professor	COP.ANIM:PROX	Panare

<sup>23</sup> Ou do próprio português, com as cópulas “ser” e “estar”.

“O panare (aqui) é professor.” (GILDEA, 1993)

b.

maestro	něj	e’ñapa
professor	COP.ANIM:DIST	Panare

“O panare (lá) é professor.” (*ibid.*)

c.

e’chipen	mën	manko
fruta	COP.INAN	manga

“A manga é uma fruta.” (*ibid.*)

A presença, portanto, de distinções mais minuciosas e detalhadas na marcação de cópulas aumenta o número de partes do sistema – sua complexidade absoluta – por meio de maior sobrespecificação.

As medidas e conceitos de flexão contextual e complexidade fenogramatical-verbosidade, como formas de sobrespecificação, serão tratados detalhadamente no capítulo 4. A seguir, apresentamos os outros dois elementos da taxonomia da complexidade de McWhorter.

### 3.1.2. Elaboração estrutural

McWhorter define *elaboração estrutural* como o grau em que uma dada língua se diferencia de outra com base no número de regras fonológicas e sintáticas ou de elementos do inventário fonêmico que são necessários para se gerarem formas de superfície. A complexidade de uma língua, portanto, cresce à medida que cresce o número de regras necessárias para se formarem as sentenças em sua gramática. O autor também traz uma lista exemplificativa de casos que aumentam a elaboração estrutural numa língua: riqueza de processos morfofonêmicos, inventário fonêmico com mais elementos marcados, existência de classes de gênero gramatical/declinação e conjugação, alternância de ergatividade com alinhamento nominativo/acusativo e ordem heterogênea de palavras.

Para exemplificar, utilizarei os exemplos de McWhorter (2007) para riqueza de processos morfofonêmicos. Por mais que todas as línguas naturais apresentem tais processos,

o exemplo trazido, das línguas célticas, é caracteristicamente rico em mutações consonantais. É esse o caso do galês:

#### 12. Galês (País de Gales: indoeuropeia)

eu	cath	[kaθ]	“o gato deles”
fy	ngath	[ŋaθ]	“meu gato”
ei	gath	[gaθ]	“o gato dele”
ei	chath	[χaθ]	“o gato dela”

(MCWHORTER, 2007)

Em galês, a forma de citação do substantivo “gato” é *cath*, acompanhado de *eu* “deles”. Mas como visto em 12., sua consoante inicial sofre diferentes mutações de acordo com o pronome possessivo que o acompanha. É de se salientar que apenas a mutação consonantal diferencia o gênero entre *ei gath* “o gato dele” e *ei chath* “o gato dela”; este é um caso em que as fronteiras da fonologia e da morfologia se borram.

Igualmente, tem-se o desencadeamento da mutação consonantal em objetos:

#### 13. Galês

a.

gwelodd	ci
viu	cachorro

“O cachorro viu.” (MCWHORTER, 2007)

b.

gwelodd	Alun/ef/ø	gi
viu	Alun/ele/ele	cachorro

“Alun/ele viu o cachorro.” (*ibid.*)

O autor aduz ser óbvio que a morfofonêmica do galês é mais complexa que a do inglês, uma vez que o galês tem mais alternâncias morfofonêmicas, que são condicionadas por mais interfaces com a gramática que as do inglês – indo além da mera eufonia –, e não raro são foneticamente imprevisíveis, necessitando de memorização. Ademais, segundo ele, a existência desses processos é prontamente perceptível a qualquer falante de galês, enquanto

os falantes de inglês necessitam de instrução formal para entenderem do que se trata a morfofonêmica.

### 3.1.3. Irregularidade

O último aspecto da complexidade entre os três definidos por McWhorter é a irregularidade. A presença de irregularidades num sistema implica uma maior lista de itens separados na gramática que devem ser memorizados e acionados caso a caso; aumenta-lhe as partes e o tamanho de sua descrição, logo, sua complexidade absoluta. A lista demonstrativa trazida no livro inclui irregularidades de gênero gramatical e classes nominais, de marcação de plural, de formas das raízes e presença de supleção.

Um exemplo expressivo de irregularidade como elaboração de uma língua trazido por McWhorter é a existência de plurais irregulares, como por exemplo no inglês *child-children* (criança-crianças) e *person-people* (pessoa-pessoas). Nesse aspecto, o russo é mais complexo, por ter uma classe de substantivos com plural irregular ainda maior – *syn* "filho" e *synov'ja* "filhos" (em vez do regular *\*syny*) e *doč* "filha" e *dočeri* "filhas" (em vez de *\*doči*). O alemão revela-se portador de uma irregularidade ainda maior em seus plurais, que no masculino podem ser formados ora com o sufixo *-e*, ora por *umlaut*: *der Besuch* "a visita", *die Besuche* "as visitas", mas *der Gast* "o hóspede", *die Gäste* "os hóspedes". Por fim, o árabe e seus plurais quebrados são um grau extremo de irregularidade; a despeito de haver um plural regular sufixal nessa língua, ele é secundário. Os substantivos mais importantes formam seu plural de acordo com longas listas de interações entre a moldura triconsonantal e as vogais inseridas, e a qual padrão cada substantivo se subsume deve ser memorizado. São exemplos *kalb* "cachorro" e *kilāb* "cachorros", *kitāb* "livro" e *kutub* "livros" e *dawla* "país" e *duwal* "países" (MCWHORTER, 2007, pp. 33-34).

### 3.1.4. Medição da elaboração estrutural/irregularidade

Elaboração estrutural e irregularidade foram aqui medidas de acordo com supleção/alomorfia observados nas gramáticas para cada significado glosado, conforme se vê na tabela 6:

<b>morfema</b>	<b>freq</b>	<b>língua</b>	<b>PRESO</b>	<b>EE</b>
NON.FUT	414	daaichin	SN	N
DEM	314	daaichin	N	S
ERG	252	daaichin	SN	N
S.AGR:3S	174	daaichin	N	N
PL	161	daaichin	SN	S
CF	138	daaichin	SN	N
LOC	130	daaichin	SN	S

Tabela 6. Classificação de morfemas em livres *versus* presos (PRESO) e elaboração estrutural/irregularidade (EE); exemplos do daai chin.

Para cada morfema da tabela, indicados por significado, a gramática de sua respectiva língua foi consultada à procura de alomorfia, supleção e/ou irregularidades na sua forma, com o auxílio de ferramentas de busca. Caso encontradas diferentes formas de realização de um mesmo significado, a célula correspondente recebeu valor positivo (S); do contrário, valor negativo (N).

Estas medidas aumentam a complexidade de um sistema aumentando-lhe o número de itens. Maior grau de elaboração estrutural envolve um maior número de regras para gerar formas de superfície, indicando a existência de processos morfofonêmicos e alomorfes. Irregularidade e supleção também aumentam o tamanho da descrição do sistema por aumentarem o número de itens que devem ser memorizados. Como exposto, no presente estudo, medimos tais parâmetros por meio da presença de alomorfia, supleção e diferentes formas para um mesmo significado. A elaboração estrutural e irregularidade, doravante EE, também mostrou uma correlação negativa com a proporção de morfemas livres, conforme se depreende da análise do gráfico na figura 6:

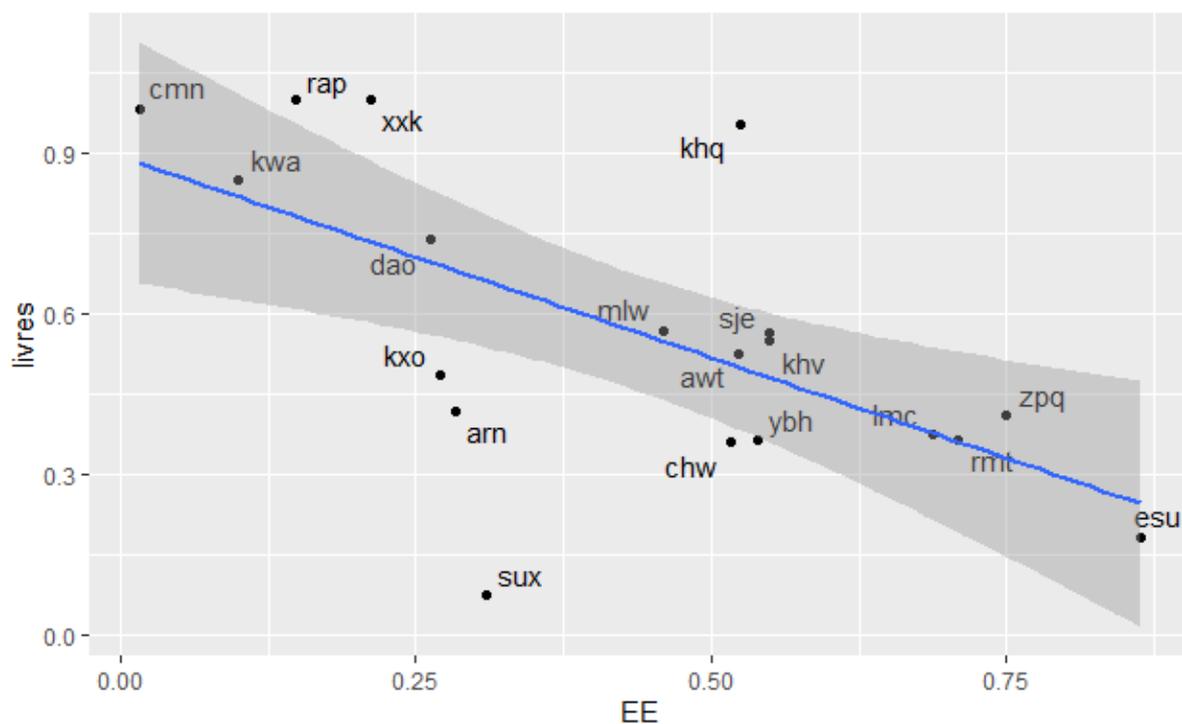


Figura 6. Relação entre proporção de morfemas livres e grau de elaboração estrutural (EE).

*Grosso modo*, vê-se que as línguas GML também apresentaram de baixa a baixíssima proporção de elaboração estrutural e irregularidade – o mandarim com 1,6%, o dêw com 10%, o rapa nui com 14%, o kéo com 21% e o daai chin com 26%. Uma notável exceção sendo o koyra chiini (com 52%), que apresenta extensa alomorfa e irregularidades em seu quadro de morfemas gramaticais, a exemplo dos pronomes. Vejam-se:

#### 14. Koyra-chiini (Mali: nilo-saariana)

forma irregular	forma regular-esperada	glosa
ye	ay ma	1SG.S.SUBJ
yene	ay se	1SG.DAT
yee	ay o	1SG.S.IPFV
ma	ni ma	2SG.S.SUBJ
mana~mane	ni se	2SG.DAT
ma na	ni na	2SG + NEG

(HEATH, 1999, pp. 41-42)

Como consequência, o koyra chiini encontra-se numa coordenada consideravelmente avançada em respeito ao eixo x, juntamente com línguas como o moloko (45,9%), o yakkha

(53,9%), o saami de Pite (sje, com 54,9%), o cuwabo (51%) e o khwarshi (khv, com 54,9%), amplamente descritas como *flexionadas* na literatura e de tipos aglutinantes ou fusionais (KHALILOVA, 2009; WILBUR, 2014; GUÉROIS, 2015; SCHACKOW, 2015; FRIESEN *et al.*, 2017).

### 3.2. A verbosidade de Dahl

A contribuição de Östen Dahl para o debate acerca da complexidade das línguas e sua tipologia é considerável. Em seu livro *The growth and maintenance of linguistic complexity* (2004), Dahl oferece uma série de conceitos e distinções a fim de munir os linguistas do vocabulário necessário para uma abordagem razoável e clara dos muitos problemas à frente.

Primeiramente, ele diferencia as noções de *custo*, *dificuldade* e *complexidade*, auxiliando no deslinde de uma confusão que os próprios linguistas fizeram por bastante tempo. *Custo* é a quantidade de recursos que um determinado agente gasta para atingir algum objetivo. Por recurso, entende-se aqui energia, tempo, dinheiro, ou qualquer outra coisa que seja investida para que o objetivo seja alcançado, e a relação entre custo e benefício – e seus *trade-offs* – é axial para muitas das descrições do comportamento comunicativo.

Por *dificuldade*, Dahl tem em mente um conceito que não se aplica no vácuo, e sim a tarefas, sempre relativamente a algum agente. Qualquer coisa – no caso, línguas – pode ser difícil para *alguém*. Mas dificuldade e complexidade para ele não são sinônimos; a dificuldade, por óbvio, varia de acordo com quem emite sua opinião sobre determinada língua, independentemente do inventário morfêmico, fonêmico ou das regras sintáticas da língua. Acaba sendo de maior relevância a exposição do indivíduo, a proximidade da língua com sua própria língua-mãe, dentre outros fatores subjetivos que não necessariamente dizem algo sobre a língua analisada. Dahl igualmente relaciona dificuldade com custo, pelo fato óbvio de que se tornam difíceis quaisquer tarefas com alto custo de recursos, ou que forcem o agente a ir além de suas capacidades de investimento.

A maior contribuição, porém, se dá com sua definição do conceito de *complexidade*. Diferentemente das noções de *custo* e *dificuldade*, que são relativas a um dado agente, Dahl afirma que a complexidade é mais *objetiva*, ao menos em princípio. Para compreender a proposta de Dahl, faz-se necessária outra distinção – a de *recursos* e *regulamentações* de uma língua. Qualquer língua tem disponível a si uma gama de *recursos*, mormente seu léxico, que diz respeito ao que nela pode ser expresso. Por complexidade de recursos, muitas vezes se entende o tamanho do léxico da língua – línguas com maiores vocabulários, por exemplo.

Mas não é este o foco da proposta de Dahl, e sim a complexidade das regulamentações de uma língua, ou seja, não o que *pode* ser expresso, mas o que *deve* ser expresso nela. Também pode ser entendida como a maneira que a língua expressa o que será expresso, alinhando-se com o conceito de complexidade de autores como McWhorter (2001), que têm em mente não a vastidão do léxico de uma língua ao utilizarem o termo *complexidade*, mas sim a riqueza de suas regulamentações – isto é, sua complexidade gramatical.

### 3.2.1. Tectogramática *versus* fenogramática

Ao abordar a complexidade, portanto, Dahl lança mão de mais uma distinção – dessa vez, proposta por Haskell Curry em 1961 – a de tectogramática *versus* fenogramática. Em suas palavras, Dahl aduz:

A distinção entre estrutura de escolha e estrutura de saída (*output*) é intimamente relacionada à introduzida pelo lógico Haskell B. Curry (1961), entre dois níveis de gramática, tectogramática e fenogramática: – tectogramática – “o estudo da estrutura gramatical em si mesma”, por exemplo, como “algo independente da maneira como é representada em termo de expressões”; – fenogramática – como a estrutura gramatical é representada em termo de expressões. Por exemplo, as expressões *two pound butter*, *two pounds butter* e *two pounds of butter* podem ser vistas como formas diferentes de realizar a mesma construção gramatical, logo diferindo apenas com respeito a sua fenogramática. Da mesma forma, as regras de estrutura sintagmática  $S \rightarrow NP VP$  e  $S \rightarrow VP NP$  não difeririam tectogramaticalmente, mas apenas fenogramaticalmente. (2004, p. 49, tradução nossa)

Na linguística pós-Chomsky, não é novidade o conceito que a gramática vai além do que se pode ver imediatamente, embora tal ideia remonte a Heráclito – *ἀρμονίῃ ἀφανῆς φανερῆς κρείττων* *harmoniē aphanēs phanerēs kreittōn*, “a estrutura latente é mestra da estrutura visível” (AFONASINA, 2012; KORNAI, 2020) –; em Noam Chomsky, dos anos 50 aos 80, trata-se da distinção entre estrutura superficial e estrutura profunda. Curry, porém, não tem tal distinção em mente ao sugerir a sua própria entre tectogramática e fenogramática, chegando a criticar os primeiros trabalhos de Chomsky por confundir as duas áreas da gramática (DAHL, 2004).

A tectogramática (do latim *tectus*, “coberto”, em oposição à fenogramática, do grego antigo *φαίνω φαίνῶ*, “aparente, visível”) preocupa-se com a combinação lógica de elementos, sendo também referida como sintaxe interna (*inner syntax*) por Kornai (2020). Este último autor, elaborando sobre os conceitos de Curry, traz como exemplo a sentença inglesa em 15.:

15. *the shooting of the hunter was terrible*

“o tiro do caçador foi terrível”/”o caçador ter recebido tiros foi horrível”

Nessa sentença, não é claro se o que é terrível é a habilidade do caçador (a) ou o fato de ele ter sido alvejado (b); no primeiro caso, o caçador é o agente que executa a ação, e no segundo, é o objeto/alvo do tiro. Embora ambas as proposições, em inglês, possuam a mesma forma aparente, tectogramaticalmente elas são diferentes. Para demonstrar, Kornai considera ser terrível (*be\_terrible*) um predicado unário, e inicialmente dispõe as duas proposições como:

- (a) hunter shoot be\_terrible
- (b) shoot hunter be\_terrible

Em seguida, ele lança mão de uma operação que utiliza dois elementos invisíveis, as noções de nominativo (NOM) e acusativo (ACC); daí, temos, tectogramaticalmente:

- (c) hunter NOM shoot
- (d) hunter ACC shoot

Em muitas línguas, estes elementos são abertamente marcados na morfologia sob a forma de afixos ou partículas indicando caso; por exemplo:

16. Rapa Nui (Ilha de Páscoa: austronésia)

ina	ko	kai	i	te	kai	mata
NEG	NEG.IPFV	comer	ACC	ART	comida	crua

“Não coma comida crua.” (KIEVIET, 2017)

Já em inglês (e em português), tal distinção é codificada pela posição; o nominativo vem antes do verbo, e o acusativo, depois. Tais diferenças de codificação da mesma proposição são diferenças fenogramaticais entre as línguas; ou seja, a fenogramática diz respeito à estrutura visível dos constituintes e sua ordem, enquanto a tectogramática se refere à estrutura função-argumento.

Outro exemplo útil para se compreender a distinção, também dado por Kornai (2020), é o dos *phrasal verbs* em inglês. Tais verbos são tectogramaticalmente similares a demais

verbos transitivos na medida em que selecionam sujeito e objeto (SUF verbo OBJ), mas fenogramaticalmente se distinguem pelo fato de seu objeto vir no meio no material fonológico, e não no fim – *call NP up* ou *put NP down* (respectivamente, “ligar para NP” e “abaixar NP”).

### 3.2.2. Complexidade fenogramatical e verbosidade

Tendo em vista que tectogramaticalmente, uma proposição em diferentes línguas codifica o mesmo conteúdo, o nível que se presta à análise da complexidade ao longo das várias línguas do mundo é o nível fenogramatical; é nele em que aparecem as diversas formas que as línguas têm para exprimir o que deve ser expresso por meio de mecanismos aparentes, frequentemente morfológicos. É no grau de elaboração desses mecanismos que as línguas divergem, e, segundo a concepção de Dahl, mostram-se mais ou menos complexas. Logo, o que tem interesse tipológico é a complexidade fenogramatical de uma língua.

A complexidade fenogramatical, portanto, é a medida na qual a gramática de uma dada língua desvia de um sistema hipotético de “concatenação irrestrita” – justaposição dos elementos sem ordem estabelecida ou adição de elementos fixos, como no exemplo *two pounds butter* –, que representaria a estrutura tectogramatical por baixo da estrutura aparente. Há dois grandes tipos de desvios de um sistema de concatenação irrestrita nomeados por Dahl: restrições na ordem dos elementos e verbosidade. A restrição na ordem dos elementos é encontrada em todas as línguas em certa medida. Por exemplo, em português, um numeral tem de preceder o núcleo nominal, então *três cachorros* é bem formado, mas *cachorros três* não. Da mesma forma, em inglês, *one dog* é bem formado; *dog one*, não. O autor afirma que é mais comum que padrões gramaticais nas línguas obedeçam a alguma forma de restrição na ordem dos elementos do que tenham ordem completamente livre.

O segundo desvio é de maior importância para a presente dissertação, e representa mais um conceito formulado pelo próprio Dahl – a verbosidade. Em suas palavras:

[...] Em inglês, palavras quantificadoras e substantivos incontáveis podem ser unidos por simples justaposição, como em *much snow* ou *little snow*. Em francês, por outro lado, uma palavra como *beaucoup* “muito” não pode simplesmente ser justaposta ao substantivo *neige* “neve”; em vez disso, deve-se inserir a preposição *de* entre os dois: *beaucoup de neige* “muita neve” [...] Do ponto de vista do inglês, porém, o elemento *de* parece redundante. Podemos dizer que as construções quantificadoras francesas são caracterizadas por verbosidade, definidas como tendo maior peso fonético (contendo mais material) do que seria minimamente necessário (logo, um caso especial de redundância). (2004, p.53, tradução nossa)

Dahl define concisamente a verbosidade como sendo a “dispensabilidade através das línguas” (*cross-linguistic dispensability*). Para usar o mesmo exemplo de *one dog* dado acima, podemos comparar o sintagma em inglês com seu equivalente em mandarim. Temos:

17. Mandarim

yī      zhī      gǒu  
um      CL      cachorro  
“Um cachorro”

Vê-se que tanto *one dog* quanto *yī zhī gǒu* desviam-se da concatenação irrestrita (simples justaposição de seus elementos) por terem uma ordem fixa, não livre; o numeral precede o substantivo quantificado. No entanto, a expressão em mandarim possui um elemento que sua equivalente inglesa não possui – o classificador nominal *zhī*. Portanto, *sensu* Dahl, ela é mais *verbosa*. Caso comparemos *three dogs* e seu equivalente em mandarim, teríamos:

18.

a. Inglês

three dog-s  
três cachorro-PL  
“Três cachorros”

b. Mandarim

sān      zhī      gǒu  
três      CL      cachorro  
“Três cachorros”

Percebe-se que embora o inglês não possua o classificador nominal, o sintagma apresenta material fonético que seu equivalente em mandarim não possui: a desinência de plural *-s*. Poderíamos dizer, portanto, que no tocante a essa construção, ambas as línguas se expressam com o mesmo grau de verbosidade. Destes exemplos, tiramos a conclusão lógica de que a verbosidade se aplica às construções das línguas, e não às línguas em si. Nesta dissertação, portanto, falaremos de construções mais verbosas que outras, e morfemas que

participam de construções mais verbosas que outros, conforme veremos no capítulo seguinte<sup>24</sup>.

### 3.3. As complexificações de Trudgill

Por fim, Peter Trudgill (2011), num capítulo de *The SAGE Handbook of Sociolinguistics*, elenca quatro processos de complexificação pelos quais uma língua pode passar decorrentes de situações de baixo contato e relativo isolamento de seus falantes. Os quatro processos descritos são similares aos descritos por Booij (1993, 1996), McWhorter (2007) e Dahl (2004), e consistem em 1) desenvolvimento de categorias morfológicas, similar à sobrespecificação, pois a língua passa a codificar por meio de morfologia certos significados antes não expressos morfológicamente; 2) aumento na redundância sintagmática, o que pode envolver desenvolvimento de morfologia flexional contextual e aumentar a verbosidade de construções; 3) aumento na opacidade morfológica, envolvido no surgimento de alomorfia e polixponência; e 4) irregularização, como abordado em McWhorter (2007).

Nesta dissertação, pensando explicitamente um dos processos de complexificação trazidos por Trudgill, o aumento na opacidade morfológica, utilizamos como um dos parâmetros de complexidade o conceito introduzido por Balthasar Bickel e Johanna Nichols (2007) de *polixponência*, que diz respeito à veiculação de mais de uma categoria gramatical por um mesmo morfema. Determinamos, para cada língua, a proporção de morfemas polixponentes – uma vez que a opacidade morfológica representa desvio do *princípio da transparência* como enunciado por Kusters (2003, p. 21) – “a relação entre forma e significado é a mais transparente possível [...] cada significado único é expresso numa forma separada”. Veja-se o exemplo de três morfemas do domari (tabela 7):

<b>morfema</b>	<b>freq</b>	<b>lingua</b>	<b>POLI</b>
PRG	51	rmt	N
OBL.F	50	rmt	S
PRED.SG	49	rmt	S

Tabela 7. Classificação dos morfemas por polixponência (POLI); exemplos do domari.

O morfema PRG, veiculador apenas do significado de “aspecto progressivo”, recebeu valor negativo (N) na coluna referente à polixponência. Já os morfemas seguintes, OBL.F

<sup>24</sup> Por envolverem exemplificação e descrição metodológica mais extensa, as medidas de sobrespecificação – flexão contextual e verbosidade – serão abordadas num capítulo à parte, o 4.

(veiculador das categorias caso: oblíquo e gênero: feminino) e PRED.SG (veiculador do significado de predicador e da categoria número: singular), são polixponentes, recebendo o valor positivo (S).

Compreende-se que a polixponência contribui para o aumento da complexidade do sistema à medida em que causa maior opacidade morfológica, reduzindo a correspondência um-a-um entre forma e significado (TRUDGILL, 2011). Tradicionalmente, a presença de morfemas polixponentes foi usada para descrever o perfil *fusional* na tipologia holística, como as línguas indoeuropeias (SCHLEGEL, 1818; SCHWEGLER, 1990). Muito embora línguas dessa família apresentem grande número de morfemas polixponentes (DRYER; HASPELMATH, 2013), tal fenômeno também ocorre em línguas de diversas outras famílias – e na nossa análise, as que apresentaram maior grau de polixponência foram não-indoeuropeias, como o exemplo do saami de Pite (26,7%), no qual o morfema *muvne* veicula as noções de primeira pessoa, número singular e caso inessivo:

19. Saami de Pite (Suécia e Noruega: fino-úgrica)

<i>muvne</i>	<i>lä</i>	<i>bijjla</i>
1SG.INESS	ser.3SG.PRS	carro.NOM.SG

“Eu tenho um carro.” (WILBUR, 2014)

Assim como os parâmetros anteriores, a polixponência apresentou correlação negativa com a proporção de morfologia gramatical livre (figura 7).

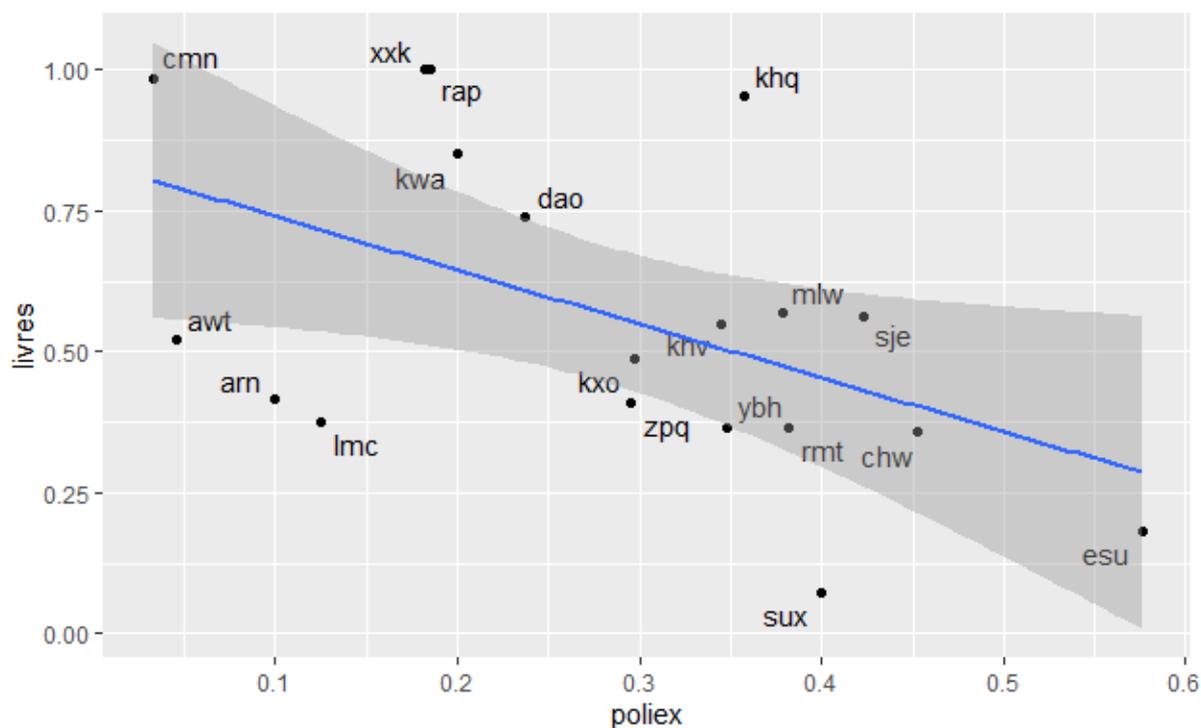


Figura 7. Relação entre proporção de morfemas livres e proporção de morfemas veiculando mais de uma categoria gramatical (poliex).

Línguas do grupo GML – mandarim (3%), rapa nui (18%), kéó (18%), dêw (20%) e daai chin (23%) –, juntamente com o araweté (4%), limilngan (12%) e mapudungun (10%), apresentaram baixa proporção de morfemas gramaticais poliexponentes. No entanto, uma língua GML, koyra chiini, apresentou proporção consideravelmente maior (35%), esta última comparável aos níveis do moloko (37%), língua afroasiática, e do domari, uma língua indoeuropeia<sup>25</sup> (38%). As línguas que apresentaram maior proporção de poliexponência foram, em ordem decrescente, o yupik (57%), o cuwabo (45%), o saami de Pite (42%) e o sumério (40%).

<sup>25</sup> É bem verdade que o domari é uma língua de contato, minoritária e sem *status* de língua oficial em todos os lugares em que é falada, e com forte intrusão léxico-gramatical do árabe; outras línguas indoeuropeias, caso amostradas, podem apresentar outras cifras de poliexponência; o uso delas em pesquisas subsequentes é algo a ser considerado.

### 3. MEDIDAS DE SOBRESPECIFICAÇÃO E ELABORAÇÃO ESTRUTURAL

Uma vez definido o conceito de complexidade com o qual esta dissertação trabalha – a complexidade absoluta, que não é a dificuldade de uma língua, mas sim o número de peças de seu sistema –, bem como expostas a taxonomia da complexidade de McWhorter (2007), a medida de complexidade fenogramatical de Dahl (2004) e as complexificações de Trudgill (2011), agora descrevemos o conceito de flexão contextual *sensu* Booij (1993, 1996) e a verbosidade *sensu* Dahl como medidas de sobrespecificação, logo, de complexidade absoluta das línguas, e como tais parâmetros foram medidos nas 19 línguas usadas no nosso estudo.

#### 4.1. Flexão contextual

No *Yearbook of Morphology* de 1993, num capítulo intitulado *Against split morphology*, Geert Booij aborda a distinção entre flexão e derivação para a morfologia – ponto de vista conhecido como teoria da morfologia cindida (*split morphology* em inglês). Inicialmente citando Stephen Anderson (1982), para quem a morfologia flexional seria aquela “relevante à sintaxe”, Booij argumenta contra a dicotomia flexão/derivação em duas categorias discretas e apartadas uma da outra; seu principal ponto é o fato de algumas formas de flexão serem capazes de alimentar processos formadores de palavras – logo, seriam mais próximas da morfologia tida por “derivacional” que os demais componentes flexionais.

O próprio Anderson, em 1988, havia oferecido uma classificação das categorias flexionais em quatro tipos: (i) categorias configuracionais (como marcação de caso em substantivos), atribuídas com base na estrutura maior na qual a palavra ocorre; (ii) categorias de concordância (aspectos da forma de uma palavra determinados por outras palavras na mesma estrutura); (iii) categorias sintagmáticas (como marcação de caso num sintagma nominal realizada no núcleo nominal); e (iv) categorias inerentes (como número em substantivos). Booij (1993) alude a esta análise e elabora sua própria classificação; para o autor, as categorias (i-iii) são *flexão contextual* e a categoria (iv), *flexão inerente*. A morfologia flexional inerente seria a forma de flexão mais próxima do que, considerando a morfologia cindida, se chama de *derivação*; sob esta categoria, estão propriedades inerentes tais qual número, para substantivos, e tempo/aspecto, para verbos. Em suas palavras:

Há uma clara diferença em conteúdo semântico entre flexão inerente e flexão contextual. Flexão inerente expressa, assim como a derivação [expressa], uma certa quantidade de informação independente, enquanto a informação expressa por flexão

contextual é redundante, e apenas reflete certos aspectos da estrutura sintática da sentença. Por exemplo, marcação de um substantivo como plural não é previsível com base na estrutura sintática, enquanto a marcação de plural do verbo finito o é, uma vez o sujeito identificado. (Booij, 1993, p. 30, tradução nossa)

Muito embora Booij tenha se concentrado no holandês na primeira publicação sobre o tópico, outras línguas serviram de exemplo e argumentação em seus subsequentes trabalhos (1996), bem como de diversos outros autores que adotaram ou trabalharam com a distinção flexão inerente *versus* flexão contextual (KIHM, 2003; LUÍS, 2010; HASPELMATH; SIMS, 2013; MCWHORTER, 2016, 2019; BISANG, 2020; GARDANI, 2014, 2018, 2020). A flexão inerente, portanto, compreende a realização de categorias que, embora possam ser relevantes à sintaxe, veiculam informação independente, como os já citados número em substantivos e tempo/aspecto verbal, mas também a exemplo de casos gramaticais portadores de informação semântica, não ditada pelo contexto sintático. Servem de exemplo o locativo (ex.: turco *ev-de* [casa-LOC] "na casa"), ablativo (ex.: quéchua de Huallaga *mayu-pita* [rio-ABL] "[vindo] do rio") e instrumental (ex.: russo *nož-om* [faca-INS] "com uma faca"). Já a flexão contextual consistiria em categorias dependentes do contexto sintático, a exemplo dos casos estruturais – como nominativo e acusativo –, e demais morfemas requeridos em regência ou concordância, exprimindo informação em grande parte redundante (HASPELMATH e SIMS, 2013). Nas palavras de Anna Kibort:

A especificação de uma característica inerente pode ser pensada como exprimindo uma informação que logicamente pertence ao – ou surge a partir do – interior do elemento no qual ela se encontra, enquanto a especificação de uma característica contextual pode ser pensada como exprimindo uma informação que se origina logicamente fora do elemento na qual ela se encontra (na concordância, chamamos essa informação de “deslocada” [*displaced*], e na regência, de “marcação” [*brand mark*]). Então, características encontradas em controladores de concordância são características inerentes, enquanto características encontradas em alvos de concordância e em [elementos] regidos são contextuais. (2008, tradução nossa)

Como ilustração, veja-se:

#### 20. Yakkha (Nepal: sino-tibetana)

ka	kheʔ-ma	mit-a-ŋ-na
1SG	ir-INF	pensar-PST-1SG-NMLZ.SG

“Eu quero ir.”(SCHACKOW, 2015)

Em 20., o morfema *-ŋ*, realizador de concordância verbal de primeira pessoa do singular, é um exemplo de flexão contextual, por exprimir uma informação que se origina logicamente fora do elemento – o verbo – em que ele se encontra; no caso, a origem é o constituinte *ka*, “eu”.

Além dos critérios apresentados, a morfologia inerente se aproxima da derivação quanto à emprestabilidade de seus morfemas veiculadores; Gardani (2014, 2018, 2020) demonstrou que formativos realizadores de flexão inerente estão muito mais sujeitos a serem emprestados de/para outras línguas que os de flexão contextual. Como exemplo, traz o empréstimo do morfema de plural *-s* do espanhol no quéchua boliviano meridional (*algu-s* [cachorro-PL] “cachorros”), em comparação às demais línguas quéchua (ex.: quéchua de Cajamarca-Cañaris *allqo-kuna* [cachorro-PL] “cachorros”), que preservaram o morfema plural nativo (GARDANI, 2018).

#### 4.2. Quantificação da morfologia flexional contextual

Cada morfema amostrado das 19 línguas utilizadas foi classificado de acordo com sua determinação ou não pelo contexto sintático – isto é, se fazia parte da morfologia flexional contextual ou não. Duas colunas foram preenchidas; a primeira, FC, diz respeito à flexão contextual *sensu* Booij, mais restrita, por levar em consideração o conceito de *palavra*, logo, apenas os morfemas que são afixos de raízes lexicais. A segunda coluna, CS, foi baseada num conceito mais amplo, considerando como flexão a realização no plano morfológico de uma categoria gramatical – entrando, portanto, nessa definição, morfemas grafados como *partículas*, *clíticos*, *palavras independentes* e também afixos, uma vez que nela se desconsidera o conceito de palavra. Como exemplo, mostramos alguns morfemas do daai chin na tabela 8:

<b>morfema</b>	<b>freq</b>	<b>língua</b>	<b>FC</b>	<b>CS</b>
NON.FUT	414	daaichin	N	N
DEM	314	daaichin	N	N
ERG	252	daaichin	N	S
S.AGR:3S	174	daaichin	N	S
PL	161	daaichin	N	S
CF	138	daaichin	N	S
LOC	130	daaichin	N	N

Tabela 8. Classificação de morfemas quanto a serem veiculadores de flexão contextual *sensu* Booiij (FC) ou serem veiculadores de flexão contextual encarando flexão como realização de conteúdo gramatical (CS).

Vemos que, sob a acepção mais estrita, nenhum dos morfemas apresentados acima seria considerado flexional contextual, uma vez que todos são trazidos pela gramática (SO-HARTMANN, 2009) como *partículas* ou *palavras independentes*. No entanto, a acepção que adotamos nesta dissertação, desconsiderando a *palavra*, revela mudanças para os morfemas da língua em questão<sup>26</sup>; os morfemas caso ergativo ERG, concordância de sujeito de 3ª pessoa S.AGR:3S, plural verbal PL e final de constituinte CF, veiculadores de informação flexional contextual, receberam o valor S na coluna CS.

Confrontamos, portanto, os valores de flexão contextual *stricto sensu*, doravante FC, e *lato sensu*, doravante CS. A flexão contextual *stricto sensu* segue a definição clássica de Booiij, isto é, com um entendimento de flexão como processos que ocorrem por meio de afixos ligados a palavras. Já a flexão contextual *lato sensu* considera todo e qualquer morfema que veicule informação gramatical determinada pelo contexto sintático – a definição original –, mas independentemente de serem afixos ou morfemas livres (*partículas*). Esta acepção segue entendimentos como o de Alain Kihm (2003, p. 335), que afirma que “[m]orfologia flexional é a realização explícita de elementos funcionais”, não necessitando se manifestar como morfologia presa.

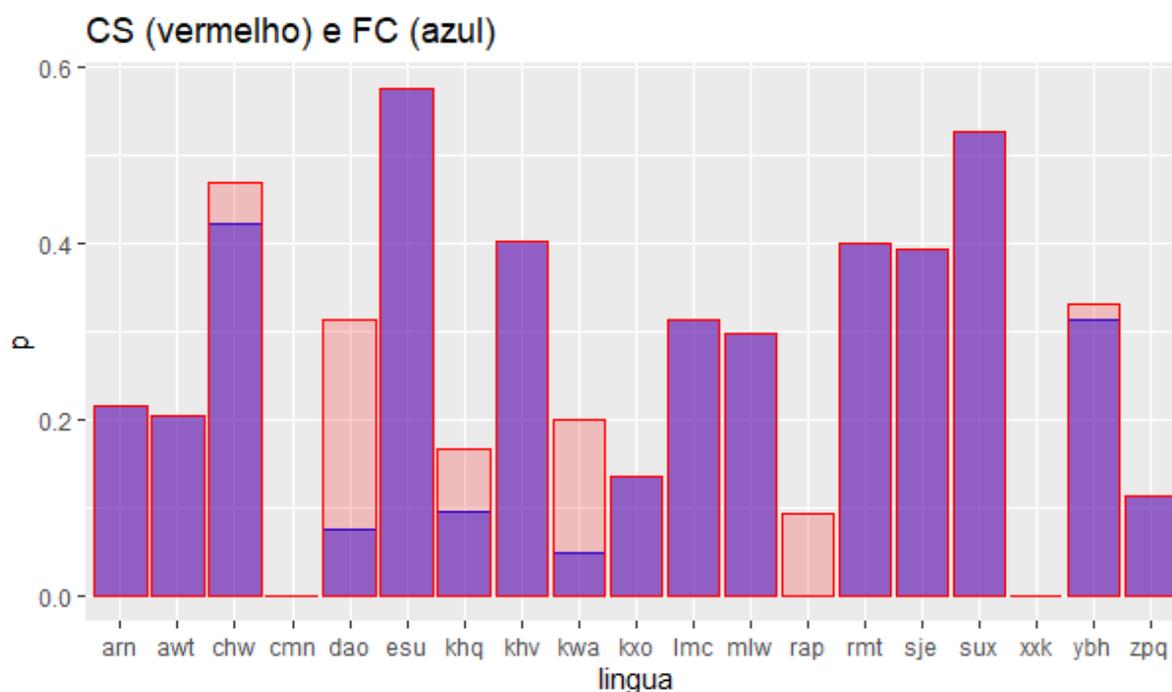


Figura 8. Comparação entre flexão contextual *sensu stricto* (FC), em azul, e *sensu lato* (CS), em vermelho.

<sup>26</sup> E de fato o daai chin foi a língua que mais apresentou mudança de resultados entre as duas acepções, FC e CS.

A figura 8 mostra a comparação gráfica dos dois parâmetros. CS revela-se, não surpreendentemente, mais inclusivo; todas as línguas que apresentam morfemas veiculadores de FC também apresentam CS, embora o contrário não seja verdadeiro – o rapa nui apresenta flexão contextual apenas na medida em que a consideramos em sentido amplo. Caso limitássemos flexão a processos que ocorrem em afixos, o rapa nui não teria flexão contextual, por exemplo. Com a exceção do cuwabo (chw), que já apresentava uma proporção significativa de morfologia flexional contextual mesmo sob a acepção mais estrita (passou de 42% a 46,8%), todas as línguas que obtiveram aumento nas taxas de morfemas veiculadores de tal morfologia ao se adotar a acepção ampliada pertencem ao grupo GML – o daai chin (dao, de 7,5% a 31%), o koyra chiini (khq, de 9% a 16,6%), o dâw (kwa, de 5% a 20%) e o rapa nui (rap, de 0% a 9%).

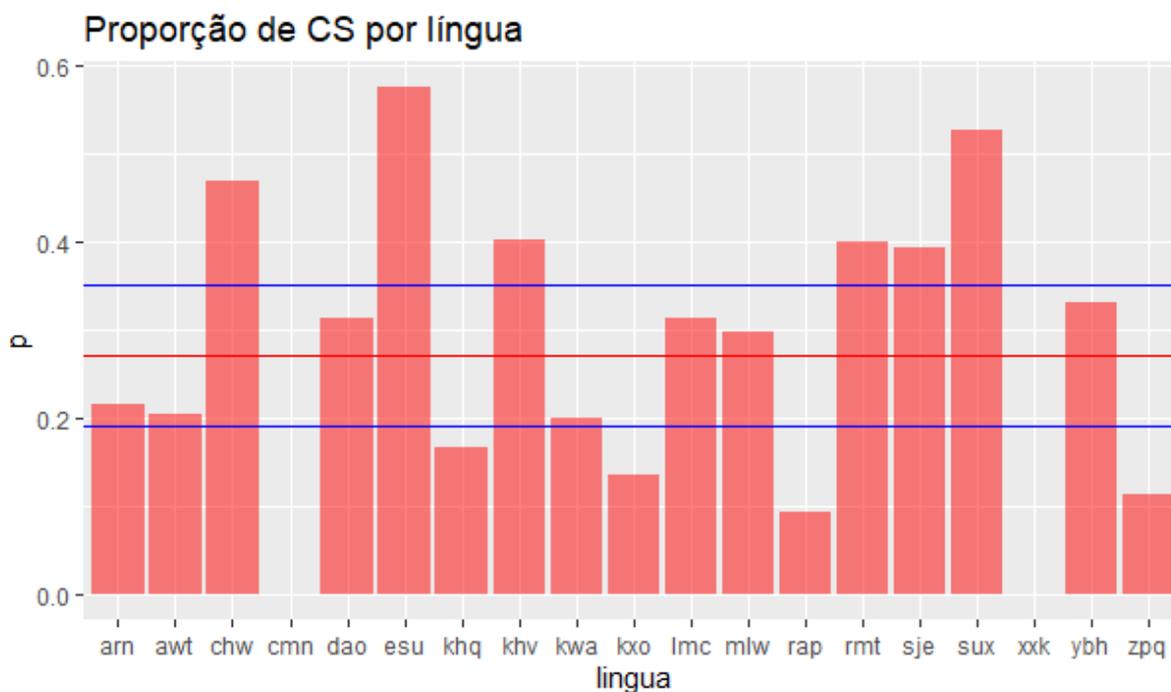


Figura 9. Proporção de morfemas veiculadores de informação gramatical determinada sintaticamente, sem considerar palavra como item de análise (CS). Em vermelho, a média da proporção; em azul, o intervalo de confiança.

Da análise da figura 9, depreende-se que dez das 19 línguas apresentaram proporção estatisticamente significativa de morfologia flexional contextual – o cuwabo (chw), o daai chin (dao), o yupik (esu), o khwarshi (khv), o limilngan (lmc), o moloko (mlw), o domari (rmt), o saami de Pite (sje), o sumério (suc) e o yakkha (ybh). Das demais nove línguas, cinco

são do grupo GLM – koyra chiini (khq), mandarim (cmn), kéó (xxk) e dâw (kwa) –, que apresentaram alta proporção de morfemas gramaticais livres, bem como baixos índices de complexidade no geral; o comportamento inusitado dessas línguas, bem como explicações com base em possíveis eventos de contato e aquisição não-nativa em suas histórias são abordados no capítulo 5. As outras quatro, mapudungun (arn), araweté (awt), kanoê (kxo) e zapoteca (zpq), são línguas indígenas das Américas; muito embora não tenham apresentado proporção estatisticamente significativa para esse parâmetro, obtiveram níveis consideráveis em outras medidas de complexidade aqui adotada – como o kanoê, com alto grau de complexidade fenogramatical-verbosidade e poliexponência, e o zapoteca, com altos níveis de elaboração estrutural. Para os propósitos desta dissertação, consideraremos flexão contextual como o parâmetro CS.

Duas das línguas GML, mandarim e kéó, não apresentam morfemas veiculadores de flexão contextual sob nenhum dos dois parâmetros. Tais línguas são descritas por McWhorter (2016, 2019) como tendo gramáticas *telegráficas*, e como tendo sofrido processos de aquisição não-nativa em larga escala durante suas histórias<sup>27</sup>, resultando em gramáticas mais *simplificadas* em alguns aspectos – a ausência de flexão contextual sendo um deles.

#### 21. Mandarim

nǐ	xiǎng	nà	yǒu	chéngxiào	ma
2	pensar	aquilo	ter	efeito	QUE

“Você acha que isso terá efeito?”

Das demais línguas GML, o rapa nui apresenta pouca morfologia contextual, e apenas ao se aplicar a acepção mais inclusiva. Koyra chiini e dâw, mesmo após a acepção mais inclusiva de flexão contextual, apresentam igualmente baixa proporção desses morfemas. Já o daai chin, porém, ao levarmos em consideração o conceito *sensu lato* de flexão, passa a ter uma proporção consideravelmente alta de morfologia flexional contextual, o que a diferencia das demais línguas GML.

Comparem-se agora o parâmetro CS com a proporção de morfemas gramaticais livres das línguas, conforme demonstrado na figura 10:

---

<sup>27</sup> É bem verdade que, segundo McWhorter (2001, 2007, 2016, 2019), línguas descritas como *isolantes* e com baixa complexidade absoluta são suspeitas de terem sofrido aquisição adulta em larga escala em algum momento de suas histórias – pode ser esse também o caso das demais línguas GML, como será debatido à frente.

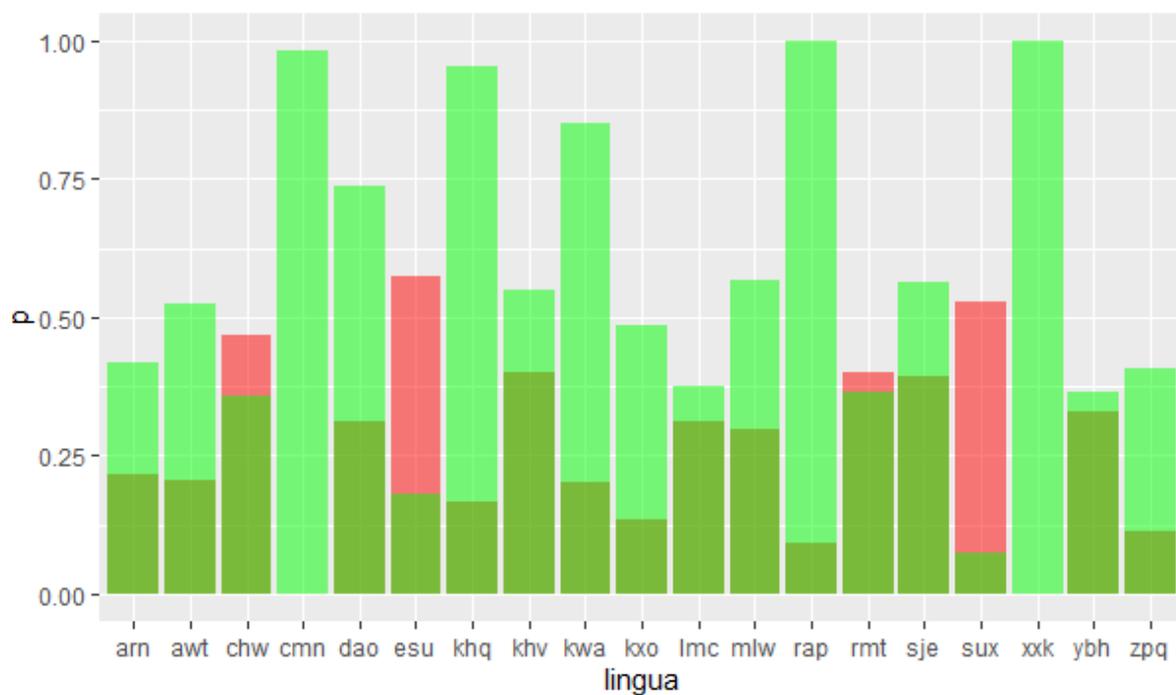


Figura 10. Proporções de CS (vermelho) e morfemas livres (verde) plotadas uma sobre a outra para cada uma das línguas.

A correlação entre morfologia gramatical livre e morfologia gramatical dependente de contexto sintático também pode ser visualizada na forma de um gráfico de reta (figura 11):

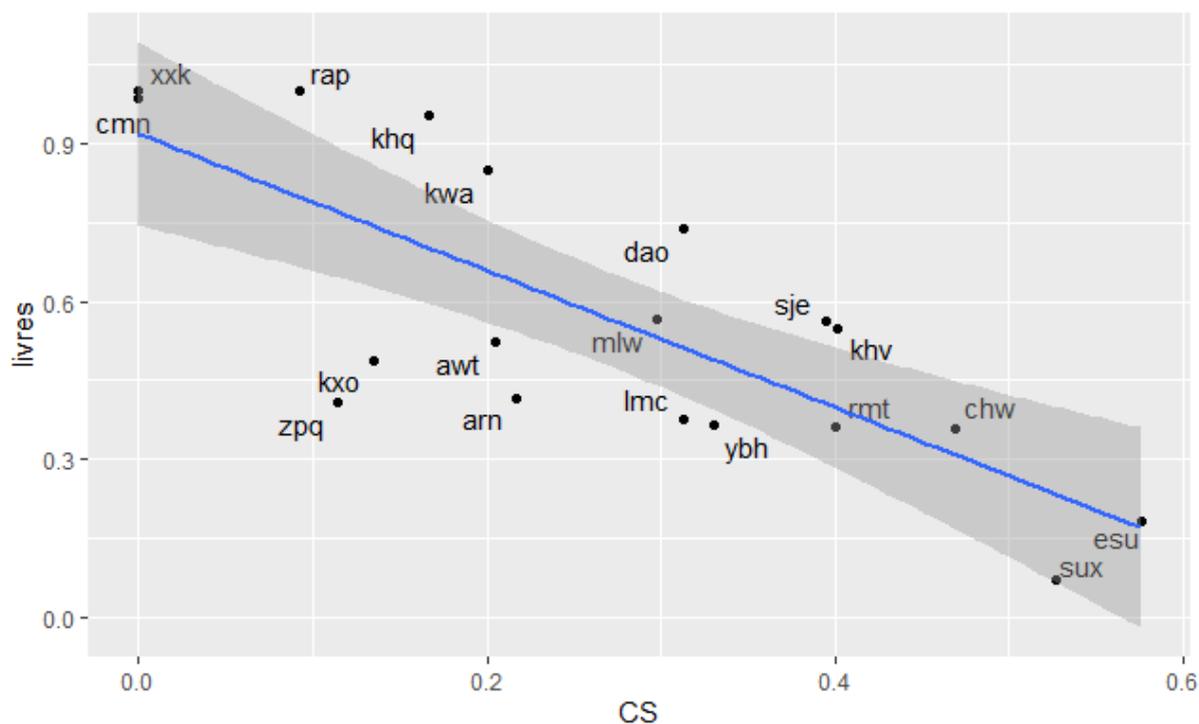


Figura 11. Relação entre proporção de morfemas livres e proporção de morfemas veiculadores de flexão contextual (CS)

Podemos ver, portanto, que há uma correlação negativa entre a proporção de morfemas livres no inventário de morfemas gramaticais de uma língua e a proporção de morfemas gramaticais veiculadores de informação flexional contextual. As duas línguas que mais apresentaram morfemas livres, mandarim (cmn) e kéó (xxk), foram as que menos apresentaram flexão contextual, por exemplo. Já as duas mais caracterizadas por terem alta proporção de morfologia flexional contextual, yupik (esu, 57,5%) e sumério (sux, 52,7%) também apresentaram menor proporção de morfologia gramatical livre – isto é, maior proporção de morfologia gramatical presa. O yupik é descrito como sendo uma língua *polissintética* (MIYAOKA, 2012):

22. Yupik do Alasca central (Alasca: esquimó-aleúta)

yu-urte-llr-an	allami-ku-ani
peessoa-virar-NMLZ-REL.3SG.SG	ano.passado-ASP-LOC.3SG.SG

yu-urte-llru-unga  
 peessoa-virar-PST-IND.1SG

“Eu nasci um ano depois de ele ter nascido.” (*ibid.*)

Já o sumério, tradicionalmente descrito como *aglutinante*, é alvo de uma lúcida discussão sobre a validade da classificação holística por critérios morfológicos em sua gramática (JAGERSMA, 2010). Apesar disso, o trabalho termina por descrever a morfologia do sumério como *aglutinante*. Importante ressaltar que o sumério, diferentemente das demais línguas, está extinto desde 1800 antes da era comum, e teve como fonte de morfemas uma gramática que, mesmo moderna e declaradamente descritiva, lança mão de dados obtidos de antigas inscrições cuneiformes, registrando um uso da língua que pode ter sido diferente do utilizado diuturnamente pela população – muito embora o autor analise mais de mil anos de inscrições, computando em seu trabalho inclusive variações geográficas e diacrônicas no uso de certos morfemas.

### 23. Sumério (Mesopotâmia: isolada)

ud inannak-e igi namtil-ak-ane-Ø  
 dia Inanna-ERG olho vida-GEN-SEU-ABS

Ø-mu-n-ši-n-bar-Ø-'a-'a

VP-VENT-3SG-para-3SG.A-estar.fora-3N.S/DO-NOM-LOC

“Quando Inanna o olhara com seus olhos provedores de vida.” (*ibid.*)

Outras duas línguas que combinam alta proporção de morfemas veiculadores de flexão contextual com baixa proporção de morfemas livres são o cuwabo, língua bantu que como os demais membros da família, é descrita como *aglutinante* e dispõe de um extenso sistema de classes nominais, e o domari (40% de CS), língua indoeuropeia.

### 4.3. Complexidade fenogramatical

No terceiro capítulo, abordamos o conceito de verbosidade, e como este conceito diz respeito a quanto uma construção desvia do mínimo necessário para exprimir uma ideia em termos de material fonético adicionado; isto é, o quanto de complexidade fenogramatical está presente e é dispensável de um ponto de vista comparativo. Com o fito de medir o grau de verbosidade das construções nas quais os morfemas gramaticais das línguas tomam parte, foi confeccionada uma tabela (Tabela 9) com uma gradação de complexidade fenogramatical para cinco grandes categorias gramaticais: tempo-modo-aspecto (doravante, TAM), número-pessoa

ou gênero (doravante, NPG), posse, sintagmas adposicionais e papel sintático (mormente sujeito/objeto).

grau	tempo-modo-aspecto	número-pessoa e gênero	posse	sintagmas adposicionais	papel sintático
<b>0</b>	NP verbo NP	verbo	-	-	-
<b>1</b>	NP + advérbio + verbo	pessoa + verbo	justaposição	adposição + NP	ordem dos elementos
<b>2</b>	NP + advérbio + verbo-TAM	pessoa + verbo-pessoa	N1 + POSS + N2	adposição NP-caso	NP1-NOM + verbo + NP2
<b>3</b>	NP + advérbio + verbo1-TAM1 + verbo2-TAM2	pessoa + verbo-pessoa + AUX-pessoa	N1 + POSS + N2-pessoa	-	NP1-NOM + verbo-pessoa + NP2

Tabela 9. Gradação de complexidade fenogramatical

O grau zero refere-se à ausência de marcação das cinco categorias; isto é, não se marca tempo-modo-aspecto, nem número-pessoa/gênero, posse, ou papel sintático, tampouco há adposições. Não entra na presente análise pela simples ausência de material a ser classificado, servindo apenas como marco zero para os demais graus.

#### 4.3.1. Grau 1

Em se tratando do grau 1, já há observações a serem feitas. Este é o primeiro grau de desvio da concatenação irrestrita, podendo ser marcado por presença de material morfológico em algumas categorias (TAM, NPG, sintagmas adposicionais), ou por justaposição e ordem dos elementos para outras (posse, papel sintático).

Para TAM, o exemplo é de uma sentença que contenha apenas um advérbio de tempo como morfema gramatical daquela categoria. Veja-se:

24. Kéo (Ilha de Flores: austronésia)

kau négha ti'i 'imu joi o  
 2sg já dar 3sg dinheiro que

“Você deu o dinheiro a ele?” (BAIRD, 2002)

A mesma sentença em 24. também serve de exemplo para o grau 1 na categoria número-pessoa e gênero; veja-se que para veicular o significado de segunda pessoa do singular, há apenas um morfema – *kau*, sem qualquer espécie de concordância ou repetição deste significado. Perceba-se que sentenças como a seguinte em 25. contam também como sendo de primeiro grau para número-pessoa:

25. Mandarin

tā	shuō	tā	zhǎo-dào	le	tā
3SGi	falar	3SGi	procurar-RES	PFV	3SGj

“Elei falou que elei encontrou elej.”

Isto se dá pelo fato de que em cada oração, e para cada entidade extralinguística à qual o morfema de número-pessoa (o pronome) se refere, o morfema aparece uma única vez, mesmo que no período completo 3SGi apareça duas vezes (por se tratar de um período composto, não sendo um caso de redundância ou repetição do mesmo material na oração). Já 3SGj, muito embora seja igual a 3SGi, refere-se a outra entidade, a outro “ele” de quem se fala. Trata-se, portanto, não de redundância, mas do objeto, enquanto 3SGi se referia ao sujeito.

Na categoria posse, temos como exemplo do grau 1 a sentença trazida em 26., na qual vê-se apenas justaposição do possuidor com o possuído, nesta língua, necessariamente nesta ordem:

26. Dâw

nũx	ʔãay	ʔãx	paar	beey
curupira	fêmea	pedir	saber	ITER

"A mulher do curupira perguntou de novo." (CARVALHO, 2016)

Para a categoria sintagmas adposicionais, um exemplo do grau 1 é:

27. Kéo

koti	'imu	pu'u	pusu	muku	ta	ti'i	muri	pati	'imu
pião	3SG	de	coração	banana	REL	dar	vida	dar	3SG

“O pião dele era feito do coração de bananeira que lhe tinha dado vida.” (BAIRD, 2002)

Neste grau, a adposição *pu'u* acompanha o substantivo *pusu muku* sem reger caso algum, e ocorrendo apenas uma instância do significado por ela veiculado.

Por fim, no primeiro grau, a categoria de papel sintático se dá por meio da ordem dos elementos. Tomemos a sentença do kéo vista em 24.; o sujeito, *kau* (segunda pessoa do singular), ocorre anteposto ao verbo. O objeto, ‘imu, posposto a ele.

#### 4.3.2. Grau 2

O segundo grau de complexidade fenogramatical, para TAM, dá-se quando há um marcador de tempo, aspecto ou modo obrigatório juntamente ao advérbio, locução adverbial ou oração adverbial que já aparecia desde o grau primeiro. Veja-se:

28. Domari (Oriente Médio: indoeuropeia)

a                    zamān l-'urdunn      kil-d-om      baladiy-ē-ma  
durante          tempo DEF-Jordânia    sair-PST-1SG    município-OBL.F-LOC

“Durante o governo jordão, eu me juntei ao município.” (MATRAS, 2012)

Nesta sentença, há uma oração que serve de advérbio temporal (*a zamān l'urdunn*), e uma marcação de tempo que se revela dispensável em muitas línguas, mas que o domari obrigatoriamente expressa, o morfema de passado. Esta construção, portanto, é mais verbosa que a em 24. quanto à categoria TAM.

Para a categoria 2, NPG, também usamos de exemplo uma sentença do domari:

29. Domari

dōmiy-ēni                    našī-r-e  
mulher dom-PRED.PL    dançar-PST-3PL

“Mulheres dom dançaram.” (MATRAS, 2012)

Vê-se em 29. a primeira instância do fenômeno da concordância, no caso, numeropessoal. O NP *dōmiyēni* refere-se a uma entidade de que se fala, isto é, a terceira pessoa; o morfema de terceira pessoa do plural no verbo é dispensável em muitas línguas, mas obrigatório no domari.

A sentença *koyra chiini* em 30. serve de exemplo para o segundo grau de complexidade fenogramatical quanto à categoria posse:

30. *Koyra chiini*

a wane gaabi di

3SG POSS força DEF

“A força dele.” (HEATH, 2011)

Compare-se tal sentença com a sentença *dâw* presente em 26. Em 26., o único desvio de um estado de concatenação irrestrita é a ordem dos constituintes (possuído-possuidor); já em 28., além da ordem, temos a presença de material morfológico obrigatório – *wane*, indicando posse.

Para o segundo grau, a categoria sintagma adposicional também pode ser ilustrada por uma sentença do domari:

31. Domari

qrara aha šōna mang-id-a ihi domi-yē min bōy-is-ki

beduíno este.Mgaroto pedir-PST-M este.F dom-OBL.F de pai-3SG.OBL-ABL

“O garoto beduíno pediu a mão da garota dom ao pai dela.” (MATRAS, 2012)

Pode-se perceber que a preposição *min* “de” rege o caso ablativo de *bōy* “pai”. Portanto, há uma dupla sinalização do conteúdo gramatical veiculado, tanto por parte da adposição, quanto pelo caso gramatical que ela obrigatoriamente rege na língua domari. Este caso é mais verboso que a mera justaposição da adposição com o NP, como na sentença *kéo* em 27.

Já o papel sintático no segundo grau se dá quando há marcação explícita de sujeito e/ou objeto, a despeito de existir ordem preferencial dos constituintes – como é o caso da maioria das línguas. No caso de a ordem dos constituintes ser verdadeiramente livre e não-marcada em qualquer de suas combinações (SVO, SOV, VSO, etc), a existência de material morfológico explicitando papel sintático de sujeito ou objeto configuraria complexidade fenogramatical de grau 1. Não é, porém, o caso em nenhuma das línguas contempladas no presente trabalho. Portanto, a junção de ordem dos constituintes e marcação explícita de caso resulta em construções de grau 2 de complexidade fenogramatical, como a

sentença rapa nui em 16., na qual se marca morfológicamente o objeto por meio da “partícula” *i* de acusativo.

#### 4.3.3. Grau 3

No tocante à categoria TAM, o terceiro (e último grau desta análise) se expressa por meio do fenômeno conhecido na literatura como *consecutio temporum*; isto é, um tempo-modo verbal numa oração subordinada, sintaticamente dependente do tempo da oração principal, caso acompanhado de um terceiro elemento que marque a subordinação.

#### 32. Yakkha

khiŋ-nuŋ      em-ma-niŋa      lak-nuŋ      leks-a-bi  
esse tanto-COM inserir-INF-CTMP      salgado-COM      tornar-se3SG-SBJ-IRR

“Se inserissem esse tanto, ele ficaria salgado.” (SCHACKOW, 2015)

Veja-se que o morfema de subjuntivo ocorre numa construção de terceiro grau de verbosidade; é sintaticamente dependente do infinitivo na oração principal e é acompanhado por um morfema de modo *irrealis*. Logo, são três morfemas juntos que contribuem para o sentido veiculado pelo morfema de subjuntivo.

Na categoria NPG, temos como exemplo a seguinte sentença da língua kanoê:

#### 33. Kanoê

aj      mo-õ-kÿj      ã-e-re      ña      munaw-ni      aj  
1SG      dormir-1-olho      1-DECL-AUX      POSS1SG      rede-OBL      1SG

“Eu durmo na minha rede.” (BACELAR, 2004)

Em 33., vemos que para expressar *eu durmo*, o kanoê utiliza três morfemas que veiculam a informação de primeira pessoa. Perceba-se que o morfema que acompanha *munawni* (*munawni aj* “minha rede”) não conta como uma quarta instância, pois seu significado contribui para fazer referência à rede (a rede da primeira pessoa do discurso), e não ao ato de dormir.

Como exemplo de uma construção de terceiro grau exprimindo posse, temos a sentença 34. do yakkha:

#### 34. Yakkha

ak-ka            paŋ-be

1SG.POSS-GEN   casa-LOC

“Na minha casa” (SCHACKOW, 2015)

Em 34., vemos que além da justaposição na ordem possuidor-possuído (primeiro grau), temos marcação tanto por meio do uso de um possessivo de primeira pessoa, *ak*, quanto pelo morfema de genitivo, *-ka*.

Finalmente, a sentença em 35. presta-se ao papel de exemplo de uma construção de terceiro grau quanto a papel sintático:

#### 35. Yupik

arna-t                    iqva-llru-ut

mulher-ABS.PL            catar frutos-PST-IND.3PL

“As mulheres catavam frutos.” (MIYAOKA, 2012)

Há desvio de um estado de concatenação irrestrita por meio da ordem dos constituintes – sujeito antes do verbo –, que configuraria o primeiro grau na escala de complexidade fenogramatical, e também por meio da marcação de caso absolutivo e concordância verbal de número-pessoa com o sujeito.

#### 4.3.4. A análise

Tendo em vista a gradação estabelecida para análise da complexidade fenogramatical-verbosidade, classificamos cada morfema individualmente de acordo com sua participação em construções de cada um dos três graus:

- Morfemas que participam em apenas construções de primeiro grau para seu significado – isto é, cujo significado é veiculado em apenas uma instância – receberam o valor S na coluna grau1, e o valor N nas demais, como o morfema *pu’u* “de” do kéo em 27., ou o *then* do cuwabo na tabela 10.
- Morfemas cujo significado é expresso tanto em construções de primeiro grau, quanto em construções de segundo grau – isto é, cujo significado pode aparecer uma vez

apenas, ou até duas – receberam S nas colunas grau1 e grau2, e N na grau3, como o morfema *-d-* PASSADO do domari em 28., ou o *LOC* do cuwabo na tabela 10.

- Morfemas cujo significado apareceu em construções dos três graus receberam S nas três colunas, grau1, grau2 e grau3, como o morfema *-õ-* de primeira pessoa do kanoê em 33., ou o *PFV.DJ* do cuwabo na tabela 10.

Caso um morfema não tenha aparecido em construções de grau mais baixo, mas apenas em graus acima, ele recebe N no grau baixo em que não apareceu, e S nos demais; é o caso do morfema *OMI* do cuwabo, que recebeu N-S-S por aparecer apenas em construções de segundo grau acima.

morfema	freq	língua	grau1	grau2	grau3	TAM	GEN	AGR	ADP	FG
PFV.DJ	71	cuwabo	S	S	S	3				
NAR	60	cuwabo	N	S	S	3				
OM1	54	cuwabo	N	S	S			3		2
IPFV.CJ	48	cuwabo	S	N	N	1				
APPL	33	cuwabo	S	N	N				1	
then	33	cuwabo	S	N	N					
LOC	31	cuwabo	S	S	N				2	

Tabela 10. Análise por graus e categorias gramaticais em tabela; exemplos do cuwabo.

Para cada morfema, os valores referentes a cada grau foram somados – 1 para o primeiro grau, 2 para o segundo e 3 para o terceiro; caso um morfema tenha recebido S nos três, conta com um valor total de 6, por exemplo. Tais valores foram somados e divididos pelo número total de morfemas por língua, avaliando dessa forma a média de complexidade fenogramatical-verbosidade para cada uma das 19 estudadas. Além disso, também para cada morfema, registramos o nível máximo de verbosidade de construções para veicular seu significado em cada categoria gramatical dentre TAM, genitivos, concordância (AGR), sintagmas adposicionais (ADP) e função gramatical (FG). Por exemplo, o morfema OM1 do cuwabo veicula a primeira classe nominal em construções de verbosidade até 3, enquanto veicula função gramatical (indicando o sujeito do verbo) em construções de até o nível 2.

Entendendo, portanto, o conceito de complexidade fenogramatical-verbosidade como *dispensabilidade translinguística*, classificamos cada morfema de acordo com o grau em que sua ocorrência faz a construção da língua desviar de um estado inicial de concatenação

irrestrita para transmitir uma mensagem. A figura 12 mostra a proporção de morfemas envolvidos em cada um dos três graus por língua:

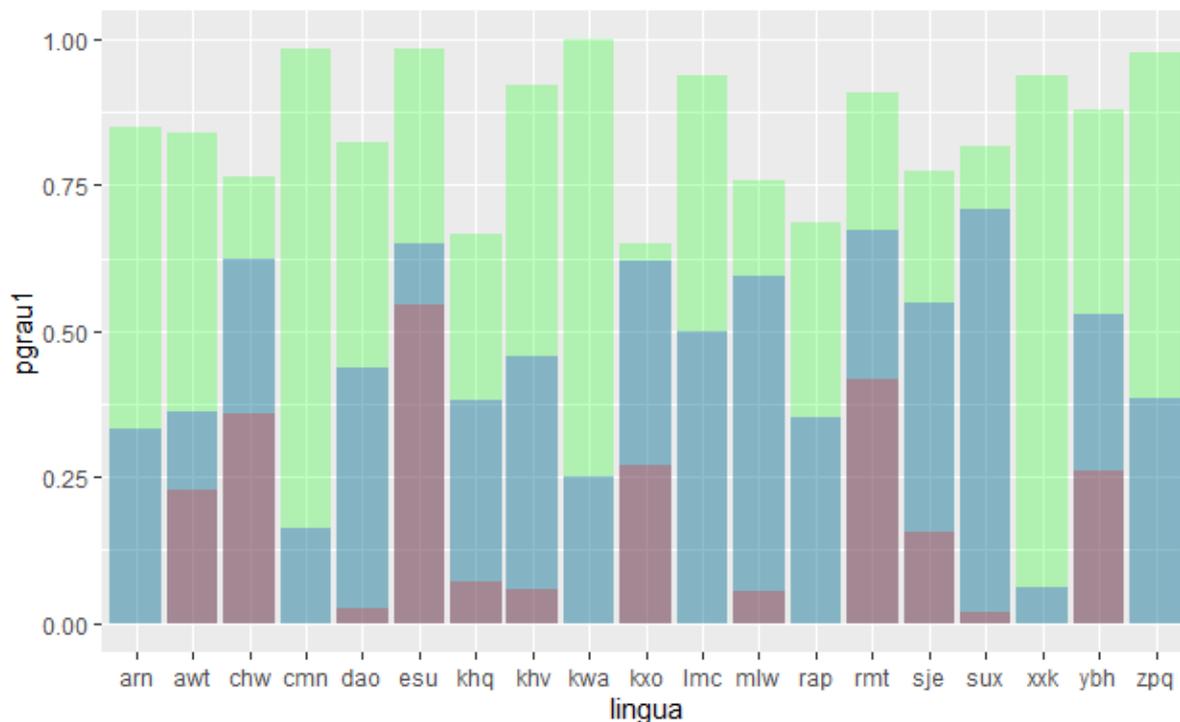


Figura 12. Proporção de morfemas presentes em construções dos três graus de complexidade fenogramatical-verbosidade nas línguas. Em verde, proporção de morfemas envolvidos em construções de primeiro grau; em azul, proporção de morfemas envolvidos em construções de segundo grau; em roxo, proporção de morfemas envolvidos em construções de terceiro grau.

Dentro do grupo GML, vemos que apenas duas línguas apresentaram morfemas de construções de terceiro grau – o daai chin (dao) e o koyra chiini (khq). As demais línguas do grupo, juntamente com três línguas não-GML – o mapudungun (arn), o limilngan (lmc) e o zapoteca (zpq) –, não apresentaram sentenças de grau superior ao segundo, logo, nenhum morfema gramatical que fosse usado em tais construções.

A proporção de morfemas gramaticais livres também revelou ter correlação negativa com a média de complexidade fenogramatical-verbosidade calculada para cada língua, conforme o gráfico na figura 13. As línguas com menores médias de complexidade fenogramatical-verbosidade, o kéo (xxk, 1,06), o mandarim (cmn, 1,3), o rapa nui (rap, 1,38) e o dâw (kwa, 1,5) são todas do grupo GML, com alta proporção de morfologia gramatical livre.

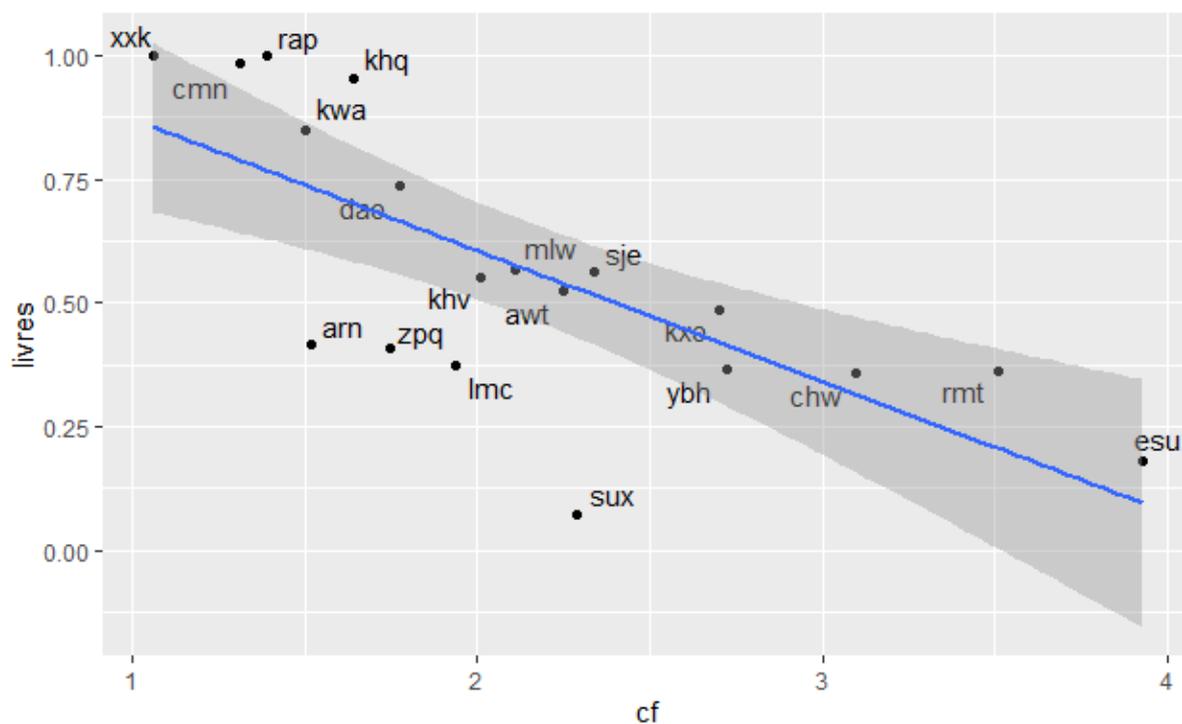


Figura 13. Relação entre proporção de morfemas livres e média de complexidade fenogramatical-verbosidade nas línguas (cf).

Percebe-se, por conseguinte, que as línguas com mais alto grau de complexidade fenogramatical-verbosidade, calculada por média, também possuem baixa proporção de morfologia gramatical livre – são o yupik (esu, 3,92), novamente, e o domari (rmt, 3,5), acompanhados do cuwabo (chw, 3,09), do yakkha (ybh, 2,72), do kanoê (kxo, 2,7) e do saami de Pite (sje, 2,33). O yakkha, língua kiranti falada nos Himalaias, por exemplo, apresenta sentenças com repetidas instâncias de um mesmo significado:

### 36. Yakkha

kanciŋ sok-khusa ca-ya-ŋ-ci-ŋ

1.DU olhar-RECIP COMER.AUX-PST-EXCL-DU-EXCL

“Nós [dual, exclusivo] olhamos um para o outro.” (SCHACKOW, 2015)

Conforme exposto no capítulo 4, a complexidade fenogramatical-verbosidade possui correlação positiva com o parâmetro CS – um tanto lógico, uma vez que flexão contextual frequentemente envolve repetição de um mesmo significado sob a forma de concordância.

### 37. Cuwabo (Moçambique: bantu)

namárógolo o-lí odhúlú wa muyére  
1a.lebre 1-ser 17.TOPO 17.CON 3.árvore.sp

ó-ni-ó-íbá wénéwo  
1-IPFV.DJ-15-cantar 17.DEM

“A lebre está no topo da árvore.sp, ela está cantando lá em cima.” (GUÉROIS, 2015)

O exemplo em 37. é do cuwabo e seu sistema de concordância por classe nominal, tão característico de sua família<sup>28</sup>.

A relação entre a proporção de morfemas flexionais contextuais e o grau de complexidade fenogramatical-verbosidade das línguas também foi investigada por meio de um gráfico de correlação. Como aduzido, cada um dos níveis foi convertido em valores numéricos (de acordo com o nível, 1, 2, 3), que foram somados para cada morfema. Enfim, para cada língua, a média de verbosidade foi tirada, somando-se os valores de seus morfemas e dividindo-os pelo número total de morfemas. A correlação pode ser vista no gráfico da figura 14:

---

<sup>28</sup> Nas glosas do cuwabo, números de 1-20 desacompanhados de número (plural ou singular) se referem às classes nominais, como os prefixos em 7.

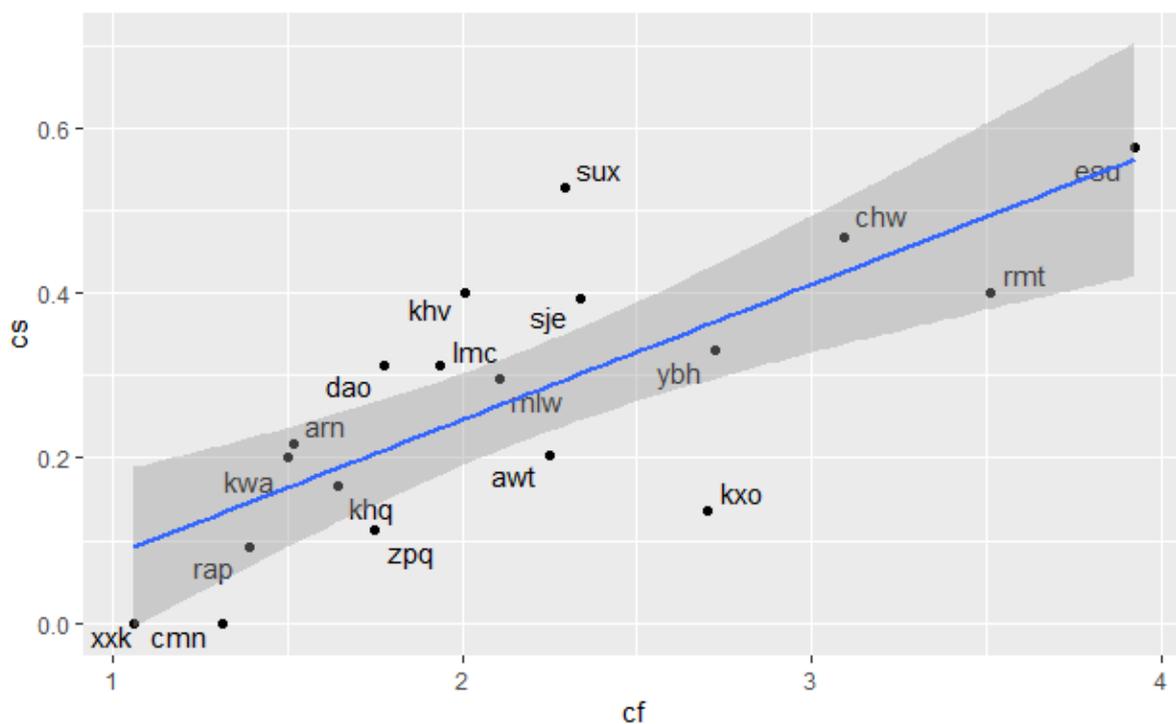


Figura 14. Correlação entre complexidade fenogramatical-verbosidade (cf) e proporção de morfemas flexionais contextuais (cs) nas línguas.

Vemos que há uma correlação positiva entre o nível de complexidade fenogramatical-verbosidade de uma língua e sua proporção de morfemas gramaticais que veiculam significados determinados pelo contexto sintático – mostrando que ambas são medidas de complexidade de uma língua, em especial o tipo de complexidade que costuma ser perdida em línguas cujas histórias contam com eventos de contato intenso e adoção em massa por adultos<sup>29</sup>. Ademais, ambos os parâmetros podem frequentemente ser vistos como o resultado de um mesmo processo, o tipo de complexificação chamado por Trudgill (2011) de *aumento na redundância sintagmática*.

<sup>29</sup> Para uma discussão acerca do tópico, ver o capítulo sexto.

## 4. IMPLICAÇÕES

Usamos um método quantificável e testável estatisticamente para mensurar a complexidade das línguas estudadas, bem como a proporção de morfologia gramatical livre, e comparar tais parâmetros. Os resultados sugerem alguns pontos de vista quanto às próprias propostas que basearam a dissertação, quanto ao estudo da complexidade e quanto à história e classificação de algumas línguas.

### 5.1. Análise de correspondência (CA) e dendrogramas

Análises de clusterização nos permitem perquirir os dados em busca de alguma estruturação no agrupamento das diferentes línguas tendo em base sua semelhança nos diferentes parâmetros e identificar possíveis *outliers*. De maneira a melhor analisar a semelhança de comportamento das línguas de acordo com seus inventários de morfemas gramaticais, os dados de cada língua foram convertidos em tabelas de contingência, às quais se aplicou a análise de correspondência (CA) do pacote FactoMineR para a linguagem R, e subsequente agrupamento por *clustering* hierárquico na forma de dendrogramas. A análise de correspondência forneceu as distâncias entre as diferentes línguas, e os dendrogramas permitiram a visualização hierarquizada das relações entre elas e a estrutura de seus agrupamentos no formato de árvore. Foram feitas três análises; na primeira, a distância entre as línguas foi mensurada de acordo com suas proporções de morfemas gramaticais livres (figura 15). Na segunda, analisamos a combinação dos quatro outros parâmetros – CS, CF, EE e polixponência (figura 16), e na terceira, a combinação dos cinco parâmetros (figura 16). Em seguida, os três dendrogramas e os agrupamentos formados foram comparados.

#### 5.1.1. Agrupamento por morfologia livre

No primeiro dendrograma, por proporção de morfologia gramatical livre, vemos que a maior parte das línguas do grupo GML se manteve unida em um clado – dâw (kwa), koyra chiini (khq), mandarim (cmn), rapa nui (rap) e kéó (xxk).

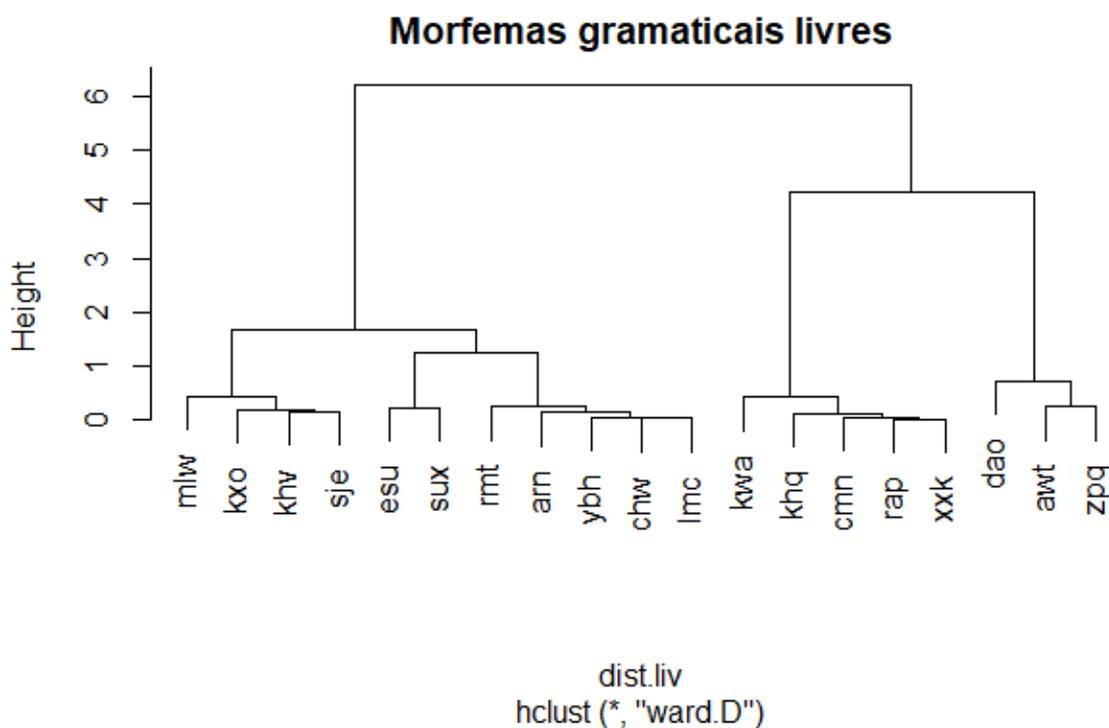


Figura 15. Agrupamento em dendrograma por *clustering* hierárquico de acordo com a proporção de morfemas gramaticais livres.

O grupo das línguas GML aparece associado a outro menor, reunindo a última língua GML, daai chin (dao), juntamente com outras duas: o araweté (awt) e o zapoteca (zpq), descritas tradicionalmente como *aglutinantes* e *polissintéticas*. As demais línguas aparecem mais próximas entre si, reunidas num terceiro grupo com dois subgrupos, um formado por moloko (mlw), kanoê (kxo), khwarshi (khv) e saami de Pite (sje), e outro com as demais; note-se que sumério (sux) e yupik (esu), dotados de baixíssima proporção de morfologia gramatical livre, aparecem extremamente próximos um do outro.

### 5.1.2. Agrupamento pelos parâmetros de complexidade

O segundo dendrograma, combinando os quatro demais parâmetros, obteve um resultado curioso; do grupo GML, quatro línguas se mantiveram próximas – dâw (kwa), mandarim (cmn), rapa nui (rap) e kéó (xxk) – mais afastado das outras línguas do estudo, caracterizado por baixa complexidade absoluta sob os parâmetros aqui analisados. O kanoê (kxo), porém, língua descrita como *polissintética* em sua gramática, apareceu associado a essas quatro línguas – embora bastante distante. O kanoê apresenta altos índices de poliexponência e complexidade fenogramatical, mas, diferentemente da outra língua descrita

como *polissintética* no estudo, o yupik (esu), apresenta baixíssima proporção de morfologia flexional contextual, o que talvez explique a associação com as línguas GML.

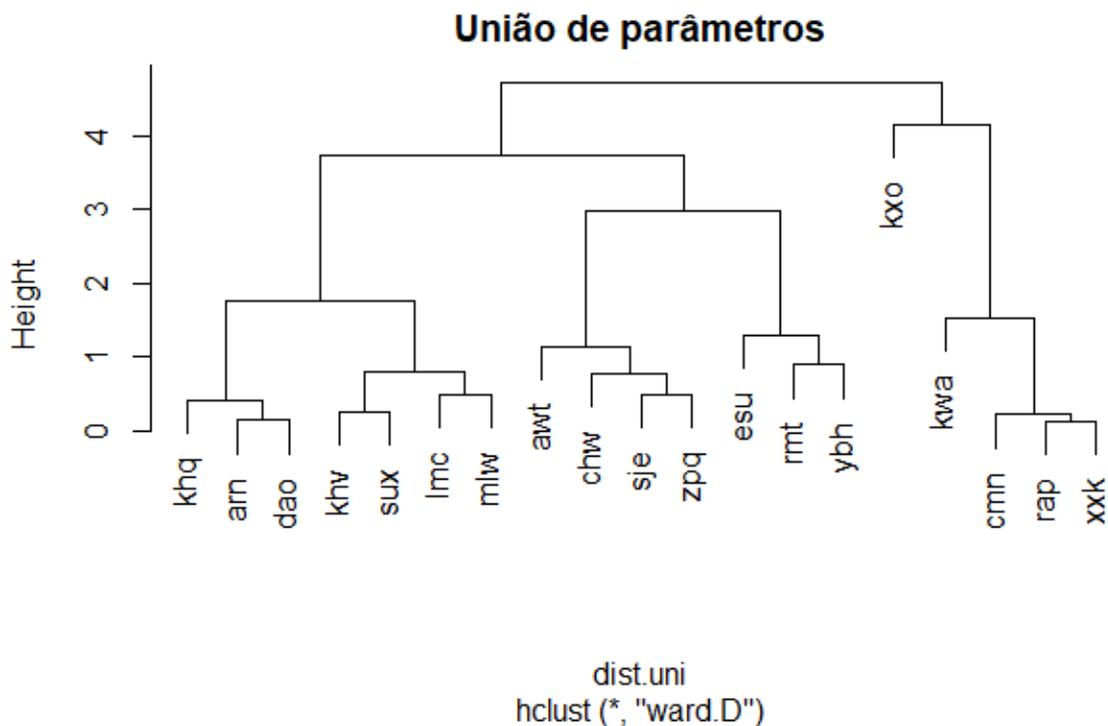


Figura 16. Agrupamento em dendrograma por *clustering* hierárquico de acordo com os quatro demais parâmetros.

Digno de nota é que o koyra chiini (khq) e o daai chin (dao), associados ao mapudungun (arn), encontram-se próximos a demais línguas com níveis maiores de complexidade, como khwarshi (khv), sumério (sux), limilngan (lmc) e moloko (mlw).

### 5.1.3. Agrupamento pela união de todos os parâmetros

Por fim, um terceiro dendrograma, combinando todos os cinco parâmetros – os quatro de complexidade e a proporção de morfemas gramaticais presa – foi elaborado (figura 17). Nele, vemos que o koyra chiini (khq) voltou a se associar às línguas GML, mas o daai chin (dao) não – este último associou-se ao araweté (awt) e ao zapoteca (zpq).

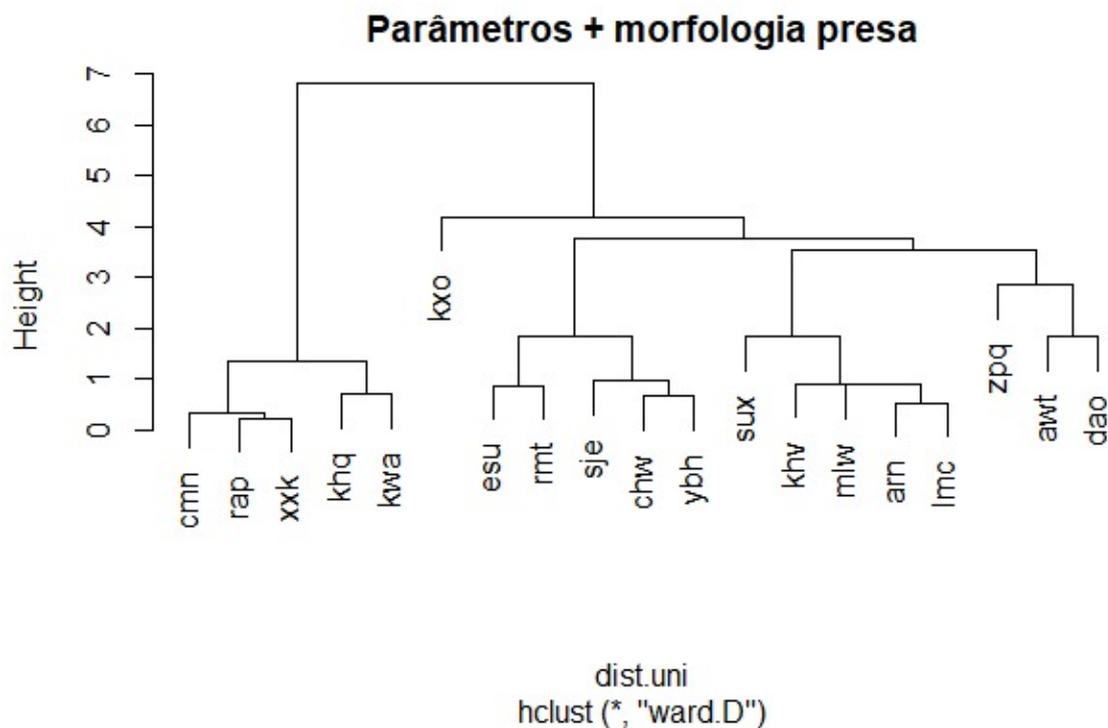


Figura 17. Agrupamento em dendrograma por *clustering* hierárquico de acordo com os quatro parâmetros de complexidade e a proporção de morfologia presa.

O kanoê (kxo) se desassociou das línguas GML; diferentemente das demais línguas nativas das Américas usadas no estudo (à exceção do yupik, que não é uma língua *ameríndia*<sup>30</sup>), apresentou proporção baixa de flexão contextual e altos níveis de complexidade fenogramatical-verbosidade e poliexponência; isso pode ser evidenciado pela quantidade de instâncias usadas para veicular a segunda pessoa do singular na sentença abaixo, por exemplo;

### 38. Kanoê

mi tsero itæ mi-e-re mi  
 2SG chicha beber 2-DECL-AUX 2SG  
 “Você bebe chicha.” (BACELAR, 2004)

Outras línguas indígenas da América, como o zapoteca (zpq), o araweté (awt) e o mapudungun (arn), a despeito de também apresentarem baixa proporção de morfologia gramatical livre, tiveram complexidade fenogramatical-verbosidade muito mais baixa – no caso do zapoteca e do mapudungun, atingindo apenas o segundo nível.

<sup>30</sup> Ver capítulo 7.

Finalmente, dois pequenos grupos de línguas com maiores níveis de complexidade foram formados – um contendo yupik (esu), domari (rmt), saami de Pite (sje), cuwabo (chw) e yakkha (ybh), e outro com sumério (sux), khwarshi (khv), moloko (mlw), mapudungun (arn) e limilngan (lmc).

## 5.2. Relação entre morfologia presa e complexidade

Ao longo dos capítulos anteriores, vimos como todos os parâmetros de complexidade absoluta adotados no presente estudo tiveram correlação negativa com a proporção de morfemas gramaticais livres de uma língua. McWhorter (2007), considerando *flexão* como processos envolvendo morfologia presa, afirma que a flexão corporifica os três tipos de complexidade por ele trazidos de tal forma que não podem ser vistas “como meras alternativas presas a morfemas alhures manifestados como formas livres” (p. 35); é dizer, línguas *flexionadas* segundo sua aceção serão mais complexas. Ele compara a mesma sentença em duas línguas, o suaíle, trazida por ele como apresentando *flexão aglutinante*, e o saramacano, considerada *isolante*:

39.

a. Suaíle (África oriental: bantu)

ni-li-taka

1SG-PST-querer

“Eu quis.” (MCWHORTER, 2007)

b. Saramacano (Suriname: crioulos de base inglesa-portuguesa)

mi bi ké

1SG PST querer

“Eu quis.” (*ibid.*)

Segundo o autor, o fato de os morfemas na sentença suaíle serem presos e os da sentença do saramacano serem livres não implica uma maior complexidade da primeira em relação à segunda; uma língua *isolante* não é necessariamente menos complexa que uma *aglutinante*, em suas palavras. No entanto, sentenças que apresentem *disrupções fonéticas ou paradigmáticas* – processos morfofonêmicos e supleção – acabam por apresentar níveis mais altos de complexidade, desencadeados pela existência de flexão.

Por mais que a contribuição de McWhorter para o estudo translinguístico da complexidade seja considerável, algumas ressalvas precisam ser feitas, tendo em vista o enquadramento teórico desta dissertação: ele utiliza os conceitos da tipologia holística, classificando línguas inteiras como isolantes, aglutinantes ou fusionais. Ademais, baseia-se também no conceito de palavra, ao qual seu conceito de *flexão*, mais estrito, está inextricavelmente ligado. Há nisso alguns problemas; o primeiro e mais óbvio é a validade do conceito de palavra, que não se mostra translinguisticamente consistente. O segundo é que no registro e feitura de gramáticas de línguas consideradas *aglutinantes*, como o próprio McWhorter coloca, é documentado o curioso fenômeno de uma mesma língua ser grafada de forma *disjuntiva*, com seus morfemas considerados palavras livres e separadas, ou de forma *conjuntiva* – e exemplos em línguas da mesma família do suaíle, bantu, não são desconhecidos (MATTOS; LAZZARINI-CYRINO, 2020). Por fim, a proposta do autor, embora bastante detalhada, não oferece os meios de abordar o grau de morfologia presa ou solta de uma língua sem aludir à palavra, tampouco de o quantificar, bem como acaba por confundir a análise desse parâmetro e a da complexidade.

Nossa análise usa a proposta de McWhorter como ponto de partida, mas: (i) desconsiderando *palavra*; (ii) desconsiderando a classificação da tipologia holística; (iii) estendendo a ocorrência de *flexão* a todas as línguas, uma vez que realização morfológica de categorias funcionais; (iv) mensurando numericamente o grau de morfologia livre ou presa das línguas; e (v) analisando separadamente este parâmetro e a complexidade com o uso de estatística. De fato, complexidade e morfologia presa parecem andar lado a lado, *grosso modo* acompanhando as afirmações de McWhorter; as línguas com menores escores de parâmetros de complexidade tinham as maiores proporções de morfemas gramaticais livres. No entanto, tirar conclusões analisando apenas poucas sentenças, sem quantificar parâmetros, pode não ser a melhor medida.

A língua bantu analisada nesta dissertação, por exemplo, é o cuwabo; de perfil tradicionalmente descrito como *aglutinante*, e da mesma família que o suaíle, o cuwabo apresentou complexidade bastante alta em todos os quatro parâmetros com que trabalhamos – 46,8% de flexão contextual, média de 3,09 de complexidade fenogramatical-verbosidade, 51,5% de elaboração estrutural/irregularidade e 45% de poliexponência. Dificilmente as medidas de complexidade do saramacano seriam as mesmas apenas com a diferença no grau de morfologia presa/livre, sendo o saramacano uma língua de gramática *telegráfica* como são os crioulos. Veja-se em 40. a mesma sentença de 39., mas agora em cuwabo:

40. Cuwabo

*ddiifúná* [dʒiifúná]

ddi-hí-fúná

1SG-PFV.DJ-querer

“Eu quis.” (GUÉROIS, 2015, adaptado)

Perceba-se a existência de processos morfofonológicos que causam a queda da consoante inicial do morfema de perfectivo *-hí-* por conta da vizinhança com o morfema de primeira pessoa do singular *ddi-*; a sentença já demonstra um nível de complexidade não visto nem no saramacano, nem no suaíle; processos como este são frequentes nessa língua; Além disso, existem dois morfemas diferentes de aspecto perfectivo, um disjuntivo e um conjuntivo, cujo uso é determinado sintaticamente – o que não ocorre no saramacano e no suaíle, e demonstra mais uma complexidade do cuwabo. Igualmente às quatro medidas de complexidade mensuradas para o cuwabo neste trabalho, todos esses fatores passariam despercebidos caso alguém se deparasse ou utilizasse apenas a classificação holística do cuwabo como *aglutinante*.

### 5.3. Mudança por contato

Segundo autores como McWhorter (2001, 2007, 2016, 2019) e Trudgill (2011), o estado natural e ininterrupto das línguas humanas é o de acumular complexidade, devido ao fato de a maior parte de seus falantes adquirirem-nas na fase infantil – as crianças são extraordinariamente capazes de adquirir uma língua com todas as suas dificuldades e complexidade sem maiores problemas, contanto que expostas a ela durante uma determinada fase. Foi essa a situação da maioria das línguas humanas durante a história: grupos pequenos, com número reduzido de falantes e alto grau de conhecimento de mundo compartilhado, caracterizando o que McWhorter, usando o conceito de William Thurston (1987), chama de *gramáticas esotéricas*, e Wray e Grace (2007) chamam de usadas em *comunicação esotérica*. Raramente aprendidas por adultos, tais línguas acabariam favorecendo a compacidade, com elisão de vogais, redução fonêmica e fusão dos morfemas entre si (THURSTON *et al.*, 1987; MATTOS; LAZZARINI-CYRINO, 2020). Das línguas aqui utilizadas, muitas são minoritárias ou indígenas, relativamente isoladas ao longo de suas histórias, com altos níveis

de complexidade mensurados de acordo com os parâmetros escolhidos – a exemplo do yupik, o yakkha, o saami de Pite, o khwarshi e o cuwabo.

Diferente é a situação de línguas com gramáticas *exotéricas*, caracterizadas por interrupção na normalidade de sua transmissão; em algum momento do seu passado, tais línguas teriam sido aprendidas por uma grande massa de adultos aloglotas. O adulto, não tão apto a aprender todas as filigranas e complexidades quanto as crianças, agiria para simplificar a gramática da língua (MCWHORTER, 2007, 2016). Esse, segundo o autor, não é o estado livre de distúrbios da mudança linguística, e sim a exceção, característica de sociedades pós-neolíticas com grandes movimentos populacionais em suas linhas do tempo, conquista, aculturação e outros cenários decorrentes de contato intenso. As gramáticas esotéricas, portanto, seriam grandemente simplificadas (“*streamlined*”), uma vez que “é de grande ajuda ao aprendiz [adulto] se cada morfema puder ser facilmente analisado e identificado” (THURSTON *et al.*, 1987, p.58).

Algumas observações merecem ser feitas, haja vista a escolha por esse enquadramento teórico: (i) a situação descrita não é uma dicotomia entre línguas radicalmente simplificadas e línguas que mantiveram sua complexidade; o que ocorre na realidade é um *continuum* de perda de complexidade e situações de contato mais ou menos intensas, com proporção maior ou menor de aloglotas a adotar a língua – os casos extremos são os crioulos, mas casos intermediários de simplificação menos dramática existem, a exemplo de línguas como o inglês, o persa, o português popular brasileiro e os árabes vernaculares urbanos (MCWHORTER, 2007); (ii) não delimitamos como sendo um único mecanismo envolvido no contato, seja a adoção por parte de aloglotas adultos, seja algum efeito areal, podendo agir os dois ou outro mecanismo ainda por se elucidar<sup>31</sup>; e (iii) nada impede que uma língua submetida a uma simplificação radical de sua gramática, situação de contato intensa e/ou uso em comunicação exotérica em dado momento de sua história torne a ser usada de forma isolada, esotérica, e recomeça a acumular complexidade. Em verdade, é o esperado, haja vista a situação natural de uma língua seria de desenvolver complexidade na forma de irregularidades, sobrespecificações e outras filigranas morfológicas com o tempo; é o que se teoriza ter acontecido com muitas línguas siníticas, que teriam sofrido simplificação radical por contato no seu passado, mas, à exceção do mandarim (que teria passado por subsequentes contatos e simplificações), tornaram a desenvolver algum grau de complexidade. Os

---

<sup>31</sup> Para McWhorter, a aquisição não-nativa é o único mecanismo capaz de simplificar tão radicalmente uma gramática; para uma discussão acerca deste ponto com o observado a partir dos dados do dâw, ver a subseção 5.4.1 mais à frente.

contornos exatos dessa retomada rumo à complexidade, porém, necessitariam de investigação para serem delineados.

#### 5.4. Crioulos e perda da flexão contextual

Uma das medidas de complexidade, a dicotomia flexão inerente/contextual, também se prestou a embasar argumentos de áreas como a crioulistica, que se debruça sobre línguas frequentemente vistas como “desprovidas de flexão” ou de “pouca complexidade morfológica”, as línguas crioulas. Sobre a suposta ausência de morfologia flexional em crioulos, Kihm (2003) argumenta:

A tradicional conexão entre morfologia flexional e morfologia afixal<sup>32</sup> – a razão profunda pela qual se pensavam as línguas crioulas como não possuindo flexão [...]. Morfologia flexional é a realização explícita de elementos funcionais, para se usar um termo tão neutro quanto possível, da estrutura sintática. Não há necessidade *a priori* de estes elementos se configurarem como afixos; eles podem muito bem ocorrer como formas livres, ou clíticos [...]. **O que línguas crioulas de fato parecem ser desprovidas é de flexão contextual**, como marcação de casos estruturais em sujeitos e objetos diretos (exceto raras formas objeto em pronomes) e concordância sintática substantivo-adjetivo ou verbo-argumento. (p. 335, tradução e grifos nossos)

Arends *et al.* (2006) avaliaram a presença/ausência de morfemas flexionais em quatro crioulos de base latina – angolar, palenquero, papiamentu e seychellois –, chegando à conclusão de que toda a morfologia flexional que os quatro crioulos haviam preservado era do tipo inerente. Luís (2016), estudando o indoportuguês de Korlai, um crioulo falado no estado de Maharashtra, na Índia, argumenta que, ao descartar a morfologia de concordância numeropessoal em verbos, mas reter marcadores de tempo e aspecto e classe conjugacional, o crioulo de Korlai teria seguido a tendência das línguas crioulas de manter apenas a flexão inerente, e perder a contextual. Para Bisang (2020) e McWhorter (2016, 2019), não só as línguas crioulas, bem como diversas outras línguas tradicionalmente tidas por *analíticas* ou *isolantes*, como o iorubá e o mandarim, não apresentariam virtualmente nenhuma flexão contextual devido a características históricas e populacionais de suas comunidades de falantes – envolvendo aquisição adulta em larga escala. Vimos aqui que tal afirmação não é distante da verdade – as línguas deste perfil que foram analisadas de fato apresentaram baixíssima proporção de morfologia flexional contextual, entre outros parâmetros de complexidade,

---

<sup>32</sup> Por “morfologia afixal”, Kihm entende processos como sufixação, prefixação, infixação, apofonia, etc; em suma, qualquer processo que resulte na modificação significativa de um “item”. (p. 333)

como será exposto na seção a seguir. Essas línguas estes autores chamaram de línguas *radicalmente analíticas*, e suas gramáticas, de *telegráficas*.

### 5.5. As línguas GML

As análises apontam para a similaridade de quatro línguas do grupo GML – doravante línguas GML nucleares – quanto a seu inventário morfêmico gramatical – o dâw, o kéó, o mandarim e o rapa nui. Tais línguas possuem alta proporção de morfemas gramaticais livres, acompanhada de baixíssimas proporções dos quatro parâmetros de complexidade absoluta aqui utilizados, a flexão contextual, a complexidade fenogramatical-verbosidade, a elaboração estrutural/irregularidade e a poliexponência. Ao menos duas delas – o kéó e o mandarim – são diretamente referidas por McWhorter (2016, 2019) como sendo *radicalmente analíticas*, desprovidas das “filigranas gramaticais” não estritamente necessárias à comunicação que são características de diversas línguas com maior complexidade absoluta. As quatro línguas se mantiveram próximas na análise por dendrogramas tanto considerando os quatro parâmetros citados, quanto por proporção de morfemas livres, e apresentam possíveis eventos de contato ou aquisição adulta em massa em suas histórias que podem servir para explicar ao menos em parte a natureza menos complexa de sua morfologia gramatical.

#### 5.5.1. Dâw

Língua nadahup falada na região do alto rio Negro, estado do Amazonas, Brasil, o dâw assim é descrito:

Numa perspectiva tipológica, o dâw pode ser considerada uma língua isolante analítica com apenas poucos processos de sufixação. Seu léxico consiste de itens monossilábicos com baixa taxa de afixação. Categorias verbais (por exemplo, tempo e aspecto) e nominais (plural e gênero) são predominantemente indicadas por palavras livres. (OBERT, 2019, p. 45, tradução nossa)

O dâw está inserido na periferia da *Sprachbund* do Vaupés. Essa área linguística se caracteriza pela coexistência de línguas tucano orientais, uma língua aruaque (o tariana), e algumas línguas nadahup, em intenso contato. Este contato tem contornos bastante únicos, decorrentes da estrutura da convivência entre tais grupos linguísticos; a área é caracterizada pela ocorrência de exogamia linguística – isso é, membros de uma determinada comunidade de falantes podem se casar apenas com um membro de outra comunidade de falantes

(MARTINS e MARTINS, 1999; AIKHENVALD, 2002, 2012). Essa situação, porém, se dá apenas entre falantes das línguas tucano e o tariana, que são sedentários, cultivam roças e habitam as margens dos grandes rios; os falantes das línguas nadahup, pejorativamente chamados de *makú*, não participam do sistema de exogamia, evitando os grandes cursos d'água e habitando o interior da mata como caçadores e coletores. Há uma simbiose entre os grupos sedentários e os nadahup, que trocam cultivos, como a mandioca, por produtos silvestres, numa espécie de clientelismo (*ibid.*).

É hipotetizado que as línguas nadahup originalmente tinham um perfil bastante distinto da maioria de seus atuais integrantes, assemelhando-se ao do nadëb, a única língua nadahup fora do *Sprachbund* do Vaupés. O nadëb é descrito como *altamente ergativo e aglutinante*, e amplamente prefixal em sua afixação (AIKHENVALD, 2002). As línguas nadahup na área do Vaupés, como o dâw, são fracamente sufixais, tendo perdido quase inteiramente o aparato prefixal da família e o alinhamento ergativo – sob pressão das línguas tucano orientais, que são fortemente sufixais e nominativo-acusativas (MARTINS; MARTINS, 1999; AIKHENVALD, 2002). Um exemplo de perda de complexidade no domínio da marcação de posse por conta da participação das línguas nadahup na área do Vaupés é trazido por Alexandra Aikhenvald:

As línguas makú faladas na área do Vaupés, sob dominância das línguas tucano orientais, podem também ter sofrido alguma reestruturação induzida por contato na sua marcação de posse. Posse é marcada via justaposição de termos na família makú [...] Mas o nadëb – [outra língua makú] falada fora da área do Vaupés – tem uma estrutura com um classificador de possuídos para animais e plantas, do tipo “Subih sua-planta banana”, para dizer “a banana de Subih”. Essa estrutura, que não tem equivalente em línguas tucano orientais, não é encontrada nas línguas makú do Vaupés. (2002, p. 81, tradução nossa)

Veja-se como, por influência das línguas tucano orientais, as nadahup na área do Vaupés eliminaram sua provável construção original, com maior verbosidade, e adotaram a simples justaposição dos elementos. Saliente-se que a despeito de os empréstimos lexicais na área do Vaupés serem extremamente mal-vistos, o empréstimo de padrões gramaticais é intenso, caracterizando uma zona de convergência para certas características tipológicas (AIKHENVALD, 2002, 2012). Veja-se em 41.a. a estrutura de posse com animais no dâw, consistindo apenas em justaposição na ordem possuidor-possuído, em contraste com a estrutura de posse em 41.b., do nadëb, mais complexa e necessariamente incluindo um classificador:

41.

a. Dâw

abug            rid        'yãm            waan            dôo'            'yãm xũ'  
 DISC.CONJ    3PL    cachorro       seguir.animal   AUX:fonte      onça

“E então o cachorro deles começou a seguir a onça.” (OBERT, 2019)

b. Nadēb (Brasil: nadahup)

maʃãh                            ʔĩ:ʔ                            awa:r  
 CL:animal.estimacão    1SG.NSUJ                    cachorro

“Meu cachorro” (EPPS; OBERT, 2022)

É igualmente digno de nota que o dâw possui fraca sufixação, possivelmente ainda estando em processo de cooptar morfemas lexicais, oriundos de verbos, para desempenhar papéis de marcação de aspecto, pospostos ao verbo principal (MARTINS; MARTINS, 1999; CARVALHO, 2016; OBERT, 2019). Veja-se:

	<b>auxiliar aspectual</b>	<b>etimologia</b>
imperfectivo	?	?
progressivo	yoow	ser reto
contínuo	dâr raam	<i>rãm</i> "ir"
perfectivo	yût	matar
completivo	rũ'	terminar
habitual	tâg, weer, pũn'	saber? adicionalmente sempre?
télico	rãm/-ãm	ir
durativo	xâd	passar
iterativo	beey	retornar
incoativo	dôo'	tomar

Tabela 11. Morfemas aspectuais do dâw e suas origens em itens lexicais (OBERT, 2019).

Isso aponta para a possibilidade de a língua, após ter perdido seu aparato de prefixação, estar adotando a estratégia dominante no Vaupés, seus morfemas de aspecto estando no início do processo de gramaticalização. Aqui, vê-se uma instância de simplificação gramatical por conta da participação da comunidade de falantes numa *Sprachbund*; as pressões areais, decorrentes de um perfil tipológico bastante distinto do que as línguas



ocidental da ilha da Nova Guiné (*New Guinea* no mapa). Sobre a paisagem linguística da área, Klamer afirma:

A maioria das línguas austronésias da Indonésia oriental e do Pacífico têm sistemas morfológicos que são menos elaborados que o das línguas austronésias faladas em Taiwan, Filipinas ou Indonésia ocidental [...] **Empobrecimento morfológico extremo é encontrado em línguas faladas em Flores central e oriental**, incluindo manggarai, ngada, lio e kéó. (2012, p. 88, tradução e grifos nossos)

De fato, o kéó apresentou baixíssima complexidade absoluta nos parâmetros aqui utilizados. Comparem-se duas sentenças de línguas austronésias, o tukang besi, que não passou pelo “empobrecimento” morfológico extremo descrito, e o kéó:

42.

a. Tukang besi (Célebes: austronésia)

ko-hu'u-aku	te	ika	na	iaku	te	iko'o
2SG.IRR-dar-1SG.O	CORE	peixe	NOM	2SG	CORE	1SG

“Você vai me dar o peixe.” (DONOHUE, 1999)

b. Kéó

péa	nio	gnada	ti'i	wawi
resto	coco	capaz	dar	porco

“(eles) podem dar os restos de coco ao porco.” (BAIRD, 2002)

É bastante óbvio o maior grau de complexidade absoluta do tukang besi – que apresenta marcação de sujeito de segunda pessoa do singular, modo *irrealis*, objeto de primeira pessoa do singular e artigos marcadores de constituinte nuclear não-nominativo –, enquanto a sentença kéó é inteiramente desprovida de marcação morfológica de categorias gramaticais. A sentença tukang besi, portanto, é mais sobrespecificada e mais verbosa que a kéó.

Estudos em genética de populações humanas revelam uma história populacional altamente dinâmica, com extensos movimentos de mudança populacional na Wallaceia tendo ocorrido em dois momentos – um por volta do último máximo glacial (cerca de 15.000 anos atrás), e outro após o contato austronésio (cerca de 3.000 anos atrás) (PURNOMO, 2021). Além das evidências genéticas, diversos autores (DONOHUE, 1999; ROSS, 2003; KLAMER

et al., 2008; KLAMER, 2002, 2012) sugerem que mudanças tipológicas nas línguas da região, tanto austronésias quanto não-austronésias, são atribuíveis a intenso contato entre os falantes de línguas austronésias, mais recentes na região, e os habitantes mais antigos, prováveis falantes de línguas papuanas; falantes de tais línguas ainda existem espalhados principalmente no leste de Nusantara, hoje minoritários (KLAMER *et al.*, 2008; ROBINSON, 2015). Veja-se:

Falantes pré-austronésios desde as Filipinas até Flores passaram a falar línguas austronésias na antiguidade [...] Quando os austronésios chegaram na região de Timor por volta de 3.800 anos atrás (SPRIGGS, 2011), eles podem ter sido tecnologicamente superiores aos povos indígenas de Flores, que passaram a falar línguas austronésias. (ROBINSON, 2015, p. 30, tradução nossa)

Na mesma direção:

É muito provável que em vários lugares as populações originais de falantes papuanos foram confrontadas com grupos menores mas mais poderosos de invasores falantes de línguas austronésias. Os povos indígenas aprenderam a língua austronésia **imperfeitamente**, mantendo algumas de suas rotinas papuanas. (KLAMER *et al.*, 2008, p. 97, tradução e grifos nossos)

McWhorter (2011), sobre o povoamento da ilha de Flores e a baixa complexidade atestada nas línguas do centro desta ilha, hipotetiza ter havido uma terceira migração, relativamente recente, após a chegada dos falantes de línguas austronésias na região. Com base em itens do vocabulário, o autor sugere que grupos das atuais Célebes, falantes de algo próximo ao tukang besi, teriam chegado em Flores e imposto sua língua aos indígenas – que, ao a aprenderem de maneira defectiva, teriam drasticamente simplificado sua gramática.

Alexander Elias (2020), no entanto, discorda de McWhorter, postulando que as línguas de Flores central são, para usar a terminologia que ele próprio usa, *isolantes* por conta de uma forte influência de substrato proveniente de uma hipotética língua também *isolante* já extinta. Esta língua extinta teria as características da área linguística Mekong-Mamberano proposta por Gil (2015), como uso obrigatório de classificadores, maior simplificação morfológica, adjetivos verbais, dentre outras.

Tendo em vista ser um elemento comum entre as diferentes reflexões acerca da realidade linguística de Flores, não é difícil compreender a maior simplificação do inventário morfêmico gramatical do kéó sob a luz do contato entre populações falantes de diferentes línguas.

### 5.5.3. Mandarim

O mandarim é a língua com maior número de falantes nativos do planeta. O ramo em que está inserido, o das línguas siníticas, faz parte da família sino-tibetana; apesar disso, tipologicamente, as línguas siníticas pouco têm a ver com as demais línguas às quais são aparentadas. As línguas sino-tibetanas não-siníticas, chamadas de línguas tibeto-birmanesas<sup>34</sup>, variam de perfis mais conservadores e com morfologia intrincada, como as rGyalrong, kiranti e nung, a línguas consideradas *aglutinantes*, como o tibetano moderno (DELANCEY, 2013). Em contraste, as línguas siníticas são caracteristicamente desprovidas de complexidade morfológica, possuindo o que muitas vezes na literatura se chamou de *morfologia criouloide* (MATTHEWS; ANSALDO, 1998; DELANCEY, 2013; MCWHORTER, 2016) – outra denominação para gramáticas telegráficas ou radicalmente analíticas.

Em verdade, a disparidade entre as línguas siníticas e as demais sino-tibetanas é tão grande que por muito tempo pôs em prova a relação entre os dois grupos; o perfil sinítico é tão mais afim do de línguas de outras famílias do sudeste asiático, como a austroasiática, a hmong-mien e a tai-kadai, que relações genéticas entre tais grupos já foram aventadas (LUO, 2004; DELANCEY, 2013). Comparem-se, para efeito de ilustração, as sentenças em 41.; duas delas, 43.a. e 43.e., são de línguas sino-tibetanas – a primeira, o yakkha, é uma língua kiranti considerada mais conservadora, e a outra é do próprio mandarim. Em 43.b., 43.c. e 43.d., temos respectivamente sentenças de uma língua austroasiática, de uma língua hmong-mien e uma tai-kadai.

43.

a. Yakkha

kaniŋ	kei	kheps-u-m-na.
1PL.ERG	tambor	ouvir-3.P.PST-1PL.A-NMLZ.SG

“Nós ouvimos o tambor.” (SCHACKOW, 2015)

b. Vietnamita (Vietnã: austroasiática)

nó	đã	sửa	xe	ròi.
3SG	PFV	consertar	carro	já

“Ele já consertou o carro.” (PHAN; DUFFIELD, 2019)

<sup>34</sup> Embora sejam tratadas como um grupo, existem poucas evidências de que as línguas tibeto-birmanesas sejam de fato um grupo coeso, e sim apenas uma reunião parafilética das línguas sino-tibetanas que não são siníticas.

c. Miao (Sudeste asiático: hmong-mien)

tus    dey                    tom                    tus    npua  
ANIM cachorro            morder            ANIM porco

“O cachorro mordeu o porco.” (CLARK *et al.*, 1989)

d. Tailandês (Tailândia: tai-kadai)

chǎn   pháa                    tææŋmoo            mây    ʔòk  
1SG    cortar.ao.meio            melão            NEG    sair

“Tentei cortar o melão ao meio mas não consegui.” (THEPKANJANA;  
UEHARA, 2004)

e. Mandarim

wǒ    chī    le    tián    dōngxī  
1SG    comer P<sub>RV</sub>    doce    coisa

“Eu comi algo doce.”

Percebe-se grande similaridade entre o mandarim e as línguas de três outras famílias; a ordem dos constituintes nas sentenças declarativas não-marcadas é SVO, em contraste com a ordem SOV do yakkha, mais característica das línguas sino-tibetanas não-siníticas. Além disto, enquanto a sentença yakkha apresenta concordância de sujeito e objeto e marcação de caso, as sentenças de todas as demais línguas não demonstram tal acúmulo de material morfológico, apresentando níveis muito mais baixos de sobrespecificação, elaboração estrutural e complexidade fenogramatical-verbosidade.

É hipotetizado, portanto, que o proto-sinítico é resultado da adoção em massa de uma língua sino-tibetana por falantes de outras línguas, numa situação bastante complexa em que os aloglotas seriam de mais de uma família linguística (DELANCEY, 2013; MCWHORTER, 2016). Scott DeLancey chega a aventar a possibilidade de que o proto-sinítico seria o resultado da conquista da dinastia Shang pelos Zhou – esses últimos, falantes de uma língua sino-tibetana –, por volta do ano 1000 antes da era comum. Uma vez que a própria dinastia Shang já teria sido responsável por conquista, escravização e movimentos populacionais em larga escala na planície central da China, possivelmente tendo favorecido a existência de uma ou mais *lingue franche* no território do império, a subsequente imposição de uma língua

sino-tibetana teria aumentado consideravelmente a complexidade da paisagem linguística, que já deveria contar com falantes de línguas austroasiáticas, austronésias, hmong-mien e tai-kadai.

Nós vemos, no chinês antigo, padrões sintáticos SOV e SVO, ao ponto que Cheng (1983) fala em “duas sublínguas coexistindo no chinês arcaico”, um estrato SOV mais antigo e uma sintaxe SVO inovativa, e Xu (2004), em uma língua “tipologicamente misturada”. Isso seria, essencialmente, o “puro” Zhou sino-tibetano com sintaxe SOV, e a *lingua franca* influenciada pelo Zhou, com o padrão SVO das línguas bai yue<sup>35</sup> (DELANCEY, pp. 91-92, tradução nossa)

O chinês Zhou tardio, como é chamado o estágio do chinês pós-conquista dos Shang, era ainda mais desprovido de filigranas morfológicas que o mandarim atual, sendo inusitadamente *telegráfico* – com autores inclusive situando-o como tendo menor complexidade que línguas crioulas (MCWHORTER, 2016):

#### 44. Chinês Zhou tardio† (China: sino-tibetana)

yī                bù     sān     shì                bù     fù     qí     yào  
médico        NEG   três   geração        NEG   tomar   3.POSS remédio

“Se o médico não for de terceira geração, não tome o remédio dele.” (HERFORTH, 2003)

Adicionalmente, evidências arqueológicas e genéticas indicam não só grande diversidade humana no sul da atual China a partir de Shandong e o rio Yangtze (LUO, 2008), como também contato e miscigenação entre falantes do sinítico provenientes do norte e os indígenas (BLENCH, 2008; HE *et al.*, 2020; WANG *et al.*, 2021), que até hoje se manifesta como um gradiente genético norte-sul (CHEN *et al.*, 2009). Delineia-se, desta forma, um cenário de intenso contato entre falantes de línguas diferentes entre si, que passaram por subsequente sinicização e absorção na matriz étnica han, reforçada por posterior expansão rumo ao sul, levando uma língua sinítica significativamente diferente da sua ancestral sino-tibetana.

A tipologia das línguas siníticas pode representar o resultado de crioulização antiga, ou ao menos mudança induzida por contato envolvendo um subgrupo dos fatores envolvidos na crioulização. A situação multidialetal de longa data na China, como

---

<sup>35</sup> Do chinês 百越 *bǎiyuè*, consistem em povos testemunhados desde a antiguidade no sul da China, falantes de línguas não-sino-tibetanas – possivelmente austroasiáticas, austronésias, hmong-mien e tai-kadai –, que sofreram posterior sinicização, embora ainda restem minorias étnicas na região até hoje.

descrita por Escure (1997), com uma língua alvo falada por uma parte pequena e altamente educada da população e uma grande variedade de línguas locais faladas pelas massas apresenta um cenário sócio-histórico não muito diferente daquele no qual os crioulos se situam. Essas circunstâncias podem ser responsáveis pela manutenção, senão pela origem, das características criouloides nas línguas siníticas. (MATTHEWS; ANSALDO, 1998, tradução nossa)

Além da possível criouliização (ou situação análoga à criouliização) envolvida na diferenciação das línguas siníticas a partir de um ancestral sino-tibetano, o mandarim é ainda mais simplificado em relação às demais línguas siníticas, não possuindo muitas das complexidades das línguas mais ao sul (MCWHORTER 2007; ARCODIA, 2015). McWhorter (2007) sugere que o desenvolvimento do mandarim nas planícies do norte da China, em intenso contato com falantes de línguas mongólicas, coreânicas e tungúsicas, pode ter agido como uma instância adicional de simplificação, desnudando sua gramática de muitas das filigranas que poderiam ter se acumulado naturalmente<sup>36</sup> desde a simplificação inicial a que o proto-sinítico já tinha sido submetido.

#### 5.5.4. Rapa nui

Língua da família austronésia, pertencente ao ramo oceânico – e dentro deste, do grupo polinésio –, é falada na remota Ilha de Páscoa, que se encontra sob a jurisdição chilena, e é a aresta mais a leste do vasto triângulo formado por ela, a Nova Zelândia e o Havaí, delimitando a região da Polinésia. A morfologia do rapa nui é caracteristicamente polinésia – grupo de línguas que, ao contrário de muitas das línguas austronésias, é consistentemente descrito como sendo de línguas *isolantes* (BIGGS, 1960; MASSAM, 2020). Comparem-se sentenças transitivas de três línguas austronésias em 45.: novamente o tukang besi, uma língua malaio-polinésia nuclear (não-polinésia), e duas polinésias, o rapa nui e o havaiano:

45.

a. Tukang besi

no-kiki'i-ko	na	iko'o	te	beka
3R-morder-2SG.o	NOM	2SG	CORE	gato

“O gato te mordeu.” (DONOHUE, 1999)

---

<sup>36</sup> E de fato em McWhorter (2007) vê-se uma comparação com línguas siníticas mais ao sul, como o min e o cantonês, que se apresentam com maiores níveis de sobreespecificação, elaboração estrutural e irregularidades que o mandarim.

b. Rapa nui

ā'ana            te        kai     i        te        me'e    ririva            ta'ato'a  
POSS.3SG.A    ART    comer ACC    ART    coisa    boa:RED        todas

“(foi ele quem) comeu todas as melhores coisas.” (KIEVIET, 2017)

c. Havaiano (Havaí: austronésia)

ua        'ai        ke        kanaka            i        ka        poi  
PFV    comer ART    homem            ACC    ART    poi

“O homem comeu o poi.” (ELBERT; PUKUI, 2001)

Vemos que a sentença *tukang besi* apresenta morfologia gramatical robusta, com marcação de objeto de segunda pessoa e de terceira pessoa (o agente) no verbo, bem como um artigo nuclear não-nominativo marcando o objeto e um marcador de nominativo, enquanto o verbo em *rapa nui* e *havaiano* não possui quaisquer morfemas gramaticais associados, e ambas as línguas marcam apenas o acusativo. É dizer, segundo os parâmetros adotados nesta dissertação, a sentença *tukang besi* é mais sobrespecificada, mais elaborada estruturalmente, e tem maior complexidade fenogramatical-verbosidade.

Por que isso se daria? Friedrich Müller, em sua *Grundriss der Sprachwissenschaft* (1884) já nota a maior simplicidade morfológica das línguas polinésias; sobre as conclusões de Müller, William Grace comenta:

Ele considerou **as línguas polinésias como as mais simples em estrutura fonológica e morfológica**, as melanésias intermediárias, e as indonésias as mais complexas. Ele presumiu que a proto-língua teria sido mais semelhante às polinésias em morfologia. Nenhuma razão foi dada, mas uma assunção de evolução do simples ao complexo parece estar implicada. (GRACE, 1958, p. 6, tradução e grifos nossos)

Perceba-se que ao se referir às línguas *indonésias*, mais complexas, Müller provavelmente se referia às línguas malaio-polinésias nucleares, não-oceânicas – ou a algum grupo que as incluísse –, nas quais se insere o *tukang besi*, por exemplo. Note-se também a concepção de evolução das línguas como sempre partindo do simples ao complexo, sendo o simples considerado o primitivo – alinhado à episteme do século XIX. A aceção aceita nessa dissertação, porém, é outra.

Segundo a hipótese de McWhorter (2001, 2007, 2016, 2019), uma língua não se desnudaria de suas filigranas morfológicas simplesmente por deriva; tais casos seriam

explicados por intenso contato e adoção por adultos aloglotas, como exposto. A história das populações polinésias parece apontar para importantes eventos de contato, de fato; da origem comum dos falantes das línguas austronésias na atual ilha de Taiwan – indicado como *Formosa*, no canto superior esquerdo do mapa na figura 18 –, passando pelas Filipinas ou pela Indonésia, os falantes do proto-oceânico (subgrupo das línguas austronésias que inclui a maior parte das línguas melanésias e as línguas polinésias) teriam se deslocado por território já ocupado por prováveis falantes de línguas papuanas, sofrendo considerável contato com essas populações (PAWLEY, 1973; WHYTE *et al.*, 2005; KLAMER, 2012, PURNOMO, 2021). Adicionalmente, os falantes do proto-polinésio poderiam também ter entrado em intenso contato e miscigenação com melanésios após a separação dos dois grupos, antes da famosa migração polinésia pelo Pacífico sul e povoamento de suas ilhas, as áreas mais remotas – e últimas a serem atingidas – ocupadas pelos seres humanos fora da África. Desse contato entre polinésios e melanésios – a chamada hipótese do “barco lento” –, têm-se evidências lexicais, antropológicas e genéticas (STEPHEN; RICHARDS, 2001; WHYTE *et al.*, 2005; KAYSER *et al.*, 2000, 2006, 2008; BARTH, 2022;).

Vê-se claramente que as línguas polinésias, grupo do qual faz parte o rapa nui, têm sua morfologia consideravelmente simplificada (“*streamlined*”) em comparação à vastíssima maioria das demais línguas austronésias. Tal simplificação possivelmente ocorreu após movimentações humanas que puseram os falantes da língua ancestral polinésia em contato com populações aloglotas melanésias e/ou papuanas, adicionando a baixa complexidade absoluta de sua morfologia gramatical como mais uma evidência – que se soma às lexicais, antropológicas e genéticas – de miscigenação e contato íntimo entre tais populações na sua história remota.

#### 5.5.5. Daai chin e koyra chiini

Por fim, das seis línguas originalmente pertencentes ao grupo GML, duas, daai chin e koyra chiini, revelaram-se mais afastadas das demais no *clustering* hierárquico que mediu as distâncias entre as línguas com base na combinação dos quatro parâmetros de complexidade absoluta (figura 16). Ambas apresentaram morfemas envolvidos em construções de terceiro grau de complexidade fenogramatical-verbosidade – em contraste com as demais GML, que apresentaram apenas o primeiro e o segundo grau –, bem como níveis consideráveis de morfemas veiculadores de mais de uma categoria gramatical (figura 7), bastante acima das

GML nucleares. O daai chin especificamente apresentou proporção alta de morfemas flexionais contextuais, a par com línguas como o moloko, o limilngan e o yakkha (figura 11); isso pode ser percebido pela presença de marcação morfológica de caso e concordância que a língua possui, diferentemente das outras GML:

#### 46. Daai chin

mhnam pa:    noh    aksük ksük    a    thi:ngj nu:    ah  
 deus.criador    ERG    primeiro    CFIN    árvore grande    S.AGR:3s

msom ma

criar    fazer.primeiro

“O deus criador antes de tudo criou uma grande árvore.” (SO-HARTMANN, 2009)

O daai chin é uma língua minoritária de Myanmar, falada por cerca de 45.000 pessoas no estado de Chin, oeste do país, que habitam uma área montanhosa e remota, organizando-se em aproximadamente 160 vilas. As montanhas variam de 1.500 a 3.000 metros de altitude e são cobertas por densa floresta tropical, tornando a região em que as vilas se assentam de difícil acesso até o presente momento, isolando o povo daai dos demais birmaneses (SO-HARTMANN, 2009). Apesar de ser descrita como uma língua *isolante* em sua gramática, o daai chin organiza seus verbos de acordo com complexos de fronteiras bem definidas, podendo conter direcionais, aspecto, aplicativos, causativos, modais, tempo, número e outras categorias gramaticais – e nisso se assemelha a outra língua chin, o lai, usado na literatura como exemplo para ilustrar o conceito de *palavra gramatical* formada por várias palavras unidas em ordem rígida (BICKEL e NICHOLS, 2007).

Além disso, a língua daai chin possui uma série de filigranas gramaticais e complexidades não encontradas em línguas como o mandarim, o káo, o dâw e o rapa nui: é uma língua de alinhamento ergativo – que é mais sobrespecificado que o alinhamento nominativo/acusativo (MCWHORTER, 2007); assim como a maioria das línguas chin<sup>37</sup>, o daai exhibe alternância de tronco verbal em 20% de todos os seus verbos, algumas orações selecionando o tronco A, e outras, o tronco B, podendo ainda tais padrões serem alterados pelo emprego de causativos, aplicativos e mudanças de foco – tudo isso revelando alto grau de elaboração estrutural e irregularidade, como também possuindo um sistema de

<sup>37</sup> Línguas tibeto-birmanesas aparentadas ao daai chin e faladas ao redor da área em que o daai chin é falado.

concordância intrincado, com concordância obrigatória de sujeito e objeto (SO-HARTMANN, 2009). De fato, em nossas análises, o daai chin demonstrou ter mais complexidade absoluta que as línguas GML nucleares em todos os quatro parâmetros. Tudo isso aponta para o fato de que embora seja *grafado* de forma *disjuntiva*, o daai chin não se comporta como línguas *radicalmente analíticas* ou de gramáticas *telegráficas* (MCWHORTER, 2016).

Igualmente, conforme evidenciado, o koyra chiini, ao contrário das línguas do grupo GML nucleares, possui altos níveis de elaboração estrutural/irregularidade, sob a forma de alomorfa, processos morfofonológicos, irregularidade e poliexponência. É possível que o grau de elaboração estrutural do koyra chiini seja ainda maior, uma vez que, por escolha do linguista que elaborou essa gramática em específico, alguns dos processos morfofonêmicos, como contrações e elisões de vogais, não foram registrados na glosa. São casos como os seguintes:

#### 47. Koyra chiini

##### a. Contrações com o morfema de imperfeito:

	pronúncia	glosa
ŋgu <sub>o</sub>	[ŋgo:]	LOGO.SGS IPFV
ŋga <sub>o</sub>	[ŋgo:]	SFOC IPFV

##### b. Com pronomes:

	pronúncia	tradução
ay <sub>č</sub> erbu ga <sub>i</sub> se	[at:ferbugi:se]	“Eu o mostrei a eles”

##### c. Com verbos:

	pronúncia	tradução
ay noo <sub>i</sub> se X	[ajni:se]	“Eu dei X a ele/a”

(HEATH, 1999, pp. 31-33)

Apesar disso, o koyra chiini continua sendo menos verboso e com menos morfologia flexional contextual que a maioria das línguas não-GML; das outras 18 línguas do estudo, 13 apresentaram maiores proporções de flexão contextual (figura 11) e de complexidade fenogramatical-verbosidade (figuras 12 e 13). No dendrograma que combinou os quatro parâmetros de complexidade e a proporção de morfologia presa, o koyra chiini associou-se

novamente às GML. Veja-se que dentre as línguas songai, o koyra chiini é divergente – a maior parte das demais línguas têm ordem predominante SOV e são tonais, enquanto o koyra chiini é predominantemente SVO e não-tonal<sup>38</sup> (STAUFFER, 1997; HEATH, 1999) –, sendo uma língua falada em contexto urbano na antiga cidade de Timbuctu, importante centro dos impérios de Mali e Songai. É curioso que alguns autores (NICOLAÏ, 1984, 1990; STAUFFER, 1997) hipotetizem que as línguas songai sejam provenientes de um crioulo – o proto-songai, surgido do contato de falantes de línguas mande como malinké e bambara, e falantes de línguas afroasiáticas, como árabe e tamasheq –, uma vez que a região songai era encruzilhada de diversas rotas de comércio de sal, ouro e nozes de cola. Há ainda a possibilidade menos extrema de o contato ter gerado não uma forma tão radicalmente simplificada como um crioulo, e sim, por meio de moderada aquisição não-nativa adulta, uma simplificação mais leve da gramática proto-songai – similar a processos sofridos por línguas como o persa, o inglês e os árabes vernaculares urbanos (MCWHORTER, 2007), o que seria condizente com uma língua com baixos parâmetros de complexidade absoluta, mas não tão baixos quanto o dâw, o kéro, o mandarim e o rapa nui.

## 5.6. Línguas das Américas

Das línguas utilizadas no estudo, seis são das Américas: duas da América do Norte, o yupik do Alasca central e o zapoteca de San Bartolomé Zoogocho, e quatro da América do Sul, o mapudungun, o dâw, o kanoê e o araweté. Não apenas todas elas não apresentaram níveis significativos de morfologia flexional contextual à exceção do yupik, como também o yupik foi a língua com maior proporção desta morfologia entre todas as 19 línguas do estudo. Além disto, todas as línguas americanas apresentaram baixa proporção de morfemas gramaticais poliexponentes – à exceção do kanoê, e, novamente, do yupik, que foi a língua com maior proporção de morfologia poliexponencial entre todas as 19, com 57%. No entanto, o yupik, como todas as línguas esquimó-aleútas e as línguas na-dene (atabascanas), não é geneticamente próximo das demais línguas nativas do continente (por vezes agrupadas como *ameríndias*); as línguas esquimó-aleútas provavelmente chegaram ao Neártico há muito menos tempo que as línguas ameríndias, mantendo-se isoladas das demais línguas do continente até a expansão das sociedades coloniais (ARNAIZ-VILLENA, 2010; MAZIÈRES, 2011). Não é de espantar que o yupik tenha demonstrado ter um perfil diferente do das línguas

---

<sup>38</sup> Perceba-se que ambas são características listadas por McWhorter (1998, 2001) como presentes no protótipo crioulo.

ameríndias, que se caracterizam aqui por morfologia flexional contextual pouco robusta, embora tenham apresentado variações nos demais parâmetros de complexidade.

### 5.6.1. Velho Mundo *versus* áreas de colonização

Johanna Nichols, em seu livro *Linguistic diversity in space and time* (1992), utiliza certos parâmetros, como marcação de algumas categorias morfológicas relacionadas a possessivos e sujeitos, para tentativamente contabilizar a complexidade de várias línguas ao redor do mundo. Nichols obtém um índice de complexidade que, nas palavras da autora,

[...] negligencia uma boa porção da real complexidade morfológica das línguas, tanto ao omitir categorias comumente veiculadas por morfologia (como tempo e aspecto em verbos), quanto ao considerar apenas onde e se algo é marcado, mas não como (como por flexão, aglutinação, cliticização ou incorporação) [...]. É minha impressão que a complexidade como medida aqui, mesmo sendo parcial e artificial, correlaciona-se diretamente com a complexidade morfológica geral, podendo ser usado como índice de algo real. (1992, p. 64, tradução nossa)

Seu índice opera de forma similar ao cálculo para verbosidade apresentado nesta dissertação, embora, como diz a autora, de abrangência mais limitada quanto aos diversos fenômenos morfológicos; seu mérito está no grande número de línguas para as quais a medida foi calculada – 172 ao todo. Construções com marcação no núcleo, marcação em dependentes ou marcação solta pontuam, e construções sem marcação não pontuam, a soma total de pontos para cada língua determinando um valor de complexidade. Nichols estabelece três macroáreas mundiais, o Velho Mundo (a Afroeurásia), local de surgimento e expansão inicial da nossa espécie, e as áreas denominadas por ela de *colonizadas*<sup>39</sup>: o Novo Mundo (as Américas) e o Pacífico. Ao analisar a distribuição espacial de várias características – entre elas, os índices decorrentes de seu cálculo de complexidade –, a autora chega à conclusão de que as áreas colonizadas têm média de complexidade de suas línguas menor que a média do Velho Mundo. Além disso, ela estabelece uma hierarquia decrescente de complexidade linguística: Velho Mundo > Novo Mundo > Pacífico, e aduz que, embora a média da complexidade seja menor no Novo Mundo e no Pacífico que no Velho Mundo, ela é mais variável nessas macroáreas. Este fato, e o fato de que também há mais diversidade de estratégias nas zonas colonizadas em relação ao Velho Mundo, são relacionados pela autora com possíveis movimentações

---

<sup>39</sup> *Colonizadas* aqui se refere à expansão humana subsequente à ocupação da Afroeurásia que atingiu tais zonas há dezenas de milhares de anos, não tendo relação com a colonização europeia de algumas das mesmas áreas, ocorrida a partir do séc. XVI.

populacionais e linguísticas de larga escala em áreas de expansão (*spread zones*) na Afroeurásia. Essas movimentações humanas teriam modificado e/ou apagado o panorama original de maior diversidade linguística característica das línguas humanas na paisagem, panorama este preservado nas áreas colonizadas – e em *zonas de refúgio* da Afroeurásia, como o Cáucaso.

Tendo tal contribuição em vista, é digno de nota, portanto, que as línguas das Américas aqui analisadas tenham consistentemente escores baixos em pelo menos uma das medidas de complexidade, a flexão contextual (CS), e escores médios em verbosidade – à exceção do kanoê, com altos índices de verbosidade, e do yupik, a língua mais complexa em todas as medidas aqui feitas, fato este que poderia se dar por conta da referida maior amplitude de medidas atestada por Nichols (1992), a despeito de uma média de complexidade mais baixa. Igualmente, uma das línguas aqui analisadas, o rapa nui, corresponde à zona do Pacífico, e obteve escores baixíssimos de complexidade e alto nível de morfemas livres, estando entre as línguas GML.

#### 5.6.2. O yupik

Por fim, consideremos especificamente o yupik do Alasca central, em contraste com as línguas GML, que apresentou os valores máximos em todos os quatro parâmetros analisados, sendo a língua com maior complexidade absoluta entre as 19 aqui estudadas. Em sua gramática, é descrito como tendo grande quantidade de variação alomórfica por conta de processos morfofonêmicos, e “muitas idiossincrasias (por conta de fusão, supleção, etc.) específicas a sufixos flexionais em particular, e ajustes irregulares e esporádicos específicos a um número limitado de sufixos derivacionais” (MIYAOKA, 2012, p. 193). De fato, o yupik tem uma robusta morfologia de concordância e marcação morfológica de funções gramaticais, também sendo descrito como *polissintético* por conta de incorporação de raízes lexicais:

#### 48. Yupik

maa-ni	enurnar-tut	pissu-qenga-qe-tu-ke-nka
aqui-LOC	escasso-IND.3PL	caçar-NMLZ-ter.como-REG-NMLZ-ABS.1SG.PL

“Os que eu sempre caço são escassos aqui.” (*ibid.*)

No entanto, o impacto do inglês e da educação estadunidense já se faz sentir, com seu número de falantes cada vez mais reduzido e atipicidades na aquisição da língua:

Muitos falantes jovens têm dificuldade em adquirir a morfologia criativa característica da língua. Sua fala tende a ser **morfologicamente menos sintética que a fala tradicional mantida entre os mais velhos** como uma consequência óbvia da influência do analítico inglês. A sintaxe necessariamente seguiu o mesmo caminho. Como resultado, a fala dos mais novos está gradualmente se tornando algo mais ou menos difícil de ser entendido ou até alienígena para seus próprios anciões, que por sua vez a criticam por ser parecida com [fala de] bebês. Por sua vez, a língua mais conservadora dos mais velhos, **com todas as suas complexidades e riqueza**, tornou-se mais difícil de entender para os mais novos. Isso também levou a novos desentendimentos culturais entre as gerações, somando-se às lacunas culturais de longa data entre os yupik e estrangeiros. (MIYAOKA, 2012, p. 9, tradução e grifos nossos)

Do relato, depreende-se que o yupik atualmente passa por uma situação de contato intenso, que representa um vetor de simplificação de sua gramática, com potencial para permanentemente livrá-la das filigranas morfológicas acumuladas ao longo dos anos de isolamento na sua história. Não fosse o próprio desafio que sobreviver à pressão do inglês já representa ao yupik, não seria fantasioso hipotetizar que a língua poderia assumir um perfil bastante diferente no futuro, caso permanecesse existindo.

## Considerações finais

As análises e resultados apresentados e discutidos ao longo desta dissertação são provenientes de uma abordagem pouco ortodoxa – desconsiderar uma unidade tão naturalizada quanto a palavra, bem como quantificar a morfologia gramatical de uma língua de modo a compará-la com outras por meio de ferramentas estatísticas. A grande conclusão é que tudo aponta para o fato de a complexidade absoluta da gramática de uma língua ter relação inversa com a proporção do seu inventário morfêmico que é livre, e que a classificação holística não captura os diferentes perfis das línguas com eficiência, deixando de visualizar muita da diversidade de parâmetros. Destas duas conclusões, temos que (i) abandonar a classificação holística das línguas e sua terminologia pode se revelar uma direção interessante; (ii) a decisão de grafar uma língua de forma disjuntiva ou conjuntiva deve levar em consideração critérios morfofonológicos, bem como uma investigação da complexidade absoluta da língua; (iii) considerar flexão dependente do conceito de palavra é contraprodutivo e não nos permite estudar o fenômeno translinguisticamente da forma adequada.

Das línguas aqui estudadas, podemos compreender que aquelas provenientes de sociedades mais *abertas*, isto é, com intenso contato em suas histórias e estando inseridas em situações comunicativas exotéricas, obtiveram os menores escores de complexidade, de maneira a coadunar com argumentos como os de Peter Trudgill e John McWhorter. Para os parâmetros de complexidade adotados, vemos uma tendência global a ter valores médios, a não ser nas línguas que sofreram tais eventos de contato. Como corolário lógico, e após investigações mais amplas, temos que a simplificação gramatical pode ser usada como indicador de contato no passado histórico de determinada língua, mesmo na ausência de outros indícios. Além desses pontos, e especificamente sobre a língua daai chin, ressalte-se que grafá-la de forma tão disjuntiva talvez não reflita o seu comportamento; mais estudos com essa língua são necessários. Para mais, o dâw é falado por uma comunidade pequena e endógama, sob influência de natureza areal – não sendo um provável alvo de aquisição não nativa; a despeito disto, apresentou baixos índices de complexidade absoluta. Atribui-se tal fato à participação desta língua no *Sprachbund* do Vaupés, tendo estado sob pressão de línguas tipologicamente bastante distintas de seu provável ancestral. Mais investigações merecem ser feitas com os falantes de dâw e demais línguas nadahup, não só de natureza linguística; malgrado tais observações, o dâw segue como uma possível instância de

simplificação por conta de contato com línguas de perfil tipológico distinto, e não por adoção por aloglotas adultos.

O fenômeno da flexão contextual e altos níveis de morfologia associada a ele talvez sejam mais característicos de línguas da Afroeurásia; todas as línguas com alta proporção de morfologia flexional contextual são da grande massa de terra formada pela união de Ásia, África e Europa, à exceção do yupik – que é relativamente recém-chegada às Américas, vinda da Sibéria. Desta observação, tiramos três pontos: (i) o fenômeno da flexão contextual pode ser o estado básico para línguas na Afroeurásia, estas apresentando valores médios para esse parâmetro – a não ser que tenham sido sujeitas a reestruturação por contato intenso; (ii) as observações de Johanna Nichols quanto à menor complexidade média das áreas colonizadas como as Américas e Pacífico, confrontadas com a da Afroeurásia, parece ter alguma pertinência; e (iii) estudos com quantificação de mais línguas das Américas se fazem necessários.

Um debate interessante é possível de ser feito quanto à terminologia referente a línguas com baixa complexidade gramatical absoluta; como exposto, tradicionalmente foram consideradas *primitivas* e pouco elaboradas, e sua morfologia, descrita como *pobre*, *empobrecida* ou *rudimentar*. Mas, em consonância com os valores atuais, propõe-se buscar outra forma de se referir a línguas com esse perfil – *de gramática pouco robusta*, ou algum termo desprovido de conotações pejorativas e que ao mesmo tempo veicule os sentidos de simplicidade e eficiência, afinal de contas, toda língua humana é eficiente para desempenhar seu papel, e se há línguas que exprimem o mesmo com menos peças, não é por demérito delas. Termos como *otimizadas* ou *arrojadas* poderiam cobrir essa lacuna.

Investigações com línguas crioulas, medindo-se-lhes os parâmetros aqui usados, poderiam lançar luz a respeito de suas semelhanças e diferenças com outras línguas, bem como da relação entre contato e adoção adulta em massa de uma língua e sua simplificação. A quantificação da morfologia gramatical de crioulos e de mais línguas pode se revelar interessante para se tirarem conclusões mais exatas, bem como aliar a presente abordagem a estudos fonológicos, que utilizem como item de análise as unidades sonoras que o falante pronuncie/reconheça, ou estudos de rigidez (ou sua falta) na ordem da combinação dos morfemas. Além disso, analisar a mesma língua com o intuito de medir-lhe os parâmetros propostos, mas em diferentes etapas de sua história, representa outro provável caminho, a despeito do imperativo de se terem registros dessas etapas e os problemas associados a isso –

restrição a línguas de sociedades com escrita, de determinadas famílias, e disponibilidade dos dados.

## Referências

- AFONASINA, A. S.. The Birth of Harmony out of the Spirit of Tekhe. *ΣΧΟΛΗ. Φιλοσοφское антиковедение и классическая традиция*, v. 6, n. 1, pp. 68-75, 2012.
- ANDERSON, S. R. Where's morphology? *Linguistic inquiry*, v. 13, n. 4, p. 571-612, 1982.
- ANDERSON, S. R. Inflection. In M. Hammond and M. Noonan (eds.), *Theoretical Morphology. Approaches in Modern Linguistics*. San Diego etc.: Academic Press, 23-44. 1988.
- ANSALDO, U.; BISANG, W.; SZETO, P. Y. Grammaticalization in isolating languages and the notion of complexity. In NARROG, H. e HEINE, B. (eds.), *Grammaticalization from a typological perspective*, Oxford: Oxford University Press, pp. 219-234, 2018.
- ARCODIA, G. F. More on the morphological typology of Sinitic. *Bulletin of Chinese linguistics*, v. 8, n. 1, p. 5-35, 2015.
- ARENDS, J. et al. On the presence versus absence of morphological marking in four Romance-based creoles. *The structure of Creole words: Segmental, syllabic and morphological aspects*, p. 223-241, 2006.
- ARNAIZ-VILLENA, A. et al. The origin of Amerindians and the peopling of the Americas according to HLA genes: Admixture with Asian and Pacific people. *Current genomics*, v. 11, n. 2, p. 103-114, 2010.
- BACELAR, L. N. *Gramática da língua Kanoê*. Ph.D. dissertation, Katholieke Universiteit Nijmegen, 2004.
- BAIRD, L. et. al. *A grammar of Keo: an Austronesian language of East Nusantara*. 2002.
- BAKKER, D. Language Sampling. In: SONG, Jae Jung. *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press. 2010.
- BARASA, D. Inflectional forms of tense and aspect in Ateso. *The University of Nairobi Journal of Language and Linguistics*. 4, pp. 82-102. 2015.
- BARTH, D. An Interlinearized 'Family Problem' Narrative Text from Matukar Panau. *Asian and African Languages and Linguistics*, n.16, 2022
- BICKEL, B.; NICHOLS, J.: Inflectional morphology In. *Language typology and syntactic description*, v. 3, n. 2, p. 169-240, 2007.
- BICKEL, B. *Typology in the 21st century: Major current developments*. *Linguistic Typology*, v. 11, p. 239-251, 2007.

- BIGGS, B. Morphology-syntax in a Polynesian language. *The Journal of the Polynesian Society*, p. 376-379, 1960.
- 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.
- BLENCH, R. Stratification in the peopling of China. Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics, p. 105, 2008.
- BOOIJ, G. E. Against split morphology. In: *Yearbook of morphology 1993*. Springer, Dordrecht, p. 27-49, 1993.
- BOOIJ, G. E. Inherent versus contextual inflection and the split morphology hypothesis. In *Yearbook of Morphology 1995*, edited by Geert E. Booij and Jaap van Marle, 1–16. Dordrecht: Kluwer, 1996.
- CARVALHO, M. O. P. C. *Aspecto verbal na língua Dâw*. Tese de Doutorado. Universidade de São Paulo, 2016.
- CHEN, J. *et al.* Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *The American Journal of Human Genetics*, v. 85, n. 6, p. 775-785, 2009.
- CHENG, T. F. “SOV sentence patterns in Early Archaic Chinese.” 16th International Conference on Sino-Tibetan Languages and Linguistics, Seattle. 1983
- CLARK, M. *et al.* Hmong and areal South-East Asia. In: *Papers in Southeast Asian Linguistics No. 11: Southeast Asian Syntax*. Pacific Linguistics, 1989.
- CROFT, W. Lexical and grammatical meaning. In G. BOOIJ, C. LEHMANN; J. MUGDAN (eds.), *Morphology: A handbook on inflection and word formation (257–63)*. Berlin/ New York: Mouton de Gruyter, 2000.
- CROFT, William. *Typology and Universals*. Cambridge: Cambridge University Press. 2003.
- DAHL, Ö. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins, 2004.
- DELANCEY, S. The origins of Sinitic. *Increased Empiricism: New Advances in Chinese Linguistics*, p. 73-99, 2013.
- DELMER, J. F. *Analytic Languages in Mainland South-East Asia*. Term Papers, MA Linguistics Program, Payap University, Chiang-Mai, Thailand, 2016.
- DIXON, R. M. W. *The rise and fall of languages*. Cambridge: Cambridge University Press, 1997.

- DIXON, R. M. W.; AIKHENVALD, A. Y. (Ed.). *Word: A cross-linguistic typology*. Cambridge University Press, 2003.
- 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.)
- ELBERT, S. H.; PUKUI, M. K. *Hawaiian grammar*. University of Hawaii Press, 2001.
- ENFIELD, N. J. 'Case relations' in Lao, a radically isolating language. In: MALCHUKOV, A.; SPENCER, A. (eds.) *The Oxford handbook of case*. Oxford University Press. p. 808-819, 2009.
- EPPS, P.; OBERT, K.. Naduhup languages and the typology of nominal classification. *Cadernos de Etnolinguística*.
- ESQUIRE, G. *Creole and Dialect Continua: Standard Acquisition Processes in Belize and China*. Amsterdam/Philadelphia: John Benjamins. 1997.
- FENK-OCZLON, G.; FENK, A. Complexity trade-offs between the subsystems. In: MIESTAMO, M. (org). *Language complexity: Typology, contact, change*, v. 94, p. 43, 2008.
- FRIESEN, D. *A grammar of Moloko*. Berlin: Language Science Press, 2017.
- GARDANI, F. *Borrowing of inflectional morphemes in language contact*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften, 2014.
- GARDANI, F. On morphological borrowing. *Language and Linguistics Compass*, v. 12, n. 10, p. e12302, 2018.
- GARDANI, F. Morphology and contact-induced language change. In: *The Oxford handbook of language contact*, p. 96-122, 2020.
- GIL, D. How complex are isolating languages. In: MIESTAMO, M. (org). *Language complexity: Typology, contact, change*, v. 94, p. 109, 2008.
- GIL, D. The Mekong-Mamberamo Linguistic Area. In: ENFIELD, N. e COMRIE, B. (eds). *Languages of Mainland Southeast Asia: The State of the Art*. Pacific Linguistics. Berlin: De Gruyter Mouton, 2015.
- GILDEA, S. The development of tense markers from demonstrative pronouns in Panare (Cariban). *Studies in Language* 17:53–73. 1993.
- GRACE, G. W. *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family*. Columbia University, 1958.
- 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.

- GREENBERG, J. Some universals of grammar with particular reference to the order of meaningful elements. In: GREENBERG, J. *Universals of Grammar*. Cambridge, MA: MIT Press. 1963. pp. 73-113
- GUÉROIS, R. *A grammar of Cuwabo (Mozambique, Bantu P34)*. 2015. Tese de Doutorado. Université Lyon 2.
- HARRIS, Z. From morpheme to utterance. In: *Language* 22, pp. 161–183. 1946.
- HARVEY, M et. al. *A grammar of Limilngan: A language of the Mary River region, Northern Territory, Australia*. Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University, 2001.
- HASPELMATH, M; SIMS, A. *Understanding morphology*. 2ª edição. London: Hodder Education, 2010.
- HASPELMATH, M. The morph as a minimal linguistic form. *Morphology*, v. 30, n. 2, p. 117-134, 2020.
- HE, G.; WANG, Z.; GUO, J. et al. Inferring the population history of Tai-Kadai-speaking people and southernmost Han Chinese on Hainan Island by genome-wide array genotyping. *Eur J Hum Genet* 28, pp. 1111–1123, 2020.
- HEATH, J. *A grammar of Koyra Chiini: the Songhay of Timbuktu*. Berlim: Walter de Gruyter, 2011.
- HERFORTH, D. A sketch of Late Zhou Chinese grammar. In THURGOOD, G.; LAPOLLA, R. J. (eds), *The Sino-Tibetan Languages*. NY: Routledge, pp. 57–71, 2003.
- HOCKETT, C.F. *A Course in Modern Linguistics*. New York NY: Macmillan, 1958.
- HOPPER, P. On some principles of grammaticalization. In TRAUGOTT, E. e HEINE, B. (eds.), *Approaches to Grammaticalization*, vol. I, pp. 17-35, 1991.
- JAGERSMA, B. *A descriptive grammar of Sumerian*. 2010. Tese de Doutorado. Leiden University.
- JASBI, M. The meaning of the Persian object marker *rā*. *Advances in iranian linguistics*, v. 351, p. 119, 2020.
- KAYSER, M. et al. Melanesian origin of Polynesian Y chromosomes, *Curr Biol*, vol. 10, pp. 1237-1246, 2000.
- KAYSER, M. et al. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular biology and evolution*, v. 23, n. 11, p. 2234-2244, 2006.

- KAYSER, M. *et al.* The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Molecular biology and evolution*, v. 25, n. 7, p. 1362-1374, 2008.
- KHALILOVA, Z. *A grammar of Khwarshi*. 2009. Tese de Doutorado. Leiden University.
- KIBORT, A. A typology of grammatical features. *Grammatical Features*. 2008. Disponível em: <<http://www.grammaticalfeatures.net/inventory.html>>. Acesso em: 18 de agosto de 2022.
- KIEVIET, P. *A grammar of Rapa Nui*. Berlin: Language Science Press, 2017.
- KIHM, A. Inflectional categories in creole languages. In PLAG, I. (ed.): *Phonology and morphology of creole languages*. Tübingen: Niemeyer, p. 333-363, 2003.
- KLAMER, M. Typical features of Austronesian languages in Central/Eastern Indonesia. *Oceanic Linguistics* 41(2): 363-383, 2002.
- KLAMER, M. *Papuan-Austronesian language contact: Alorese from an areal perspective*. University of Hawai'i Press, 2012.
- KLAMER, M.; REESINK, G. P.; STADEN, M. East Nusantara as a linguistic area. In MUYSKEN, P. (ed). *From linguistic areas to areal linguistics*, 95-149. Amsterdam: John Benjamins, 2008.
- KORNAI, A. Phenogrammar. In: *Semantics*. Nova York: Springer, Cham, 2020. p. 127-175.
- KUSTERS, W. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Leiden: Leiden University, 2003.
- KUSTERS, W. Complexity in linguistic theory, language learning and language change. In: MIESTAMO, M. (org). *Language complexity: Typology, contact, change*, v. 94, p. 3-22, 2008.
- LANGACKER, R. W. *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford: Stanford University Press, 1987.
- LAZZARINI-CYRINO, J. P. *Tipologia linguística: métodos, generalizações e diacronia*. Macabéa – Revista Eletrônica Do Netlli, Crato, v. 8., n. 2., p. 306-322, 2019.
- LAZZARINI-CYRINO, J. P.; MATTOS, E. B. *Um estudo exploratório sobre a classificação de morfemas por agrupamento hierárquico para comparação tipológica*. Natal: Revista do GELNE, v. 22, n. 2, Dossiê Morfologia: p. 395-407 2020.
- LEFEBVRE, C; BROUSSEAU, A. M. *A grammar of Fongbe*. Berlin: Mouton de Gruyter. 2001.
- LIEBER, R; ŠTEKAUER, P. Introduction: status and definition of compounding. In: *The Oxford handbook of compounding*. p. 3-18, 2009

- LUÍS, A. R. The loss and survival of inflectional morphology: Contextual vs. inherent inflection in creoles. In: *Romance Linguistics 2009*. John Benjamins, 2010. p. 323-336.
- LUO, Y. Sino-Tai and Tai-Kadai: Another look. In: DILLER, A. V. N.; EDMONSON, J. A.; LUO, Y. (orgs.) *The Tai-Kadai Languages*. Routledge, pp. 25-44, 2004.
- MARTINS, S. e MARTINS, V. Makú. In AIKHENVALD, A. Y. (org). *The Amazonian Languages*. Cambridge University Press, pp. 251-267, 1999.
- MASSAM, D. Niuean: Predicates and arguments in an isolating language. Oxford University Press, 2020.
- MATRAS, Y. *A grammar of Domari*. Berlim: De Gruyter Mouton, 2012.
- MATTOS, E.; LAZZARINI-CYRINO, J. P. Classificação morfológica das línguas: da tipologia holística às definições circulares. *Revista Diadorim*, v. 23, p. 288-312, 2021.
- MCWHORTER, J. Identifying the creole prototype: Vindicating a typological class. *Language*, p. 788-818, 1998.
- MCWHORTER, J. The world's simplest grammars are creole grammars. *Linguistic Typology*, 5.125–166, 2001.
- MCWHORTER, J. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford University Press on Demand, 2007.
- MCWHORTER, John H. Affixless in Austronesian: Why Flores is a puzzle and what to do about it. In: \_\_\_\_\_, *Linguistic simplicity and complexity*, pp. 223-60, 2011.
- MCWHORTER, J. Is radical analyticity normal? In: VAN GELDEREN, E. (ed.). *Cyclical change continued*, Amsterdam: John Benjamins Publishing Company, p. 227-249, 2016.
- 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.
- MAZIÈRES, S. Towards a reconciling model about the initial peopling of America. *Comptes rendus biologiques*, v. 334, n. 7, p. 497-504, 2011.
- MIESTAMO, M. Grammatical complexity in a cross-linguistic perspective. In: \_\_\_\_\_ (org.) *Language complexity: Typology, contact, change*, v. 23, p. 41, 2008.
- MÜLLER, F. *Grundriss der Sprachwissenschaft*. Vienna: Alfred Hölder. 1884.
- NICHOLS, J. *Linguistic diversity in space and time*. University of Chicago Press, 1992.
- NÆSS, Å. Beyond roots and affixes: Äiwoo deverbal nominals and the typology of bound lexical morphemes. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, v. 41, n. 4, p. 914-955, 2017.

- OBERT, K. A codificação linguística do espaço em Dâw. Tese de Doutorado. Universidade de São Paulo. 2019.
- OBERT, K. Complex predicates and space in Dâw (Naduhup language, AM). *STUF-Language Typology and Universals*, v. 73, n. 1, p. 27-55, 2020.
- PAWLEY, A. Papers of the First International Conference on Comparative Austronesian Linguistics, 1974: Oceanic || Some Problems in Proto-Oceanic Grammar. *Oceanic Linguistics*, 12(1-2), 103–188. 1973.
- PAWLEY, A.; GREEN, R. C. The proto-Oceanic language community. *The Journal of Pacific History*, 19(3), 123–146. 1984.
- PHAN, T.; DUFFIELD, N. 'To be tensed or not to be tensed?' The case of Vietnamese. *Investigationes Linguisticae*, v. 41, p. 105-125, 2019.
- PURNOMO, G. A. *et al.* Mitogenomes reveal two major influxes of Papuan ancestry across Wallacea following the last glacial maximum and Austronesian contact. *Genes*, v. 12, n. 7, p. 965, 2021.
- RIDDLE, E. M. Complexity in isolating languages: Lexical elaboration versus grammatical economy. In: MIESTAMO, M. (org). *Language complexity: Typology, contact, change*, v. 94, p. 131-151, 2008.
- RIETVELD, T.; VAN HOUT, R. *Statistics in language research: Analysis of variance*. Walter de Gruyter, 2010.
- ROBINSON, L. C. The Alor-Pantar (Papuan) languages and Austronesian contact in East Nusantara. *Language contact and Austronesian historical linguistics*, p. 19-33, 2015.
- ROSS, M. Typology and language families: A comment on Klamer's "Typical features of Austronesian languages in Central/Eastern Indonesia". *Oceanic Linguistics* 42(2): 506-510, 2003.
- SAPIR, E. *An introduction to the study of speech*. Nova York: Harcourt, Brace, 1921.
- SCHLEGEL, A. W. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande, 1818.
- SCHWEGLER, A. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlim: Walter de Gruyter, 1990.
- SMEETS, I. *A grammar of Mapuche*. Berlim: De Gruyter Mouton, 2008.
- SMITH, A. *Considerations Concerning the First Formation of Languages and the Different Genius of Original and Compounded Languages*. 1759.

- SO-HARTMANN, H. *A descriptive grammar of Daai Chin*. Berkeley: Sino-Tibetan Etymological Dictionary and Thesaurus Project, 2009.
- SOLANO, E. J. B. *Descrição gramatical da língua araweté*. Ph.D. dissertation, Universidade de Brasília, 2009.
- SOLEDADE, J.; LOPES, M. Uma proposta de revisão do conceito de morfema. In: ALMEIDA, A. A. D; SANTOS, E. S.; SOLEDADE, J. (orgs.) *Saberes lexicais: mundos, mentes e usos*. Salvador: EDUFBA, p. 429-461, 2015.
- SONNENSCHNEIN, A. H. *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. University of Southern California, 2004.
- SPENCER, A. Inflection and the lexeme. *Acta Linguistica Hungarica*, v. 47, n. 1, p. 335-344, 2000.
- SPENCER, A., e LUÍS, A. R. (2012). *Clitics*. Cambridge: Cambridge University Press.
- STEPHEN, J.; RICHARDS, M. Polynesian origins: Slow boat to Melanesia?. *Nature*, v. 410, n. 6825, p. 166-168, 2001.
- SPRIGGS, M. Archaeology and the Austronesian expansion: Where are we now? *Antiquity* 85, 510-528. 2011.
- STAUFFER, D. R. *Essential elements of Songhai grammar*. The University of Texas at Arlington, 1997.
- SU, B.; XIAO, C.; DEKA, R. *et al.* Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 107, 582–590, 2000.
- SYBESMA, R. Zhuang: A Tai language with some Sinitic characteristics. In: MUYSKEN, P. *From linguistic areas to areal linguistics*, Leiden: Brill, p. 221-274, 2008.
- THEPKANJANA, K.; UEHARA, S. Directional verbs as success markers in Thai: another grammaticalization path. In: DILLER, A. V. N.; EDMONSON, J. A.; LUO, Y. (orgs.) *The Tai-Kadai Languages*. Routledge, 2004. p. 500-522.
- TRUDGILL, P. Social structure, language contact and language change. In WODAK, R.; JOHNSTONE, B.; KERSWILL, P. E. *The SAGE handbook of sociolinguistics*. Sage Publications, pp. 236-248, 2011.
- VELUPILLAI, V. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins, 2012.
- VOGEL, I. The morphology-phonology interface: isolating to polysynthetic languages. *Acta Linguistica Hungarica*, v. 55, n. 1-2, p. 205-226, 2008.

- WANG, M. *et al.* Reconstructing the genetic admixture history of Tai-Kadai and Sinitic people: Insights from genome-wide SNP data from South China. *Journal of Systematics and Evolution*, 2021.
- WILBUR, J. *A grammar of Pite Saami*. Berlin: Language Science Press, 2014.
- WHYTE, A. L. H.; MARSHALL, S. J.; CHAMBERS, G. K.. Human Evolution in Polynesia. *Human Biology*, 77(2), 157–177. 2005.
- 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, D. *Typological change in Chinese syntax*. OUP Oxford, 2006.