

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
FACULDADE DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

**Livónia dos Santos Araújo da Luz**

**Investigando as marcas de gênero do Tétum-Díli: *mane, feto, aman* e *inan*  
como classificadores nominais**

Juiz de Fora  
2025

**Livónia dos Santos Araújo da Luz**

**Investigando as marcas de gênero do Tétum Díli: *mane, fetu, aman e inan*  
como classificadores nominais**

Dissertação apresentada ao  
Programa de Pós-Graduação em  
Linguística da Faculdade de Letras  
da Universidade Federal de Juiz de  
Fora como requisito parcial à  
obtenção do título de Mestre em  
Linguística.  
Área de concentração: Linguística

Orientadora: Profa. Dra. Paula Roberta Gabbai Armelin

Juiz de Fora  
2025

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

dos Santos Araújo da Luz , Livónia.

Investigando as marcas de gênero do Tétum-Díli : mane, feto, aman e inan como classificadores nominais / Livónia dos Santos Araújo da Luz . -- 2025.

109 f.

Orientadora: Paula Roberta Gabbai Armelin

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2025.

1. Gênero. 2. Tétum-Díli. 3. Composição. 4. Derivação. 5. Classificadores Nominais. I. Gabbai Armelin, Paula Roberta, orient. II. Título.

**Livónia dos Santos Araújo da Luz**

**Investigando as marcas de gênero do Tétum-Dili: *mane, feto, aman e inan***  
como classificadores nominais

Dissertação apresentada ao  
Programa de Pós-Graduação em  
Linguística da Faculdade de Letras  
da Universidade Federal de Juiz de  
Fora como requisito parcial à  
obtenção do título de Mestre em  
Linguística.  
Área de concentração: Linguística

Aprovada em 7 de março de 2025.

**BANCA EXAMINADORA**

---

Profa. Dra. Paula Roberta Gabbai Armelin - Orientadora  
Universidade Federal de Juiz de Fora

---

Professora Dra. Ana Paula Scher  
Universidade Federal de Juiz de Fora/ Universidade de São Paulo

---

Prof. Dr. Vitor Augusto Nóbrega  
Universidade de São Paulo



Documento assinado eletronicamente por **Paula Roberta Gabbai Armelin, Professor(a)**, em 07/03/2025, às 20:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vitor Augusto Nóbrega, Usuário Externo**, em 10/03/2025, às 09:44, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ana Paula Scher, Professor(a)**, em 28/03/2025, às 11:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf ([www2.ufjf.br/SEI](http://www2.ufjf.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **2271459** e o código CRC **5732843B**.

Dedico este trabalho:

Aos meus queridos pais, Anselmo da Luz & Lucinda dos Santos Araújo e José de Araújo e Benetida da Silva, que me proporcionaram todas as oportunidades possíveis para que eu pudesse alcançar meus sonhos.

À minha querida avozinha Angelina Verdial, que agora está no Céu e que preenche um lugar especial no meu coração.

À minha excepcional orientadora, Prof. Dra. Paula Roberta Gabbai Armelin, por acreditar no meu potencial e me guiar com dedicação e sabedoria ao longo deste percurso.

A Deus, por me conceder força, coragem e saúde para enfrentar cada desafio e por estar sempre presente em minha vida.

## **AGRADECIMENTOS**

Em primeiro lugar, expresso minha maior gratidão a Deus Pai Todo-Poderoso, por me conceder força, coragem e saúde para enfrentar cada desafio e por estar sempre presente em minha vida. Agradeço imensamente aos meus protetores divinos: Nossa Senhora de Fátima, Jesus da Divina Misericórdia, Santo Antônio e São Miguel Arcanjo, por me protegerem e iluminarem cada passo do meu caminho.

Aos meus amados pais, Anselmo da Luz & Lucinda dos Santos Araújo e José de Araújo & Benedita da Silva, dedico minha eterna gratidão por terem me proporcionado todas as oportunidades possíveis para alcançar meus sonhos. Agradeço pelo amor incondicional, pelo apoio e pelos sábios conselhos que sempre me guiaram. Minha sincera gratidão aos meus queridos irmãos e sobrinhos de coração, pelo carinho e atenção durante este percurso acadêmico de dois anos longe da família. À minha querida avozinha Angelina Verdial, que agora está no Céu e ocupa um lugar especial em meu coração, minha eterna gratidão. Ao meu amado Júlio Cesar Marçal, agradeço de todo o coração pelo amor incondicional e pelo apoio constante ao longo de toda esta jornada.

Expresso meu profundo agradecimento à Profa. Irta Sequeira e ao Prof. Vicente Paulino por todo o apoio financeiro, motivacional, acompanhamento, amor e carinho durante meus percursos acadêmicos. Minha mais profunda gratidão à minha excepcional orientadora, Profa. Dra. Paula Roberta Gabbai Armelin, por todo o acompanhamento, dedicação, carinho, amizade e motivação ao longo da orientação deste trabalho, bem como pela paciência demonstrada durante os encontros realizados para a elaboração desta dissertação. Sem ela, o sucesso deste trabalho não teria sido possível.

Gostaria de expressar minha profunda gratidão aos professores da minha banca de qualificação e defesa de dissertação, Profa. Dra. Ana Paula Scher e Prof. Dr. Vitor Augusto Nóbrega, por gentilmente aceitarem nosso convite e desempenharem papéis fundamentais na construção desta dissertação.

Agradeço a todos os professores do Programa de Pós-Graduação em Linguística da UFJF, que contribuíram para a minha formação acadêmica durante este período acadêmico. Agradeço ao Coordenador Tiago Timponi Torrente e à Secretária Isabella Oliveira, ambos do Programa de Pós-Graduação em Linguística da UFJF, por estarem sempre disponíveis para atender às minhas necessidades acadêmicas.

Aos Coordenadores Clarissa e Hugo, do Programa de *Incoming Mobility and International Agreements* da Diretoria de Relações Internacionais da UFJF, minha gratidão pelo atendimento necessário, acompanhamento e colaboração desde o início da bolsa e pelo esforço para que eu pudesse chegar ao Brasil. Agradeço ao GCUB de Mobilidade Internacional pela aprovação da bolsa do Programa GCUB-Mob 001/2022. Agradeço à UFJF por me aceitar no programa de mestrado em Linguística no Brasil.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Agradeço à CAPES pela atribuição da bolsa que tornou este trabalho possível.

À UNTL, em parceria com o FDCH, agradeço pela passagem de retorno ao meu país amado, Timor. Agradeço à Professora Karin Indart pela ajuda em contactar sua amiga Aline, que me recebeu no aeroporto na minha chegada ao Brasil.

Minha gratidão à minha querida Mana Aline, Maun Jefferson e toda a família, por me receberem no aeroporto com carinho e por me acompanharem durante minha estadia no Brasil. Vocês são minha segunda família e sou grata por todo o amor, carinho e atenção que me proporcionaram. Minha gratidão estende-se também à Mana Betânia, por me apresentar os encantadores lugares de Juiz de Fora.

Agradeço aos meus amigos da UFJF: Andreia, Jéssica, Leandra, Gustavo, João, Lydsson, Karina, Dalila, Bianca e Regina, que tornaram meu mestrado um processo mais leve e alegre.

Agradeço ao Rafael por ter aceitado minha vivência na República dos Estudantes e aos meus queridos amigos da República dos Estudantes pelo carinho e colaboração durante minha vivência nesse espaço. Um agradecimento especial à Janaína, que tornou todos os meus dias uma festa.

Aos meus queridos amigos de coração, especialmente Laurentina, Isabel, Plagia, Mãezinha Clara, Senhor António, Elísio, Francisco e Fidélia, agradeço imensamente pelo apoio durante todo este tempo. Agradeço também à Mana Natália, Mana Juliana, Maun Tomé, Abelita, Marcelina, Professor Marcos, Professor Nelson, Mana Renata e Maun Dalilo, pelo carinho e amizade. Agradeço à Mana Filipa Filipe pela motivação nas leituras, em especial nas partilhas de textos linguísticos.

Aos meus amigos timorenses que estudam no Brasil, em especial ao Maun Dionísio, Maun Cláudio, Maun Nazário, Maun Vítor e Maun Jackson, minha gratidão

pelo carinho e pelos momentos de risada.

Por fim, expresso meu agradecimento mais profundo a todos que, direta ou indiretamente, contribuíram para a elaboração desta dissertação.

A todos vocês, meu muito obrigada.

## RESUMO

Esta pesquisa descreve e analisa as marcas *mane*, *feto*, *aman* e *inan* do tétum-Díli (TD), interpretadas como gênero da língua, com o objetivo de investigar seu estatuto formal e propor uma estrutura sintática que seja capaz de explicar o comportamento dessas formações. Para tanto, tomamos como ponto de partida o debate existente na literatura, que hesita entre a classificação como derivação (HULL; ECCLES, 2005; Belo e Oliveira, 2021) ou como composição (COSTA, 2015; CHOUPINA, 2011) das formações de gênero na língua. Embora derivação e composição sejam tradicionalmente vistas como processos distintos de formação de palavras, as fronteiras entre tais processos nem sempre são claras. Do ponto de vista empírico, para nos aprofundarmos nesse debate, nos baseamos nos critérios linguísticos sistematizados por Gonçalves e Andrade (2016), que trazem um conjunto de propriedades comumente associadas na literatura aos processos de derivação e composição, levando em consideração propriedades de diversas naturezas, como o estatuto das formas envolvidas no processo de formação de palavras, as propriedades estruturais das formações analisadas, as propriedades fonológicas e semânticas dos elementos linguísticos e os critérios de produtividade e produção dessas formas. A partir daí propomos que, empiricamente, *mane*, *feto*, *aman* e *inan* apresentam um comportamento ambíguo, com algumas propriedades que se aproximam da composição e outras mais próximas do processo de derivação. Propomos, então, que essa dualidade de comportamento pode ser explicada pela hipótese de que tais formadores estão em processo de gramaticalização. Mais especificamente, propomos que, nas formas em que são interpretados como gênero, tais elementos passaram de substantivos independentes a classificadores nominais. Os classificadores nominais são unidades linguísticas que categorizam o nome ao denotar alguma característica saliente da entidade a que um substantivo associado se refere (Grinevald, 2002). Nos dados do TD, a origem lexical de *mane*, *feto*, *aman* e *inan* é observada a partir do fato de que tais elementos podem ser empregados isoladamente com conteúdo conceitual distinto da interpretação de gênero. Por outro lado, o estatuto gramaticalizado dessas marcas pode ser identificado em algumas funções morfossintáticas que tais elementos assumem, como no licenciamento de correferência, retomada anafórica e leitura definida, por exemplo. Além disso, a partir das propriedades elencadas em Aikhenvald (2002) e Grinevald (2002), o estatuto de classificador de *mane*, *feto*, *aman* e *inan* é

compatível com o fato de seu emprego não ser obrigatório, de modo que tais elementos não se aplicam a todos nomes da língua. A partir daí, com base no modelo da Morfologia Distribuída (Halle; Marantz, 1993; Marantz, 1997) propomos uma formalização inicial desses classificadores nominais do TD, tratando-os sintaticamente como núcleos funcionais que atuam como adjuntos. Dessa forma, eles modificam o nome de base, mas não projetam seu rótulo e nem determinam as propriedades formais da estrutura.

**Palavras-chave:** Gênero. Tétum-Díli. Composição. Derivação. Classificadores Nominais.

## ABSTRACT

This research describes and analyzes the markers *mane*, *feto*, *aman*, and *inan* in Tetun Dili (TD), interpreted as gender in the language, aiming to investigate their formal status and to propose a syntactic structure capable of explaining the behavior of these formations. As a starting point, we draw upon the existing debate in the literature, which is divided between classifying these gender formatives as derivation (Hull & Eccles, 2005; Belo & Oliveira, 2021) or compounds (Costa, 2015; Choupina, 2011). Although derivation and compounds are traditionally viewed as distinct word-formation processes, the boundaries between these processes are not always clear. Empirically, deepen this debate, we rely on the linguistic criteria systematized by Gonçalves and Andrade (2016), which provide a set of properties commonly associated in the literature with derivational and compound formation, considering properties of various natures, such as the status of the forms involved in the word-formation process, the structural properties of the analyzed formations, as well as their phonological and semantic properties and the criteria of productivity and creation of these forms. Based on this, we propose that *mane*, *feto*, *aman* e *inan* in TD exhibit an ambiguous behavior, with some properties aligning with compounds and others closer to derivation. We suggest that this duality of behavior can be explained by the hypothesis that the markers *mane*, *feto*, *aman*, and *inan* are undergoing a process of grammaticalization. More specifically, we propose that, in the forms where they are interpreted as gender, these elements have transitioned from independent nouns to nominal classifiers. Nominal classifiers are linguistic units that categorize a noun by denoting a salient characteristic of the entity to which the associated noun refers (Grinevald, 2002). In TD data, the lexical origin of *mane*, *feto*, *aman*, and *inan* is evident in the fact that these elements can be used independently with a conceptual content distinct from the gender interpretation. On the other hand, their grammaticalized status can be identified in some morphosyntactic functions these elements take on, such as licensing coreference, anaphoric relations, and definite reading, for example. Furthermore, based on the properties outlined in Aikhenvald (2002) and Grinevald (2002), the classifier status of *mane*, *feto*, *aman*, and *inan* aligns with the fact that their use is not obligatory, meaning that these elements do not apply to all nouns in the language. To deal with the formalization of *mane*, *feto*, *aman*, and

*inan*, based on the Distributed Morphology model (Halle & Marantz, 1993; Marantz, 1997), we propose that these noun classifiers in TD are adjuncts that do not project their label in the structure. In this way, they modify the base noun, but they do not project its label or determine the formal properties of the structure.

**Keywords:** Gender. Tetun Dili. Compound. Derivation. Noun Classifiers.

## LISTAS DE ILUSTRAÇÕES

<b>Figura 1</b> – Arquitetura da gramática da MD	23
<b>Figura 2</b> – Distribuição das línguas em Timor-Leste	31
<b>Figura 3</b> – Visão geral dos sistemas de classificação no continuum da gramaticalização	82
<b>Figura 4</b> – Visão geral dos sistemas em um continuum de gramaticalização	86

## LISTAS DE TABELAS

<b>Tabela 1</b> – Pronomes pessoais	42
<b>Tabela 2</b> – Pronomes pessoais no TD	59
<b>Tabela 3</b> – Principais diferenças entre composição e derivação	66
<b>Tabela 4</b> – Comportamento das marcas sugestivas de gênero do TD	73
<b>Tabela 5</b> – Classificadores vs. sistemas mais gramaticalizados de classificação	86
<b>Tabela 6</b> – Núcleos e Adjuntos	98

## SUMÁRIO

<b>CAPÍTULO 1: APRESENTAÇÃO DA PESQUISA.....</b>	<b>16</b>
<b>1. INTRODUÇÃO.....</b>	<b>16</b>
1.1. OBJETIVOS DA PESQUISA.....	21
1.2. QUADRO TEÓRICO: MORFOLOGIA DISTRIBUÍDA.....	22
1.3. QUESTÕES DE PESQUISA E HIPÓTESES LEVANTADAS.....	26
1.4. ORGANIZAÇÃO DA PESQUISA.....	28
<b>CAPÍTULO 2: GÊNERO NO CONTEXTO DE CONTATO LINGUÍSTICO DO TIMOR-LESTE.....</b>	<b>29</b>
2.1 TÉTUM-DÍLI E O CONTEXTO TIMORENSE: UM BREVE PANORAMA HISTÓRICO.....	30
2.2 CONTATOS E INFLUÊNCIAS LINGUÍSTICAS NO TÉTUM-DÍLI.....	34
2.3 GRAMÁTICAS EM CONTRASTE: TÉTUM-DILI E PORTUGUÊS.....	38
2.4 GÊNERO E CONTATO LINGUÍSTICO: TÉTUM-DÍLI E PORTUGUÊS.....	43
2.5 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	47
<b>CAPÍTULO 3: AS PROPRIEDADES EMPÍRICAS DAS MARCAS RELACIONADAS A GÊNERO EM TÉTUM-DÍLI.....</b>	<b>48</b>
3.1. GÊNERO EM TÉTUM-DÍLI: A PERSPECTIVA DAS GRAMÁTICAS.....	48
3.2. GÊNERO EM TÉTUM-DILI: UMA VISÃO LINGUÍSTICA.....	52
3.3. GÊNERO EM TÉTUM-DILI: OUTRAS PROPRIEDADES EMPÍRICAS DETECTADAS NA PESQUISA.....	57
3.4. CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	63
<b>CAPÍTULO 4 GÊNERO NO TÉTUM-DÍLI E A FRONTEIRA DERIVAÇÃO ENTRE E COMPOSIÇÃO.....</b>	<b>64</b>
4.1. COMPOSIÇÃO E DERIVAÇÃO: PROPRIEDADES EMPÍRICAS.....	66
4.2. FORMAÇÕES DE GÊNERO NO TD: COMPOSIÇÃO OU DERIVAÇÃO?.....	73
4.3. CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	79
<b>CAPÍTULO 5: UMA PROPOSTA DE ANÁLISE PARA AS FORMAÇÕES EM MANE, FETO, AMAN E INAN COMO CLASSIFICADORES NOMINAIS.....</b>	<b>81</b>
5.1 AS PROPRIEDADES DOS SISTEMAS DE CLASSIFICADORES E OS DADOS DO TÉTUM-DÍLI.....	82

5.2. UMA PROPOSTA DE FORMALIZAÇÃO PARA OS CATEGORIZADORES NOMINAIS DO TÉTUM-DÍLI.....	91
5.2.1. RAÍZES E CATEGORIZAÇÃO NA MORFOLOGIA DISTRIBUÍDA .....	91
5.2.2 NÚCLEOS FUNCIONAIS NO DOMÍNIO NOMINAL.....	94
5.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	99
<b>CAPÍTULO 6: CONSIDERAÇÕES FINAIS .....</b>	<b>101</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>103</b>

## CAPÍTULO 1: APRESENTAÇÃO DA PESQUISA

### 1 INTRODUÇÃO

Esta pesquisa<sup>1</sup> investiga marcas *mane*, *feto*, *aman* e *inan* interpretadas como gênero da língua no Tétum-Díli (TD), uma língua austronésia falada na ilha de Timor. Também conhecido como tétum-praça, o TD é a forma do tétum que se desenvolveu em Díli durante o domínio colonial português, sendo, portanto, de uma língua fortemente influenciada pelo português.

Apresentando características de uma língua predominantemente isolante, o TD não possui flexão nos nomes (número e gênero) e nem dos verbos (tempo, aspeto, pessoa e número), como apontam Hull e Eccles (2005, p. 16-25) e Costa (2015, p. 51- 57). A esse respeito Hajek (2007, p. 168) salienta que:

O Tétum-Díli é mais bem caracterizado com tendência a ser isolante, apresentando muito pouca morfologia verdadeiramente produtiva. A língua não utiliza uma marcação de núcleo nem dependente, e as relações gramaticais são expressas pela ordem dos constituintes. A ordem dos constituintes em tétum-Díli é tipicamente AVO [Agente, Verbo e Objeto] e SV [Sujeito e Verbo], e não há sentenças passivas na língua<sup>2</sup>.

O gênero nos substantivos, foco de análise deste trabalho, se manifesta no TD através dos formadores *mane* ('homem') ou *feto* ('mulher'), indicando masculino e feminino, respectivamente, para entidades com o traço [+humano]. Já para marcar o gênero de entidades [+animadas], como os nomes de animais, empregam-se as formas *aman* ('pai/macho') ou *inan* ('mãe/fêmea'):

---

<sup>1</sup> Este trabalho se insere em um projeto de pesquisa mais amplo contemplado na chamada CNPq/MCTI n. 10/2023 – Processo 409013/2023-4.

<sup>2</sup> “Tetun Dili is best characterized as tending towards isolating, with very little truly productive morphology. The language is neither head- nor dependent marking, and grammatical relations are expressed by constituent order. Constituent order in TD is typically AVO and SV, and there is no passive.” (Hajek, 2007, p. 168) – tradução nossa.

(1)

- a. labarik-mane  
criança-homem  
'menino'
- b. labarik-feto  
criança-mulher  
'menina'
- c. asu-aman  
cachorro-macho  
'cachorro'
- d. asu-inan  
cachorro fêmea  
'cachorra'

Para os substantivos inanimados, por sua vez, majoritariamente, não há qualquer marcação de gênero. Nesse sentido, a marcação sugestiva de gênero no TD é predominantemente associada aos traços semânticos humano/não humano e animado/não animado do referente (Choupina, 2011, p. 67). Na tipologia de Cobertt (1991), o TD é uma língua em que a atribuição de gênero é predominantemente semântica, sendo, em grande parte, deduzida a partir do sexo do referente.

Em linhas gerais, gênero é entendido como a distribuição dos nomes em diferentes classes, sendo refletido na concordância com outros elementos, como determinantes e adjetivos, por exemplo (Kramer, 2015, p. 2):

O gênero é regularmente definido como a classificação de substantivos em duas ou mais classes, conforme refletido na morfologia de concordância em determinantes, adjetivos, verbos e outras categorias sintáticas<sup>3</sup>.

No TD, no entanto, uma vez que a língua não apresenta morfemas flexionais, não é possível detectar um processo de concordância. Assim, os adjetivos que acompanham os substantivos são considerados invariáveis, como exemplificado na forma adjetival *oin-furak* ('bonito'/'bonita') em (2), que permanece com a mesma realização morfofonológica, independentemente do emprego das formas de gênero

---

<sup>3</sup> "Gender is regularly defined as the sorting of nouns into two or more classes, as reflected in agreement morphology on determiners, adjectives, verbs and other syntactic categories." (Kramer, 2015, p. 2) – tradução nossa.

que identificam o referente:

(2)

- a. labarik-mane oin-furak  
criança-homem bonito  
'menino bonito'
- b. labarik-feto oin-furak  
criança-mulher bonito  
'menina bonita'
- c. asu-aman oin-furak  
cachorro-macho bonito  
'cachorro bonito'
- d. asu-inan oin-furak  
cachorro-fêmea bonito  
'cachorra bonita'

Da mesma forma, o TD não possui artigos definidos, nem artigos indefinidos que possam atuar na marcação de gênero (Costa, 2015, p.57). Há na língua apenas um marcador anafórico (*ne'e*), que pode funcionar tanto como demonstrativo ('esse/essa; este/esta; aquele/aquela'), quanto como artigo definido ('o/a,'). No entanto, como se pode ver em (3), o marcador anafórico também permanece invariável a despeito do gênero do referente:

(3)

- a. labarik-mane ne'e oin-furak  
criança-homem ANF bonito  
'o/esse menino é bonito'
- b. labarik-feto ne'e oin-furak  
criança-mulher ANF bonito  
'a/essa menina é bonita'
- c. asu-aman ne'e oin-furak  
cachorro-macho ANF bonito  
'o/esse cachorro é bonito'
- d. asu-inan ne'e oin-furak  
cachorro-fêmea ANF bonito  
'essa cachorra é bonita'

Além disso, a língua apresenta um marcador *ida*, que pode ser utilizado tanto como o numeral ('um/uma') quanto como artigo indefinido ('um/uma'). Como nos outros casos, também esse elemento não apresenta qualquer variação de forma que possa ser correlacionada como uma variação de gênero.

(4)

- a. labarik-mane    ida  
criança-homem um  
'um menino'
- b. labarik-feto    ida  
criança-mulher um  
'uma menina'
- c. asu-aman        ida  
cachorro-macho um  
'um cachorro'
- d. asu-inan        ida  
cachorro-fêmea um  
'uma cachorra'

Ainda que marcas flexionais estejam ausentes do sistema da língua, as formações de gênero do TD constituem-se como um interessante objeto de pesquisa, uma vez que elas têm estatuto ainda incerto na literatura. Aliás, como já apontado em Greenberg (1963), mesmo em línguas consideradas isolantes há uma morfologia ativa cujo estatuto precisa ser investigado.

Provavelmente não há línguas sem composição, afixação ou ambas. Em outras palavras, provavelmente não há línguas puramente isolantes. Há um número considerável de línguas sem flexões, talvez nenhuma sem composição e derivação<sup>4</sup>.

(Greenberg, 1963, p. 92)

---

<sup>4</sup> "There are probably no languages without either compounding, affixing, or both. In other words, there are probably no purely isolating languages. There are a considerable number of languages without inflections, perhaps none without compounding and derivation." (Greenberg 1963 p. 92) – tradução nossa.

Especificamente quanto à morfologia de gênero no TD, paira na literatura uma controvérsia em relação ao estatuto das formas *feto*, *mane*, *aman* e *inan*. Assim, alguns autores defendem que tais elementos são formalmente sufixos que participam de um processo de derivação (Hull e Eccles, 2005; Belo e Oliveira, 2021), sendo as formas de gênero consideradas, então, morfologicamente derivadas. Por sua vez, em outra perspectiva, alguns autores (Costa, 2015; Choupina, 2011) defendem que os marcadores de gênero são palavras independentes e que, portanto, as formações de gênero são geradas via processo de composição.

Inserindo-se nesse debate, esta pesquisa revisita as formações que envolvem a interpretação de gênero no TD, buscando descrever suas propriedades empíricas e desenvolver uma análise formal capaz de derivar seu comportamento. Mais especificamente, propomos que as propriedades dos formadores *mane*, *feto*, *aman* e *inan* do TD são mais bem compreendidas diante da hipótese de que tais elementos atuam como classificadores nominais. Mais especificamente, com base em Aikhenvald (2000), entendemos os classificadores nominais como unidades linguísticas que categorizam o nome ao denotar alguma característica saliente da entidade a que um substantivo associado se refere.

A partir daí, localizamos os formadores *mane*, *feto*, *aman* e *inan* em um contínuo de gramaticalização em que elementos com uma origem lexical passam a integrar o sistema gramatical da língua, funcionando como um dispositivo de classificação nominal. Nos dados do TD, a origem lexical de *mane*, *feto*, *aman* e *inan* é observada a partir do fato de que tais elementos podem ser empregados isoladamente com conteúdo conceitual distinto da interpretação de gênero. Por outro lado, o estatuto gramaticalizado dessas marcas pode ser identificado em algumas funções morfossintáticas que tais elementos assumem, como no licenciamento de correferência, retomada anafórica e leitura definida, por exemplo.

Para colocar em debate a estrutura sintática dessas formações, tomamos como base o modelo da Morfologia Distribuída (Halle; Marantz, 1993; Marantz, 1997), MD, e propomos uma formalização inicial dos classificadores nominais do TD, tratando-os sintaticamente como a realização de núcleos funcionais adjuntos, que não projetam seu rótulo na estrutura.

A partir dessa caracterização geral, na próxima seção apontamos, mais especificamente, os objetivos da pesquisa.

## 1.1 OBJETIVOS DA PESQUISA

O objetivo mais amplo desta pesquisa é descrever e analisar as marcas sugestivamente relacionadas à gênero no TD, propondo uma estrutura sintática que seja capaz de formalizar o estatuto desses formadores na língua. Para tanto, seguem sistematizados nossos objetivos gerais e específicos:

### **Objetivos gerais:**

1. Descrever os padrões empíricos de comportamento das marcas relacionadas a gênero no TD, a partir de diferentes níveis de análise linguística, como a morfologia, a sintaxe e a semântica;
2. Rediscutir, a partir do comportamento de *mane*, *feto*, *aman* e *inan*, as fronteiras que tradicionalmente separam os processos de formação de palavras de derivação e composição;

### **Objetivos específicos:**

1. Identificar o estatuto formal das peças morfossintáticas que participam das formações com *mane*, *feto*, *aman* e *inan* no TD;
2. Desenvolver uma proposta de estrutura sintática que seja capaz de explicar o estatuto desses formadores do TD.

Em linhas gerais, propomos que uma perspectiva sintática da formação de palavras, como a desenvolvida pela MD (Halle; Marantz, 1993; Marantz, 1997), possa nos fornecer as ferramentas necessárias para atingir os objetivos propostos. Isso porque, como veremos ao longo deste trabalho, uma proposta sintática é capaz de explicar as dificuldades de se estabelecer efetivamente uma fronteira entre derivação e composição, por um lado, e entre composição e sintagmas de outro, o que parece corroborar a ideia de que tais formações são geradas no mesmo componente da gramática. Além disso, em um modelo decomposicional como a MD, em que o estatuto das peças é definido por sua posição na estrutura sintática, podemos levantar a hipótese de que *mane*, *feto*, *aman* e *inan* são licenciados em diferentes lugares na estrutura. Assim, propomos que tais elementos podem ora ocupar uma posição de raiz que, quando categorizada, será associada a um

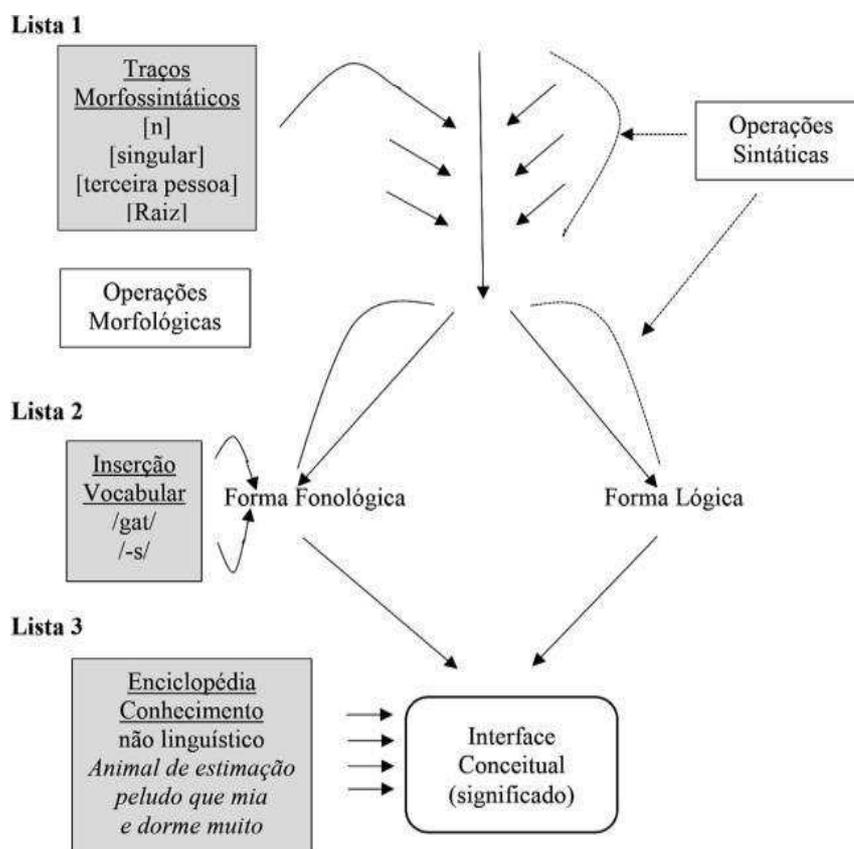
conteúdo conceitual na Lista 3, ora ocupar uma posição de núcleo funcional, atuando como uma peça gramatical da língua na forma de um classificador nominal. Na próxima seção, trazemos, então, uma exposição do quadro teórico que fundamenta o desenvolvimento da pesquisa.

## 1.2 QUADRO TEÓRICO: MORFOLOGIA DISTRIBUÍDA

A MD é um desenvolvimento teórico proposto no âmbito da Teoria Gerativa no início dos anos 1990 por Morris Halle e Alec Marantz. Esse modelo propõe uma arquitetura da gramática em que estruturas complexas, seja no nível da palavra, do sintagma ou da sentença, são geradas no mesmo componente, a sintaxe. Dessa forma, a MD surge como uma alternativa aos modelos lexicalistas, que, em linhas gerais, admitem dois componentes gerativos distintos na gramática: o léxico e a sintaxe.

Na MD, então, as informações que em modelos lexicalistas eram atribuídas ao léxico passam a ser distribuídas ao longo da arquitetura de gramática em três diferentes listas acessadas em momentos distintos da derivação, tal como sistematizado na figura a seguir:

Figura 1 – Arquitetura da gramática da MD



Fonte: Siddiqi, 2009, p. 141 – adaptada por Armelin, 2015.

A Lista 1 abriga os primitivos assumidos no modelo da MD, a saber, raízes e traços morfossintáticos, fornecendo os elementos necessários para a derivação sintática. As raízes são compreendidas no modelo como elementos destituídos de traços morfossintáticos, embora seu conteúdo fonológico, seu funcionamento formal e seu conteúdo semântico ainda esteja em debate na literatura:

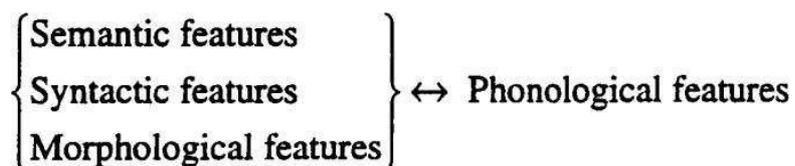
As raízes, um dos mais importantes primitivos teóricos propostos pela Morfologia Distribuída (MD), são alvo de controvérsias dentro do próprio modelo por apresentarem, ou não, três propriedades fundamentais: (i) conteúdo fonológico, (ii) projeção de estrutura argumental e (iii) conteúdo semântico. Autores se dividem ao defenderem ou negarem a presença de conteúdo fonológico nas raízes (cf. BORER, 2003, 2013; ROCHA, 2008; PFAU, 2009); outros discordam sobre o fato de as raízes serem capazes de selecionar e/ou projetar argumentos, em especial o argumento interno (cf. ALEXIADOU, 2014; BASSANI; MINUSSI, 2015; BORER, 2013; HARLEY, 2014), e também há divergência a respeito da existência e do tipo de conteúdo semântico que podem abrigar (cf. ARAD, 2003, 2005; HARLEY, 2014)

(Minussi; Bassani, 2017, p. 154)

Além disso as raízes são desprovidas de categoria sintática, que passa a ser entendida, não mais como uma propriedade do item lexical, mas como consequência da combinação entre núcleos categorizadores e raízes na sintaxe. Os traços morfossintáticos, por sua vez, são consensualmente definidos como abstratos, ou seja, desprovidos de conteúdo fonológico.

A Lista 2, também conhecida como Vocabulário, é acessada apenas pós-sintaticamente e encontra-se no ramo da Forma Fonológica (PF, no inglês *Phonological Form*). Na MD, um componente como a Lista 2 é necessário, uma vez que os traços morfossintáticos não possuem conteúdo fonológico. Mais especificamente, o Vocabulário abriga os chamados Itens de Vocabulário (IVs) que são associações entre os traços e as peças fonológicas disponíveis na língua:

(5) Item de vocabulário



(Halle & Marantz, 1994, p.275)

Os nós terminais sintáticos ganham conteúdo fonológico através da operação de Inserção de Vocabulário, que é, por sua vez, regulada pelo Princípio do Subconjunto:

Princípio do subconjunto: O expoente fonológico de um item de vocabulário é inserido em um morfema no nó terminal se o item corresponder a todos ou a um subconjunto das características gramaticais especificadas no nó terminal. A inserção não ocorre se o item de vocabulário contiver traços não presentes no morfema. Quando vários itens de vocabulário atendem às condições de inserção, deve-se escolher o item que corresponde ao maior número de características especificadas no nó terminal.<sup>5</sup>

(Halle, 1997, p. 128)

<sup>5</sup> "Subset Principle: The phonological exponent of a Vocabulary Item is inserted into a morpheme in the terminal string if the item matches all or a subset of the grammatical features specified in the terminal morpheme. Insertion does not take place if the Vocabulary Item contains features do not present in the morpheme. Where several Vocabulary Items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen." (Halle, 1997, p. 128) – tradução nossa.

Finalmente, a Lista 3, também conhecida como Enciclopédia, é um repositório que armazena o conteúdo conceitual associado às raízes em determinado contexto sintático. Na MD, a presença de um componente como a Lista 3 é necessário, uma vez que o modelo assume a decomposição plena dos objetos sintáticos:

Não há objetos complexos armazenados na memória, ou seja, todo objeto complexo deve ser derivado pela gramática<sup>6</sup>

(Embick, 2015, p. 21)

A decomposição plena implica que não há palavras prontas armazenadas no léxico. Dessa forma, todas as palavras são geradas pela sintaxe e seu conteúdo semântico só pode ser atribuído depois da derivação sintática.

Em suma, conforme sistematizado em Halle e Marantz (1994), três propriedades caracterizam a MD:

- Inserção tardia: os traços morfossintáticos são abstratos e seu conteúdo fonológico é atribuído pós sintaticamente através da operação de Inserção de Vocabulário;
- Subespecificação: para que um Item de Vocabulário seja inserido em nó sintático ele não precisa conter todos os traços especificados naquele nó. Um Item de Vocabulário pode conter um subconjunto dos traços especificados em um nó sintático;
- Estrutura sintática por toda a derivação: os nós terminais em que os Itens de Vocabulário são inseridos são organizados sintaticamente. A sintaxe é o único componente gerativo do modelo capaz de gerar estrutura complexa, seja no nível da palavra, seja no nível da sentença.

A MD abre perspectivas interessantes para o fenômeno empírico em investigação nesta pesquisa. Nas formações de gênero do TD, por exemplo, há um debate a respeito de tais elementos serem formados via composição ou derivação. Tal hesitação se coloca, justamente, porque a fronteira entre os dois processos não

---

<sup>6</sup> “No complex objects are stored in memory, i.e., every complex object must be derived by the grammar.” (Embick, 2015, p. 21) – tradução nossa.

é totalmente clara. Uma abordagem como a MD nos permite decompor as peças envolvidas na formação de palavras e derivar seu comportamento a partir da concatenação entre tais peças no componente sintático.

Partindo dessa breve descrição das propriedades do modelo teórico, sintetizamos, na próxima seção, as questões centrais da pesquisa e as hipóteses levantadas neste trabalho.

### 1.3 QUESTÕES DE PESQUISA E HIPÓTESES LEVANTADAS

O gênero no TD é realizado através dos marcadores *mane*, *feto*, *aman* e *inan* que são realizados linearmente após os substantivos. Há uma discussão na literatura a respeito do processo de formação de palavras responsável por gerar as formas marcadas em gênero na língua, se via derivação ou composição. Nosso ponto de partida nesta pesquisa é a hipótese de que uma análise dessas formações deve, no entanto, levar em conta não apenas o estatuto formal desses marcadores, mas também o ambiente sintático em que tais elementos são inseridos.

A partir daí, o recorte proposto neste trabalho contempla as seguintes questões acompanhadas das hipóteses a serem desenvolvidas nos próximos capítulos:

- a) A marcação relacionada a gênero no TD apresenta propriedades de derivação ou composição?

**Hipótese:** tomando por base as propriedades empíricas que caracterizam o processo de derivação e composição sintetizadas em Gonçalves e Andrade (2016), propomos que a marcação de gênero no TD apresenta comportamento instável, ora se assemelhando à composição, ora à derivação a depender do critério empírico que está sendo adotado na literatura. Isso, por um lado, explica o debate existente na literatura a respeito do estatuto dessas formas e aponta para a plausibilidade de um modelo como a MD, que trata tanto a derivação, como a composição na sintaxe, diferenciando-as partir do estatuto das peças envolvidas, bem como de sua altura de anexação na estrutura.

- b) Qual o estatuto das marcas interpretadas como gênero do TD?

**Hipótese:** propomos que os formadores *mane*, *feto*, *aman* e *inan* do TD estão em processo de gramaticalização. Mais especificamente, propomos que, nas formas em que são interpretados como gênero, tais elementos passaram de substantivos independentes a classificadores nominais. O estatuto gramaticalizado dessas marcas pode ser identificado em algumas funções morfossintáticas que tais elementos assumem na língua. Além disso, o estatuto de classificador de *mane*, *feto*, *aman* e *inan* é compatível com o fato de seu emprego não ser obrigatório, de modo que tais elementos não se aplicam a todos nomes da língua.

- c) Tomando os marcadores relacionados a gênero no TD como primitivos da gramática, tais elementos apresentam o estatuto de raízes ou núcleos funcionais?

**Hipótese:** a partir de um modelo decomposicional como a MD, em que o estatuto das peças é definido por sua posição na estrutura sintática, propomos que *mane*, *feto*, *aman* e *inan* são licenciados em diferentes lugares na estrutura, de modo a ter um estatuto duplo no sistema. Assim, tais elementos podem ocupar uma posição de raiz que, quando categorizada, será associada a um conteúdo conceitual na Lista 3. Da mesma forma, tais elementos podem ocupar uma posição de núcleo funcional, atuando como uma peça gramatical da língua. A nossa ideia, portanto, é que *mane*, *feto*, *aman* e *inan* estão em um processo de gramaticalização, passando de raízes para núcleos funcionais.

- d) Qual a estrutura sintática das formas marcadas com *mane*, *feto*, *aman* e *inan* do TD?

**Hipótese:** considerando que os elementos podem ser anexados na estrutura sintática de duas maneiras distintas: como núcleos ou como adjuntos, propomos que os classificadores nominais *mane*, *feto*, *aman* e *inan* do TD, são sintaticamente adjuntos. Assim, esses elementos são concatenados à estrutura, mas não por requerimentos de traços morfossintáticos. Eles atuam, portanto, como modificadores da estrutura a que se anexam, sem projetar seu rótulo na estrutura sintática.

## 1.4 ORGANIZAÇÃO DA PESQUISA

Para atingir os objetivos propostos e desenvolver as hipóteses delineadas neste capítulo, esta pesquisa se divide em mais cinco capítulos, além desta apresentação geral.

No capítulo 2 buscamos descrever alguns aspectos relacionados ao gênero no TD, levando em consideração, especialmente, o intenso contexto de contato linguístico em que o Timor-Leste historicamente está inserido. Nessa discussão, focamos, em especial, no contato entre o TD e o português que tem consequências significativas na gramática da língua, o que é especialmente interessante nas discussões sobre gênero desenvolvidas nesta pesquisa.

No capítulo 3, sistematizamos e descrevemos uma série de propriedades dos marcadores *mane*, *feto*, *aman* e *inan* do TD. Para tanto, tomamos como ponto de partida as gramáticas de Hull e Eccles (2005) e Costa (2015), buscando verificar como tais formações são descritas dessa perspectiva. Além disso, de uma perspectiva mais linguística, nos baseamos na descrição comparada de Choupina (2011) e de Belo e Oliveira (2021), que contrastaram o funcionamento das marcas interpretadas como gênero em TD e em português. Finalmente, adicionamos também outras propriedades empíricas relevantes notadas no próprio percurso da pesquisa.

No capítulo 4, revisitamos as propriedades mais tradicionalmente associadas aos processos de composição e derivação com base na sistematização fornecida em Gonçalves e Andrade (2016). A partir daí buscamos verificar, empiricamente, em que medida os dados de gênero no TD se aproximam ou se afastam das características relevantes para cada um desses processos. A nossa percepção é a de que as marcas sugestivas de gênero do TD podem ser classificadas ora como composição, ora como derivação a depender do critério aplicado, o que explica o debate saliente na literatura a respeito dessas formações.

No Capítulo 5, desenvolvemos a nossa proposta de que as propriedades dos formadores *mane*, *feto*, *aman* e *inan* do TD são mais bem compreendidas diante da hipótese de que tais elementos atuam como classificadores nominais. Dessa forma, localizamos esses formadores em um contínuo de gramaticalização em que elementos com uma origem lexical passam a integrar o sistema gramatical da língua,

funcionando como um dispositivo de classificação nominal. Além disso, para colocar em debate a estrutura sintática dessas formações, tomamos como base o modelo da MD (Halle; Marantz, 1993; Marantz, 1997) e propomos uma formalização inicial dos classificadores nominais do TD, tratando-os sintaticamente como a realização de núcleos funcionais adjuntos

Por fim, o Capítulo 6, encerra esta dissertação com algumas considerações finais e as perspectivas futuras que foram abertas a partir do percurso desta pesquisa.

## **CAPÍTULO 2: GÊNERO NO CONTEXTO DE CONTATO LINGUÍSTICO DO TIMOR-LESTE**

Neste capítulo, buscamos descrever alguns aspectos relacionados ao gênero no TD, levando em consideração, especialmente, o intenso contexto de contato linguístico em que o Timor-Leste historicamente está inserido.

Como apontam Hajek e Williams-Van Klinken (2019), a chegada dos espanhóis e portugueses à região do Timor-Leste no final dos séculos XV e XVI deu início a um longo período de contato entre suas respectivas línguas românicas e uma ampla gama de línguas austronésias. Mais especificamente, durante o período colonial português no Timor-Leste, que dura até 1975, o português foi a língua exclusiva das escolas e das instituições oficiais, o que aponta para a relevância desse contato. É interessante ressaltar ainda que o TD se abre como um campo interessante de investigação, considerando que divide com o português o estatuto de língua oficial do Timor-Leste. Isso aponta para o fato de que o contato entre as duas línguas ainda permanece intenso até os dias atuais.

Em relação à marcação de gênero, esse contato é relevante principalmente considerando que o TD é uma língua isolante que não apresenta um sistema propriamente gramaticalizado de gênero, diferentemente do português, que distingue duas classes de gênero em seu sistema gramatical, masculino e feminino. Como observa Stolz (2012), é bastante comum que o contato entre línguas com sistemas de gênero distintos afete, de alguma forma, a expressão do gênero nessas línguas.

Nessa perspectiva, Hajek e Williams-Van Klinken (2019), investigando especificamente o gênero no TD a partir das questões de contato linguístico, propõem que há indícios de um sistema incipiente de gênero na língua que pode ser visto, por exemplo, em um processo de atribuição de gênero a algumas palavras que não são empréstimos do português.

Para desenvolver uma discussão dessa natureza, este capítulo está dividido da seguinte maneira: na seção 2.1, trazemos um breve panorama histórico do TD no âmbito da diversidade linguística apresentada pelo contexto timorense; na seção 2.2, focamos especificamente nas influências e contatos linguísticos que afetam o TD; na seção 2.3, apresentamos uma caracterização geral de aspectos contrastivos entre a gramática do TD e do português; na seção 2.4 abordamos, mais especificamente

questões envolvidas na expressão do gênero a partir do contato linguístico entre o TD e o português. Por fim, na seção 2.5, concluímos o capítulo com algumas considerações finais.

## 2.1 TÉTUM DÍLI E O CONTEXTO TIMORENSE: UM BREVE PANORAMA HISTÓRICO

O Timor-Leste é um dos países mais jovens do mundo e se localiza no sudeste asiático. Proclamou a sua independência de Portugal em 28 de novembro de 1975, mas na época ficou sob ocupação indonésia, restaurando oficialmente a sua independência apenas em 20 de maio de 2002, como República Democrática de Timor-Leste. De acordo com a Constituição do país, o tétum é a sua língua nacional, juntamente com a língua portuguesa, que também é considerada língua oficial do país. No entanto, como salienta Paulino (2023, p. 1):

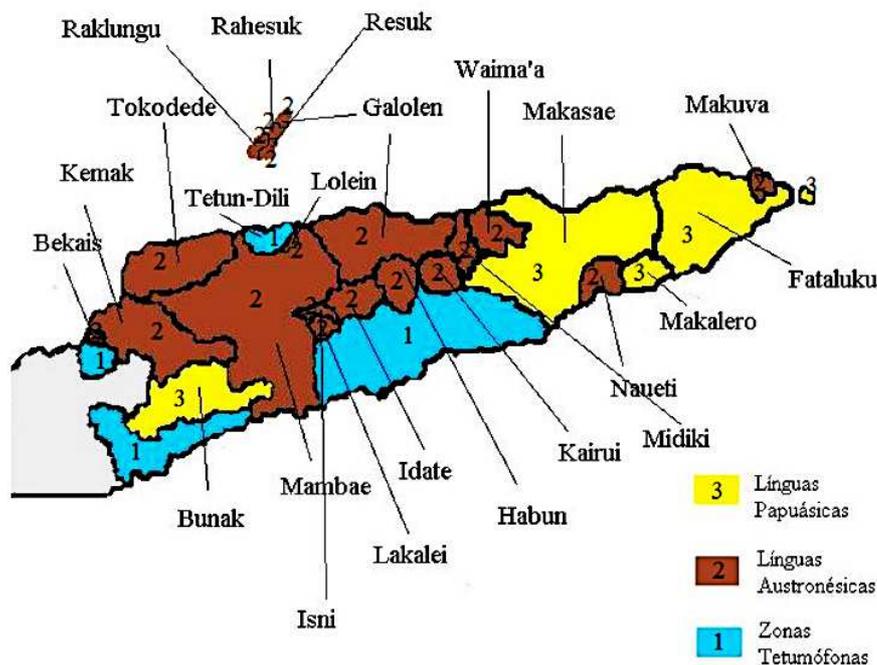
O Timor-Leste é um país que, desde o início, tem sua própria identidade cultural, identidade religiosa e identidade linguística. No que diz respeito à identidade linguística, o Timor-Leste é um país bilíngue e multilíngue.

Linguisticamente, essa jovem República possui dezesseis línguas nacionais faladas em diferentes regiões do país, que contém cerca de 1,34 milhões de habitantes. Ramos e Carmo (2020, p. 447-448), a partir da proposta de Hull (2002), agrupam tais línguas da seguinte maneira:

Essas línguas são agrupadas em dois grandes grupos. Doze de origem austronésia (tétum, habun, kawaimina, idalaka, galoli, wetar, bekais, dawan, mambae, kemak, tokodede, lovaia) e quatro de origem papua (makasae, makalero, bunak e fataluco).

Segundo, Albuquerque (2011) essas línguas estão distribuídas pela faixa territorial do Timor-Leste se acordo com o mapa a seguir:

Figura 2 – Distribuição das línguas no Timor-Leste



Fonte: Albuquerque (2011, p. 3)

Nesse sentido, é bastante comum encontrarmos no Timor-Leste falantes que dominam mais de uma língua. De acordo com Paulino (2023, p. 3):

Em Timor-Leste encontram-se pessoas a falar duas ou mais línguas com muita frequência. O mais visto é um timorense que fale fluentemente duas línguas, isto é, fala a sua língua materna e a língua cooficial, o tétum. Significa que os timorenses, em sua maioria, estão na categoria linguística de bilinguismo, contudo, alguns pertencem à categoria linguística de multilinguismo.

Partindo dessa diversidade linguística, observa-se um forte apego intergeracional às línguas maternas e às suas respectivas culturas tradicionais. Por outro lado, o TD é falado por uma quantidade significativa da população, difundindo-se em uma área vasta do território timorense.

Historicamente, o tétum já havia se estabelecido como língua franca antes da chegada dos portugueses à ilha de Timor no século XVI. O primeiro tétum falado naquela época era o tétum terik, utilizado como meio de comunicação para trocas comerciais (Hull, 2001 apud Albuquerque e Taylor-Leech, 2012, p. 156). Hoje, o tétum terik permanece falado na fronteira entre o Timor e a Indonésia e nas zonas mais interioranas do país, como Manatuto, Viqueque, Manufahi e Cova-Lima.

Por sua vez, o TD é a língua que está ligada à identidade nacional no Timor-Leste. Um dos fatores que contribuiu para isso é o fato de o TD ter sido empregado no período colonial português. Além disso, como apontado por Albuquerque (2003, p.317), o TD tornou-se também uma língua de resistência, uma vez que essa foi a língua que se manteve durante a ocupação indonésia, quando o português foi proibido. Dessa mesma forma, como salientam Ramos e Carmo (2020, p. 148):

[..] a diversidade linguística motivou o surgimento de uma língua que os falantes de diferentes grupos étnicos usam para se comunicarem entre si. Essa língua é o chamado tétum-praça (o dialeto de Díli).

Em conformidade com esse processo histórico, o uso padronizado do tétum é o TD (ou tétum-praça), designado como língua oficial do país. Nesse contexto linguístico-cultural, destaca-se a relevância da atuação do Instituto Nacional de Linguística (INL) da Universidade Nacional de Timor Lorosa'e (UNTL) criado em meados de 2001. A política de planejamento linguístico do INL se desenvolve em torno de três objetivos principais, tendo como foco a identidade linguística nacional:

(1) O desenvolvimento, modernização e disseminação do Tétum-Praça; (2) A preservação e promoção das demais línguas nacionais/ locais; (3) Subsequentemente corroborada pelo bilinguismo oficial da RDTL [República Democrática de Timor-Leste], a estratégica parceria linguística Tétum-Português.<sup>7</sup>

Em março de 2002, a Constituição da República Democrática de Timor-Leste, oficializa, em seu artigo 13º, o tétum como língua nacional do país, juntamente com o português. O padrão ortográfico do tétum, por sua vez, é estabelecido pelo decreto 1/2004, que determina a adoção do padrão oficial como sendo aquele desenvolvido pelo INL. Assim, Albuquerque e Taylor- Leech (2012, p. 158) afirmam que:

O órgão responsável para regular a alteração no corpus do tétum foi o INL (Instituto Nacional de Linguística) cuja primeira tarefa foi escolher uma variedade e padronizar a ortografia. A variedade selecionada foi o tétum praça falada em Díli, capital de Timor-Leste, e a ortografia foi padronizada após uma análise histórica das diferentes propostas ortográficas existentes para a língua e publicada pela instituição em INL (2002).

---

<sup>7</sup> Retirado de <http://untl.edu.tl/pt/centros/instituto-nacional-de-linguistica>. Acessado em 06/07/2024.

Embora o português seja a língua cooficial do Timor-Leste, junto com o TD, seu emprego está correlacionado a um grau mais alto de formalidade, como salientam Hajek e Klinken (2019, p. 65):

Enquanto o tétum e o português são línguas cooficiais, o último tem status formal mais alto, dado seu papel como a língua principal para a elaboração de leis e interpretação legal, e seu alto status dentro do sistema educacional. O grau de influência portuguesa no tétum-Díli varia não apenas pelo contexto, mas também varia amplamente entre os falantes. Os falantes do tétum-Díli estão cientes da proporção de empréstimos portugueses usados por outros, e tendem a associar julgamentos sociolinguísticos a isso [...].<sup>8</sup>

Durante o período da colonização portuguesa, de 1515 a 1974, a língua portuguesa foi considerada a língua da administração, ao lado das línguas nacionais faladas em Timor. Situação similar ocorreu com a língua indonésia durante a ocupação indonésia, de 1975 a 1999, quando o indonésio foi imposto como língua oficial no Timor-Leste e o uso do português foi proibido. Nesse contexto de contato, destaca-se também o fato de que a língua inglesa foi introduzida no país entre 1999 e 2000, devido à presença de organizações internacionais que utilizam o inglês como língua de trabalho, como agências de ajuda e desenvolvimento de diversos países, o Banco Mundial, embaixadas e sedes de ONGs internacionais. Após a restauração da independência do país em 20 de maio de 2002, a Constituição da República Democrática de Timor-Leste (C-RDTL), no Artigo 13.º, adotou oficialmente o português como língua oficial e de escolarização, juntamente com o tétum. Além disso, essa constituição, no Artigo 159.º, também reconheceu o inglês e o indonésio como línguas de trabalho (ver Carneiro, 2010; Hajek, 2017) empregadas na administração pública, juntamente com as línguas oficiais, quando necessário.

Trata-se, portanto, de um cenário geográfico, político e histórico privilegiado de contato linguístico no qual o Timor-Leste se encontra inserido, com influências importantes na constituição do TD, com exploramos na próxima seção.

---

<sup>8</sup> “While Tetun and Portuguese are co-official languages, the latter has higher formal status, given its role as the primary language for law-making and legal interpretation, and its high status within the education system. The degree of Portuguese influence in Tetun Dili varies not only by context, but also varies widely amongst speakers. Tetun Dili speakers are aware of the proportion of Portuguese loans used by others, and tend to associate sociolinguistic judgments [..].” Hajek e Klinken (2019, p. 65) – tradução nossa.

## 2.2. CONTATOS E INFLUÊNCIAS LINGUÍSTICAS NO TÉTUM-DÍLI

Um dos aspectos que desencadeiam um intenso contexto de contato linguístico no Timor é o fato desse território integrar a rota comercial do sudeste asiático, o que propicia o contato com diferentes povos, línguas e culturas. Dessa forma, o TD é naturalmente resultado de uma longa história de contatos linguísticos. Na sistematização fornecida por Albuquerque (2011, p.4-5), o autor traz um resumo desses contatos, associados a possíveis momentos históricos em que eles ocorreram:

- Em tempos pré-históricos ocorreu um contato entre os povos papuásicos, que já habitavam a ilha de Timor, com o povo que realizou a primeira migração austronésia. Este longo contato linguístico entre as línguas austronésias e papuásicas em Timor foi o fator mais importante para a formação de uma área linguística (HULL, 2001);
- A região central de Timor-Leste apresenta indícios de habitação de povos pré-austronésios distintos dos papuásicos, que já habitavam a ilha, e deram origem às línguas papuásicas. Desta maneira, as línguas da região central do país – Mambae, Tokodede e Kemak – apresentam a influência de substratos distintos que não são de origem austronésia, nem papuásica;
- A segunda migração austronésia (de origem ambônica) causou uma série de empréstimos e simplificações gramaticais, além disso, outro fato interessante é que esse povo foi totalmente assimilado à população que já habitava a ilha;
- O contato com os comerciantes de diversas origens étnicas deve ter se realizado com o uso do Bazar Melayu (VERSTEEGH, 2008, p. 175). Este fator causou uma série de empréstimos e outro processo de simplificação de algumas estruturas gramaticais, principalmente no Tétum (HULL, 2005), mas também nas demais línguas faladas em Timor-Leste (HULL, 2001);
- No século XV o reino tetumófono de Wehale iniciou sua expansão e dominação dos demais reinos. Desta forma, a língua Tétum começou a ser aprendida como a língua do reino dominador. Assim, a aquisição não nativa (MCWHORTER, 2008) do Tétum causou mais uma nova série de simplificações gramaticais;
- A mudança da capital do Timor português, no ano 1769, de Lifau – hoje a região do enclave de Oecussi – para Díli, que era uma região falante de Mambae, causou mais uma série de simplificações gramaticais;
- Nos últimos anos, o Tétum-Praça, em sua variedade Tétum-Díli, como língua urbana vem recorrendo à língua portuguesa para

realizar um grande número de empréstimos relativos ao mundo moderno.

No recorte desta pesquisa, nos focamos, mais especificamente, nas várias influências do português que caracterizam o TD ao longo do seu desenvolvimento histórico. Assim, o TD exhibe nos itens lexicais uma influência significativa do português, especialmente visível nos empréstimos ocorridos durante os tempos coloniais, mas que continuam acontecendo produtivamente até hoje. Para além disso, as consequências desse contato podem ser vistas nos mais variados níveis linguísticos, como apontado em Klinken et al (2002, p. 56) e Hajek (2017, p. 168-173) em que os pontos sistematizados abaixo foram consultados:

Fonologicamente, o TD possui vinte e dois fonemas consonantais e onze são emprestados do português (/p g v z ʃ ʒ λ ŋ r j/ e parcialmente /w/, com algum reforço do malaio). A influência do português é bastante acentuada na fonotática e nos processos fonológicos do TD. Um exemplo disso é o processo de palatalização que acontece nos empréstimos do português e mesmo em termos nativos. Por exemplo, na palavra emprestada do português: /festa/ [festa] ~ [feʃta] ('festa') e na palavra nativa do TD /has/ [ha:s] ~ [ha:ʃ] ('manga').

Algumas combinações mais fixas de itens lexicais são inclusive emprestadas do português com plena concordância gramatical, como, por exemplo em *Nasoens Unidas* ('Nações Unidas'), *primeiru ministru* ('primeiro-ministro'), *primeira klase* ('primeira classe'). Alguns empréstimos lexicais do português aparecem opcionalmente junto com itens lexicais nativos do tétum, como em *merkadu tuan* ('antigo mercado'); *eskola tuan* ('antiga escola'), em que o adjetivo (*tuan*) é item nativo, enquanto os substantivos que ele acompanha são empréstimos.

Morfologicamente, o TD parece ser mais resistente a empréstimos. Como em TD os verbos não são flexionados, os verbos oriundos do português são conjugados apenas na forma da terceira pessoa/singular do presente. Da mesma forma, há um empréstimo limitado de substantivos e adjetivos com marcação de gênero e número do português. Assim, um pequeno conjunto de substantivos e adjetivos de origem portuguesa são obrigatoriamente marcados por gênero por todos os falantes de TD. Alguns exemplos de nomes com esse padrão são: *tiu/tia* ('tio/tia'), *primu/prima* ('primo/prima'), *mestre/mestra* ('mestre/mestra'), *profesór/profesora* ('professor/professora'). Um exemplo de adjetivo com essa natureza pode ser visto no par

*bunitu/bonita* ('bonito/bonita'). É interessante apontar que a concordância de gênero feminino é variável e evitada por muitos falantes de TD em adjetivos que acompanham substantivos femininos portugueses, por exemplo: *primeira faze x faze primeiru* ('primeira fase'); *desizaun kritika/kritiku* ('decisão crítica'); *Uniaun Europea/Europeu* (União Europeia).

Quanto à marcação de número com o emprego de -s nos empréstimos, tal marcação é obrigatória em formas fixas, como vimos em *Nasoens Unidas* ('Nações Unidas'), que aparece na forma feminina e no plural da língua de origem. Em outros contextos, no entanto, o emprego de -s em empréstimos é opcional, não sendo particularmente empregado na comunicação cotidiana, tornando-se mais frequente nos registros escritos. É interessante ressaltar ainda que o -s de plural em português coincide fonologicamente com o início da forma que marca plural em TD, o elemento pós-nominal *sira*, por exemplo em *profesores kontratadus sira/ profesór kontratadu sira* ('professores contratados'). Nota-se, além disso que pode haver simultaneamente o emprego do plural oriundo de empréstimo e do plural nativo na mesma formação.

Ainda segundo Klinken et al (2002, p. 56) e Hajek (2017, p. 168-173), o único afixo que parece ter sido verdadeiramente nativizado em TD é o sufixo agentivo *-dór*, que se refere a "alguém que habitualmente faz algo", sendo, muitas vezes, pejorativo. Por exemplo, em *hemu* ('beber') > *hemudór*, que remete ao significado de 'alguém que gosta de beber (m/f)'. Esse afixo pode também ser concatenado a uma raiz nativa, sem realizar marcas de número ou gênero.

Sintaticamente, a literatura aponta diversas influências advindas do contato com o português. Dentre elas, o TD parece estar mostrando uma mudança geral na justaposição de sentenças. Mais especificamente, o TD, que apresenta uma tendência de encadeamento de sentenças sem conectores (como as conjunções) vem passando a empregar tais elementos devido à influência do português. Essa característica é especialmente notável em registros formais.

Ainda que a influência do português seja bastante significativa, é importante destacar que o grau de emprego de itens linguísticos do português é particularmente marcado pelo estatuto formal. Como salientam Hajek e Williams-Van Klinken (2019), há uma variação bastante grande entre os próprios falantes de TD em relação ao grau de influência do português, que está associado a um julgamento sociolinguístico elitizado. Segundo Ross (2017, p. 56), o emprego massivo de empréstimos do

português confere ao discurso a ideia de que o indivíduo é pertencente à “elite, [é] escolarizado e provavelmente bem-empregado, mas distante e desconectado da população timorense geral”<sup>9</sup>.

Sistematizando o grau de influência do português no uso do TD, a literatura (ver Hull e Eccles, 2001), emprega uma distinção tripartida entre as variedades acroletal, mesoletal e basiletal. Ao definir essas variedades, Hajek e Williams-Van Klinken (2019) apontam que a variedade acroletal é empregada pela elite tradicional, escolarizada em português. Os indivíduos dessa variedade são caracterizados por falarem português fluentemente e por terem, no geral, o português como língua preferida na escrita. Essa elite exerce grande influência sobre o tétum usado na mídia, como televisão, rádio e imprensa, modelando assim sua variedade de TD para um público mais amplo. Com base em uma avaliação de corpus, Hajek e Williams-Van Klinken (2019) estimam que a proporção de elementos em português nessa variedade pode facilmente atingir 35% ou mais. Da mesma forma, em artigos de jornal e textos técnicos, essa porcentagem pode ultrapassar os 40%. No outro extremo, a variedade basiletal do TD mostra pouca influência do português, refletindo uma tendência geral de menor alfabetização, ausência de conhecimento do português e pouco uso de termos técnicos e registros formais. Na estimativa de Hajek e Williams-Van Klinken (2019), a proporção de empréstimos do português nessa variedade ficaria em torno de 5%. No meio desses dois extremos está, por sua vez, a variedade mesoletal, usada pelos residentes da capital Díli que não são considerados parte da elite. Na perspectiva de Hajek e Williams-Van Klinken (2019), a porcentagem de *tokens* em português na fala urbana informal do mesoleta varia de 10% a 25%, embora essa proporção aumente significativamente quando tais indivíduos falam sobre assuntos mais técnicos.

A partir dessa visão mais geral, na próxima seção tratamos, com base em Klinken, Hajek e Nordingler (2002) e em Hajek e Williams-van Klinken (2019), de aspectos linguísticos que colocam em comparação a gramática do TD e do português.

---

<sup>9</sup> “elite, educated, and probably well-employed, but distant and disconnected from the general Timorese population”. (Ross, 2017, p. 56 – tradução nossa).

## 2.3 GRAMÁTICAS EM CONTRASTE: TÉTUM-DÍLI E PORTUGUÊS

Sintaticamente as sentenças do TD apresentam a ordem SVO como *default*, exceto em sentenças de apresentação existencial e em formações com alguns verbos intransitivos, como *mosu* ('aparecer'), em que o verbo pode preceder o único argumento devido à influência do português. Dessa forma, sujeito e objeto são reconhecidos geralmente pela sua posição na sentença, com o sujeito precedendo o verbo, enquanto o objeto segue o verbo.

(6)

- a. Ami lori kuda ida.  
 1p. bring horse one  
 'Nós trouxemos um cavalo'

(Klinken, Hajek e Nordingler, 2002, p. 51)

Há, no entanto, a possibilidade de que o objeto seja fronteado na sentença. Segundo Williams-van Klinken, Hajek e Nordingler (2002) esse fronteamento pode estar relacionado à leitura contrastiva, sendo comumente o sintagma nominal fronteado marcado com o focalizador *mak*, embora tal marcação não seja obrigatória.

(7)

- a. Só deit lingua né mak hau la hatene.  
 only just language this Foc 1s not know  
 'Era só a língua que eu não conhecia.'

(Klinken, Hajek e Nordingler, 2002, p. 54)

O objeto também pode ser fronteado na língua para marcação de um tópico discursivo, como no exemplo a seguir:

(8)

- a. Nán ita koa halo kabelak bo-bót; depois ita tetak liu deit.  
 Meat 1pl cut make flat.segment RDP-big afterwards 1PI tenderise more just  
 'Cortamos a carne em pedaços grandes e depois a amaciamos ainda mais'

(Klinken, Hajek e Nordingler, 2002, p. 54)

O TD não possui um sistema de voz, não havendo construção passiva na língua. Existem, no entanto, várias estratégias linguísticas para se criar efeitos

semelhantes à passivização. Segundo Williams-van Klinken, Hajek e Nordingler (2002) um deles é o emprego de um prefixo destransitivizador (9a), o segundo é diminuir a proeminência do sujeito omitindo-o ou usando um *ema* genérico ('pessoa, pessoas, alguém'), como (9b). Além disso, a proeminência do objeto pode ser destacada através de uma operação de fronteamento desse elemento (9c).

(9)

- a. Rai komesa nak-doko.  
earth start INTR-shake  
'A terra começou a tremer.'

(Klinken, Hajek e Nordingler, 2002, p. 20)

- b. Ema sunu bispu nia uma.  
person bum bishop POS house  
'Pessoas/alguém queimou a casa do bispo.'

- c. Balu mate, balu kikoan sira né ema sama.  
some die some little PL this person tread.on  
'Algumas (pessoas) morreram, algumas - as crianças - foram pisoteadas.'

(Klinken, Hajek e Nordingler, 2002, p. 52)

É interessante ressaltar, no entanto, que, segundo Williams-van Klinken, Catharina e John Hajek (2018) uma voz passiva desenvolveu-se recentemente no acrolecto devido ao contato linguístico com o português.

Morfologicamente, os verbos no TD não são marcados para tempo, aspecto ou modo. Tais distinções podem, em vez disso, ser reveladas por meio de uma série de marcadores de que a língua dispõe, como o perfectivo *tiha* em (10a), ou o marcador de anterioridade *ona* ('já') em (10b).

(10)

- a. Lapizeira monu tiha hosi meza.  
pen fall PRF from table  
'a caneta caiu da mesa'

- b. Ba sira nebé ke iha SMP ona, sira hatene ona bahasa Indonesia.  
for 3P REL REL LOC junior.high ANT, 3P know ANT language Indonesia  
'Para aqueles que estão no ensino fundamental, eles já sabem indonésio'

(Klinken, Hajek e Nordingler, 2002, p. 74-75)

Elementos dessa natureza, no entanto, não são gramaticalmente obrigatórios, estando ausentes em muitas sentenças na língua. Além disso, no TD, não há marcação de pessoa-número nos verbos.

Já no português, embora as sentenças geralmente apresentem a ordem SVO, há uma liberdade maior no ordenamento das palavras. Isso ocorre, segundo Hajek e Williams-Van Klinken (2019), porque os verbos são marcados com as informações de pessoa e o número do sujeito. Além disso, os adjetivos predicativos também concordam em gênero e número com o sujeito. Pronomes de terceira pessoa e determinantes também são flexionados em gênero e número, permitindo que os participantes sejam frequentemente distinguidos sem a necessidade de uma ordem de palavras estrita. Adicionalmente tempo e aspecto são marcados dentro do paradigma verbal.

No âmbito do sintagma nominal, no TD, a maioria dos modificadores segue o substantivo, com exceção dos possessivos e dos numerais do português, bem como de alguns outros empréstimos. Não há uma morfologia flexional nos sintagmas nominais da língua, ou seja, tradicionalmente, não há marcação de gênero ou número em substantivos ou adjetivos. Por outro lado, nos sintagmas nominais em português, os adjetivos normalmente seguem o substantivo, embora possam precedê-los em alguns casos. Além disso, há concordância de gênero e número entre modificadores e o substantivo no sintagma nominal. Alguns desses contrastes podem ser vistos a seguir<sup>10</sup>:

(11)

- a. Horiseik ha'u-nia organizasaun sosa motór kiik rua.  
yesterday 1SG POSS organization buy motorbike small two  
'Ontem minha organização comprou duas motos pequenas'
- b. Ontem a minha organização comprou  
yesterday DEF.F 1SG.POSS.F organization.F buy.3SG.PST  
duas motos pequenas.  
two.F motorbike.F-PL small.F-PL  
'Ontem minha organização comprou duas motas pequenas'

(Hajek e Williams-van Klinken, 2019, p.68)

---

<sup>10</sup> As glosas em inglês são retiradas da literatura citada, enquanto as traduções de português apresentadas são nossas.

O TD possui uma morfologia derivacional limitada. Muitos dos afixos encontrados em tétum-terik foram perdidos (van Klinken 1999), não ocorrendo em TD ou apresentando uma ocorrência fossilizada apenas em algumas palavras da língua. Ao sistematizar a morfologia derivacional no TD, Klinken, Hajek e Nordingler (2002) destacam os seguintes formadores:

- Prefixo *ha-*: deriva verbos causativos a partir de verbos intransitivos e adjetivos

(12)	Verbo intransitivo/Adjetivo	Verbo transitivo
	lakon 'disappear'	ha-lakon 'cause to disappear, abolish'
	manas 'hot'	ha-manas 'heat'
	moe 'ashamed, embarrassed'	ha-moe 'shame, embarrass'
	sai 'exit'	ha-sai 'remove, extract'

(Klinken, Hajek e Nordingler, 2002, p.7)

- Prefixo *nak-*: que, como vimos em (9a), deriva, a partir de um verbo transitivo, um verbo intransitivo com sujeito sintático que tem papel de tema.
- Sufixo *-dór*: anexado a um verbo X para derivar um substantivo que significa 'pessoa que habitualmente faz X'. Trata-se de um sufixo português, mas que ocorre produtivamente em palavras do TD.

(13)	Verbo	Nome
	dukur 'sleep'	dukurdór 'sleepyhead, person who sleeps a lot'
	hemu 'drink'	hemudór 'drinker (of alcohol)'
	servisu 'work'	servisudór 'one who is diligent, hard-working (but not necessarily competent)'

(Klinken, Hajek e Nordingler, 2002, p.20)

Além disso, Klinken, Hajek e Nordingler (2002) ainda destacam a existência no TD de alguns processos de formação de palavras, como a derivação zero, a reduplicação e a composição.

Em contraste, o português possui uma rica morfologia derivacional. Mesmo os timorenses que não falam português estão familiarizados com vários afixos

portugueses devido ao empréstimo de conjuntos de palavras relacionadas, especialmente no léxico técnico e abstrato (Hajek e Williams-van Klinken, 2019), como no exemplo a seguir:

(14)	Português	Tétum-Díli
	administrador	administradór
	administração	administrasaun
	administrativo/a	administrativu/a

(Adaptado, Hajek e Williams-van Klinken, 2019, p.68)

No sistema pronominal, o TD distingue entre singular e plural nas peças disponíveis no seu conjunto de pronomes pessoais, como ilustrado a seguir:

Tabela 1 – Pronomes pessoais<sup>11</sup>

Pronomes pessoais		
Pronome	Glosa	Descrição
hau	1s	'I, me'
o	2s	'you (singular)'
ita	2s.HON	'you (singular, respectful)'
nia	3s	'he, she, him, her, it'
ita	1pl	'we, us (including addressee)'
ami	1pE	'we, us (excluding addressee)'
imi	2p	'you (plural)'
sita	3p	'they (plural)'

Fonte: Klinken, Hajek e Nordingler (2002, p.26)

Não há, no entanto, qualquer distinção de gênero no sistema pronominal do TD. O português, por sua vez, além realizar uma distinção de número no seu sistema pronominal, apresenta também marcas distintas para gênero na terceira pessoa.

A partir dessa visão mais geral do sistema linguístico do TD em contraste ao português, tratamos, na próxima seção, mais diretamente de aspectos envolvidos na expressão do gênero que são relevantes na perspectiva de contato entre as duas línguas.

<sup>11</sup> Os dados na tabela 1, retirados de Klinken, Hajek e Nordingler (2002, p.26), pertencem, mais propriamente ao Tétum-DIT.

## 2.4 GÊNERO E CONTATO LINGUÍSTICO: TÉTUM-DÍLI E PORTUGUÊS

O contato linguístico pode ter um impacto significativo nos sistemas de gênero das línguas envolvidas. Segundo Stolz (2012, p. 95), quando línguas com diferentes sistemas de gênero entram em contato, é comum que tais sistemas sofram algum tipo de modificação, com sistemas de gênero existentes podendo recuar, desaparecer ou serem reformulados. Em alguns casos, línguas que anteriormente não apresentavam sinais de gênero gramatical podem desenvolver marcação de gênero devido ao empréstimo de línguas nas quais o gênero é gramaticalizado.

Diversas línguas austronésias ilustram consequências interessantes do contato com línguas que apresentam algum sistema de gênero. Explorando exemplos dessa natureza, Hajek e Williams-Van Klinken (2019) apontam que o indonésio é uma das línguas menos afetadas em relação ao gênero nesse processo de contato, tendo adotado algumas formas masculinas e femininas de nomes emprestados do sânscrito, árabe e holandês, refletindo distinções de gênero semântico. O Rapanui, por sua vez, falado na Ilha de Páscoa no Pacífico oriental, introduziu distinções de gênero natural ao adotar o diminutivo sensível ao gênero do espanhol *-ito/-ita*. Ainda segundo Hajek e Williams-Van Klinken (2019), algumas línguas austronésias vão além do empréstimo de palavras ou morfemas, adquirindo um sistema de gênero parcial. Na análise do Tagalog, Stolz (2012), por exemplo, aponta que o gênero natural é mantido nos nomes emprestados do espanhol que se referem a seres humanos, assim como em alguns adjetivos que também descrevem seres humanos.

Na mesma perspectiva de Stolz (2012), Hajek e Williams-Van Klinken (2019), investigando especificamente o gênero no TD a partir da noção de contato linguístico, propõem que, pelo menos dentro da variedade acroletal, há indícios de um sistema incipiente de gênero que pode ser visto, por exemplo, em um processo de atribuição de gênero a algumas palavras que não são empréstimos do português. Segundo os autores, devido ao contato prolongado com o português, é possível dizer que o TD começou a desenvolver um gênero “marginal”, em que a marcação de gênero via empréstimo é opcional, mas cada vez mais comum no acrolecto.

Para sustentar empiricamente essa perspectiva, Hajek e Williams-van Klinken, (2019) apontam para diversos dados do TD que refletem a marcação de gênero da gramática do português. Entre eles, estão os termos de parentesco emprestados do português, que apresentam a variação *-u/-a* na sua forma para todos os falantes,

independentemente do socioleto:

(15)

- a. kunhadu/a ('cunhado/a')
- b. padrastu / madrasta ('padrasto / madrasta')
- c. primu/a ('primo / prima')
- d. subrinhu/a ('sobrinho / sobrinha')
- e. tiu/a ('tio / tia')

(Hajek e Williams-van Klinken, 2019, p.69)

Os autores apontam ainda para a existência de alguns termos que, apesar de não indicarem relações de parentesco, ainda assim mantêm a distinção de gênero do português, tal como exemplificado a seguir:

(16)

- a. senhór / senhora ('senhor / senhora')
- b. padre / madre ('padre / freira')
- c. irmaun / irman ('irmão / irmã' [religioso/social])
- d. mestri / mestra ('professor / professora')
- e. profesór / profesora ('professor / professora')
- f. enfermeiru/a ('enfermeiro / enfermeira').

(Hajek e Williams-van Klinken, 2019, p.70)

Além disso, os autores notam que o número de substantivos femininos do português em uso no TD para se referir especificamente a mulheres está aumentando significativamente, impulsionado pelos meios de comunicação que seguem a prática do português e da elite timorense:

Por exemplo, todos os embaixadores, independentemente de serem homens ou mulheres, costumavam ser comumente referidos como embaixadór, como em "Senhora Joana embaixadór Australia" ('Senhora Joana é embaixadora da Austrália.M'). Contudo, de forma crescente, os falantes acroletais e mesoletais estão fazendo a distinção entre embaixadór (masculino) e embaixadora (feminino), preferindo agora expressões como "Senhora Joana embaixadora Australia" ('Senhora Joana é embaixadora da Austrália.F').

(Hajek e Williams-van Klinken, 2019, p.70<sup>12</sup>)

---

<sup>12</sup> So, for instance, all ambassadors, regardless of whether they were male or female, used to be commonly referred to as embaixadór, e.g., Senhora Joana embaixad'or Australia 'Mrs.

Em relação aos adjetivos, geralmente, tais elementos são emprestados na forma masculina, que no TD é utilizada para se referir tanto a homens quanto a mulheres. Os autores citam apenas algumas exceções a esse padrão, como em *bonitu/a* ('bonito/a') e *kuriozu/a* ('curioso/a') que são comumente empregados com a distinção de gênero do português. No entanto, embora seja dependente de um discurso formal, segundo os autores, a forma feminina é possível de ser realizada pelos falantes quando a língua está suficientemente influenciada pelo português:

(17)

- a. *Nia hetan isin-diak maibee sai fraka liu.*  
 3SG get body-good but become weak.F very  
 'She regained her health, but became very weak.'

Outro ponto interessante apontado pelos autores é o emprego da vogal *-a* em adjetivos que são baseados em empréstimos do português, mas que não funcionam como adjetivos femininos na língua de origem, tal como sistematizado a seguir:

- a) Adjetivos que possuem terminação *-a* no TD por serem derivados de substantivos femininos emprestados do português: *forsa* ('forte'), derivado de força ('força').
- b) Adjetivos que apresentam a terminação *-a* no TD por serem comumente empregados com substantivos femininos do português (cf. Greksakova, 2018, p. 215), por exemplo, *morna* ('morno') – geralmente acompanhada do substantivo água - e *klimátika* ('climático') – geralmente acompanhada de *mudansa* ('mudança') or *alterasaun* ('alteração')
- c) Adjetivos que possuem a terminação *-a* no TD porque derivam da forma do presente do singular da 3ª pessoa de verbos portugueses: *aperta* ('justo') do verbo português apertar ('apertar') e *keima* ('queimado') do verbo português queimar ('queimar').

(Hajek e Williams-van Klinken, 2019, p.71-72)

---

Joana is Australian ambassador.M'. Increasingly, however, acrolectal and mesolectal speakers are distinguishing between male embaixad'or and female embaixadora, thus now preferring Senhora Joana embaixadora Australia 'Mrs. Joana is Australian ambassador.F.' (Hajek e Williams-van Klinken, 2019, p.70 – tradução nossa).

Os padrões de gênero do português impactaram o TD não apenas na adoção de pares de empréstimos com marcação de gênero, mas também de maneiras mais sutis. Conforme Hajek e Williams-van Klinken (2019), uma dessas formas é na interpretação de substantivos plurais masculinos, em que há uma distinção clara entre a interpretação de substantivos emprestados e nativos. Mais especificamente, para substantivos nativos do TD que se referem especificamente ao masculino, a forma pluralizada não inclui o feminino em sua interpretação:

- (18)
- |                       |                      |
|-----------------------|----------------------|
| a. <i>maun</i>        | ‘irmão mais velho’   |
| b. <i>maun sira</i>   | ‘irmãos mais velhos’ |
| c. <i>katuas</i>      | ‘homem maduro’       |
| d. <i>katuas-sira</i> | ‘homens maduros’     |

(Adaptado de Hajek e Williams-van Klinken, 2019, p.72)

Assim, a denotação de *maun sira* (‘irmão mais velho’ DEF.PLUR) e *katuas-sira* (homem maduro DEF.PLUR), em (18), não inclui entidades femininas em sua denotação, de forma que o plural é específico para o gênero semântico do substantivo de base. Por outro lado, para substantivos emprestados do português, a forma masculina pode incluir também a interpretação do feminino quando empregada no plural. Por exemplo, *mestri sira* (professor.M DEF.PLUR) é geralmente interpretado como incluindo tanto professores homens quanto mulheres, e *deputadu sira* (deputado.M DEF.PLUR) também pode ser interpretado como um conjunto que inclui referentes masculinos em femininos. Em contraste, a forma feminina emprestada é sempre restrita ao gênero. Por exemplo, *mestra sira* (professora.F DEF.PLUR) inclui apenas professoras.

Tomados em conjunto, esses fenômenos refletem um grau de bilinguismo ou multilinguismo no Timor-Leste, onde os falantes de TD podem alternar entre o uso de formas nativas e emprestadas conforme o contexto discursivo.

## 2.5. CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo, buscamos descrever alguns aspectos relacionados ao gênero no TD, levando em consideração, especialmente, o intenso contexto de contato linguístico em que o Timor-Leste historicamente está inserido.

Em especial, o contato entre o TD e o português tem consequências significativas na gramática da língua, o que é especialmente interessante nas discussões sobre gênero empreendida nesta pesquisa. Essa influência é evidenciada na adoção de pares de empréstimos com marcação de gênero, bem como na introdução de formas inovadoras de marcação de gênero que anteriormente não existiam na língua.

Esse panorama é tomado por Hajek e Williams-van Klinken (2019) como sinal de que, o contato prolongado com o português levou ao desenvolvimento de um gênero “marginal” no TD, em que a expressão de gênero é opcional, mas se torna cada vez mais comum, especialmente no acroleto. Além disso, a presença marcante de empréstimos do português em diferentes variedades do TD fez com que muitos timorenses que não dominam a gramática portuguesa, desenvolvessem alguma noção sobre a marcação de gênero nessa língua.

Em suma, essas mudanças refletem a complexidade e a variabilidade do gênero gramatical no TD como resultado do contato linguístico com o português. Tomando como base o panorama empírico delineado neste capítulo, buscamos, no capítulo 3, ampliá-lo de modo para abarcar os marcadores relacionados a gênero *mane*, *feto*, *aman* e *inan*, que são o foco da nossa pesquisa.

### **CAPÍTULO 3: AS PROPRIEDADES EMPÍRICAS DAS MARCAS RELACIONADAS A GÊNERO EM TÉTUM-DÍLI**

Neste capítulo buscamos descrever as propriedades empíricas das formações construídas com *mane* ('homem'), *feto* ('mulher'), *aman* ('macho/pai') e *inan* (fêmea/mãe) para a marcação sugestiva de gênero em TD. Como não há muitos trabalhos a esse respeito, tomamos como ponto de partida os manuais gramaticais de Hull e Eccles (2005) e Costa (2015), buscando verificar como tais formações são descritas dessa perspectiva. Além disso, de uma perspectiva mais linguística, nos baseamos na descrição comparada de Choupina (2011) e Belo e Oliveira (2021) que descrevem o funcionamento do gênero em TD a partir do contraste com o sistema do português. Além disso, a partir dessa mesma literatura buscamos estabelecer uma das controvérsias mais reconhecidas quando se trata da marcação interpretada como gênero no TD, a saber, o estatuto do processo de formação de palavras responsável por gerar tais formações, se a derivação ou a composição.

Para tanto, esse capítulo está dividido da seguinte maneira: na seção 3.1 trazemos a descrição empírica e a posição analítica delineada nas gramáticas. Enquanto Hull e Eccles (2005), trata tais elementos como derivação, Costa (2015), por sua vez, classifica-os como composição. Já na seção 3.2, procuramos ampliar o escopo da discussão a partir de abordagens mais linguísticas, que tem na base a comparação entre o TD e o português. Mais especificamente, enquanto Choupina (2011) parte de uma ideia de composição, Belo e Oliveira (2021), por sua vez, aponta para o comportamento derivacional dessas formas. Finalmente, na seção 3.3 procuramos agregar novas propriedades empíricas, em especial relacionadas às formas inanimadas e suas possibilidades de combinação com as marcas de gênero, visando a ampliar o panorama empírico estabelecido no capítulo.

#### **3.1. GÊNERO EM TÉTUM-DÍLI: A PERSPECTIVA DAS GRAMÁTICAS**

Na gramática de Hull e Eccles (2005) os marcadores de gênero do TD aparecem mencionados em uma seção que trata de substantivos qualificativos empregados como sufixos. Na visão dos autores, portanto, as formas de gênero

*mane* ('homem'), *feto* ('mulher'), *aman* ('macho/pai') e *inan* ('fêmea/mãe') do TD, apesar de serem considerados substantivos, ou seja, formas livres na língua, devem ser analisados como sufixos quando sua função é a de expressar gênero.

Em termos empíricos, uma das observações mencionadas em Hull e Eccles (2005) é que alguns substantivos sem nenhuma marca de gênero no TD são interpretados como referentes ao masculino. Nesses casos, para ser interpretado como feminino a formação necessariamente necessita da adição de *-feto*.

(19)

a. liurai	'rei'	liurai-feto	'rainha'
b. na'i	'senhor'	na'i-feto	'senhora'
c. soldadu	'soldado'	soldadu-feto	'soldada'

(Hull e Eccles, 2005, p.23)

Por outro lado, os autores apontam que há substantivos que, quando não marcados para gênero, podem se referir semanticamente tanto ao masculino, quanto ao feminino. Nesses casos, para licenciar a interpretação de gênero masculino é preciso usar o marcador *mane*, enquanto para especificar o feminino, emprega-se o marcador *feto*, como ilustrado nos exemplos abaixo:

(20)

a. labarik	'criança'	labarik-mane	'menino'	labarik-feto	'menina'
b. oan	'filho/a'	oan-mane	'filho'	oan-feto	'filha'
c. banin	'sogro/a'	banin-mane	'sogro'	banin-feto	'sogra'

(Hull e Eccles, 2005, p.23)

Os autores descrevem também o funcionamento das formas *aman* e *inan* empregados no TD para a especificação do sexo do referente relacionado a nomes de animais.

(21)

a. fahi	'porco/a'	fahi-aman	'porco'	fahi-inan	'porca'
b. asu	'cachorro/a'	asu-aman	'cachorro'	asu-inan	'cachorra'
c. manu	'galo/inha'	manu-aman	'galo'	manu-inan	'galinha'

(Hull e Eccles, 2005, p.23-24)

Paralelamente ao que é assumido para o funcionamento do gênero em nomes [+humanos], *aman* e *inan* também são entendidos como substantivos que, na marcação de gênero, funcionam como sufixos na língua. Esse estatuto duplo detectado pelos autores é compatível com a nossa hipótese de que tais elementos estão se gramaticalizando na língua.

Já Costa (2015), ao apresentar a estrutura do substantivo no TD, afirma que a língua não possui gênero, mas reconhece sua expressão via composição. Em relação à composição, o autor aponta que:

existem muitos léxicos compostos (sintagmáticos) que se fixaram na língua, passando a funcionar como unidades simples. Neste processo, o significado do sintagma deixa de ser o seu significado composicional e passa a ser um significado simples.

(Costa, 2015, p. 104)

Para ilustrar o processo de composição, Costa (2015) recorre aos exemplos abaixo, dentre os quais podemos notar a inserção de *mane* em (22b), o que deixa transparecer a análise subjacente de que as marcas relacionadas a gênero da língua atuam como compostos do tipo N+N.

(22)

a. N+Adj: matan-dook (matan+dook)	'curandeiro'	(olho+longe)
b. N+N: oan-mane (oan+mane)	'filho'	(filho+homem)
c. N+V: hena-taka (hena+taka)	'lençol'	(pano+cobrir)
d. V+N: foti-oin (foti+oin)	'orgulhoso'	(erguer+rosto)
e. V+V: fó-hatene (fó+hatene)	'informar'	(dar+saber)
f. Adv+Adv: liu-liu (liu+liu)	'especialmente'	(mais+mais)

Em relação à marcação de gênero do TD, Costa (2015) separa as formações em três categorias: pessoas, animais e vegetais, reconhecendo a possibilidade de tais elementos se anexarem a formas não animadas. É interessante apontar, no entanto, que essas categorias são bastante compatíveis com o papel desempenhado por classificadores nominais em uma perspectiva translinguística, o que parece corroborar com a proposta que desenvolvemos nesta dissertação de que tais elementos são classificadores no TD.

Para os nomes que se referem a pessoas, o autor aponta o uso do termo

mane ('homem') e feto ('mulher') pospostos ao substantivo na expressão do masculino e do feminino, respectivamente, ilustrando com os seguintes exemplos:

(23)

- |                 |                   |              |                  |
|-----------------|-------------------|--------------|------------------|
| a. labarik mane | 'rapaz'           | labarik feto | 'rapariga'       |
| b. alin mane    | 'irmão mais novo' | alin feto    | 'irmã mais nova' |

(Costa, 2015, p. 51)

Para nomes de animais, o autor aponta o emprego dos termos *aman* ('macho') e *inan* ('fêmea') também pospostos aos substantivos na expressão do masculino e feminino, tal como nos exemplos a seguir:

(24)

- |              |          |           |        |
|--------------|----------|-----------|--------|
| a. kuda aman | 'cavalo' | kuda inan | 'égua' |
| b. busa aman | 'gato'   | busa inan | 'gata' |

(Costa, 2015, p. 51)

Já para os substantivos que se referem a vegetais, o autor aponta o emprego de *mane* ('homem') e *feto* ('mulher') para distinguir vegetais que dão somente flor daqueles que dão flor e fruto<sup>13</sup>, respectivamente:

(25)

- |                |                  |                   |
|----------------|------------------|-------------------|
| a. aidila mane | 'papaeira macho' | - dá somente flor |
| b. aidila feto | 'papaeira fêmea' | - dá flor e fruto |

(Costa, 2015, p. 51)

Como se pode notar, os manuais de gramática consultados, como é característico do gênero, não chegam a desenvolver argumentos a respeito de suas respectivas classificações para as formas *mane*, *feto*, *aman* e *inan* do TD, seja para

<sup>13</sup> Segundo Costa (2015, p. 51-52), a palavra *amak* ('macho') também pode ser empregada para fazer a distinção em (29), por exemplo: *batamamak* ('milho macho'). Além disso, em nota, o autor detecta ainda o emprego pejorativo de *amak* ('macho') e *inak* ('fêmea'), indicando desprezo, em associação à palavra *buat*, que significa "coisa", por exemplo: *buatamak* ('gajo') e *buatinak* ('gaja'). Exemplos dessa natureza não serão abordados neste trabalho, uma vez que se referem ao tétum terik.

tratá-las como derivação ou como composição. Por isso é possível detectarmos algumas incongruências que surgem da formalização dessas análises. Por exemplo, em Hull e Eccles (2005), que tratam as formas sugestivas de gênero como derivação, *mane*, *feto*, *aman* e *inan* são classificados como substantivos qualificativos, o que, por sua vez, aproxima tais formações da noção de composição. Da mesma forma, em Costa (2015), que trata as formas sugestivas de gênero como composição, há uma sobreposição entre a noção de sintagma e de composição, já que o autor entende que, no composto, o significado composicional do sintagma passa a ser um “significado simples”. Se por outro lado, a ideia de “significado simples” corresponde, em linhas gerais, ao significado de uma palavra, a linha entre composição e derivação é que passa a ser tênue.

Na próxima seção, buscamos completar o panorama empírico das formações de gênero no TD a partir de trabalho com viés mais linguístico, mais especificamente, a partir de Choupina (2011) e Belo e Oliveira (2021).

### 3.2. GÊNERO EM TÉTUM-DÍLI: UMA VISÃO LINGUÍSTICA

O trabalho de Choupina (2011) desenvolve uma perspectiva comparativa das propriedades de marcação de gênero em português europeu (PE) e em TD, fornecendo uma descrição do sistema de gênero nas duas línguas em questão.

Especificamente em relação ao TD, a autora aponta que o contraste lexical entre palavras de diferentes raízes é uma das estratégias empregadas na língua para determinar o gênero, como nos dados abaixo:

(26)

a. aman	‘pai’	inan	‘mãe’
b. mane	‘homem’	feto	‘mulher’
c. klosan	‘o jovem’	fetorán	‘a jovem’

(Choupina, 2011, p. 74)

Por completude da descrição, especificamente em relação ao dado *klosan* em (26c), ressaltamos que tal elemento pode ser empregado como um substantivo subspecificado para gênero, podendo se referir ao masculino ou ao feminino (‘o jovem’/ ‘a jovem’). Dessa forma, também é possível anexar *mane* ou *feto* (‘mulher’)

ao substantivo, como nos dados a seguir:

(27)

- a. klosan ‘o jovem’/ ‘a jovem’
- b. klosan-mane ‘o jovem’
- c. klosan-feto ‘a jovem’

Para além disso, Choupina (2011) aponta que, no TD, as marcas de gênero pode ser dada por adjetivos típicos, seja de masculino ou de feminino. Por exemplo, o adjetivo *katuas* (‘idoso’) é tipicamente masculino, como se pode ver em (28a), enquanto o adjetivo *ferik* (‘idosa’) é feminino, como ilustrado em (28b). Nesse caso, não há necessidade de marcação do substantivo, que acaba tomando o gênero do referente a partir da semântica do próprio adjetivo.

(28)

- a. liurai **katuas** laran luak                      ‘o rei idoso é bondoso’
- b. liurai (feto) **ferik** laran luak                      ‘a rainha idosa é bondosa’

(Choupina, 2011, p. 73)

Ao compararmos o dado (28a) e (28b) podemos notar que a única diferença entre elas é a forma do adjetivo. Esse contraste lexical é suficiente para desencadear a leitura de gênero que permeia a sentença toda. É interessante notar ainda que, em (28b), há a possibilidade de que a forma *feto*, indicadora de feminino, acompanhe o substantivo *liurai* (‘rei’/‘rainha’). Embora não tenha sido apontado pela autora, indicamos também a possibilidade de que o adjetivo tipicamente masculino *katuas* (‘idoso’) seja empregado simultaneamente à forma *mane*, que explicita o masculino no substantivo, ou seja, não há uma incompatibilidade na expressão do gênero do substantivo ainda que o adjetivo seja atrelado a um ou outro gênero.

(29)

- a. liurai-mane *katuas* laran luak  
‘o rei idoso é bondoso’

Considerando que *feto* e *mane* são opcionais em (28b) e (29a), sendo o gênero especificado pelo adjetivo, Choupina (2011) levanta a hipótese de que possa haver no TD algum processo sintático de marcação de gênero, para além do

processo morfológico tradicionalmente assumido pelas gramáticas. Na nossa percepção, no entanto, esse é um indício do funcionamento desses elementos como classificadores.

Na comparação com o PE, Choupina (2011) destaca ainda que, no TD, os adjetivos que não são tipicamente atrelados a gênero, permanecem invariáveis, mesmo quando modificam substantivos humanos ou animados, tal como ilustrado nos seguintes exemplos:

(30)

- |                                      |                              |
|--------------------------------------|------------------------------|
| a. labarik mane <b>aas</b> bá iskola | 'o menino alto vai à escola' |
| b. labarik feto <b>aas</b> bá iskola | 'a menina alto vai à escola' |

(31)

- |  |                            |
|--|----------------------------|
| a. liurai katuas laran <b>luak</b>       | 'o rei idoso é bondoso'    |
| b. liurai (feto) ferik laran <b>luak</b> | 'a rainha idosa é bondosa' |

(Choupina, 2011, p. 73)

Nos exemplos acima, os adjetivos em negrito, *aas* ('alto') e *luak* ('bondoso'), são invariáveis, mantendo, portanto, uma mesma forma, independentemente de se referirem ao masculino ou ao feminino. Em (30a-b), o gênero do referente é indicado pelas formas *mane* e *feto*, respectivamente, enquanto em (31a-b) o gênero é especificado pelo emprego de adjetivos típicos, *katuas* ('idoso') e *ferik* ('idosa'). Ainda assim, os adjetivos *aas* ('alto') e *luak* ('bondoso') permanecem com realizações morfofonológicas idênticas nas duas interpretações.

Em relação aos empréstimos do PE, a autora aponta que algumas palavras emprestadas entram no TD já com um contraste de gênero herdado da língua de origem, como nos exemplos abaixo:

(32)

- |            |         |
|------------|---------|
| a. kantór  | kantora |
| b. tiu     | tia     |
| c. sobriñu | sobriña |

(Choupina, 2011, p. 74)

É interessante apontar ainda que, em alguns empréstimos, o contraste de

gênero realizado permanece ativo também em alguns adjetivos emprestados do português (33a), o que se mantém mesmo quando tais adjetivos são colocados como modificadores de substantivos inanimados (33b-h).

(33)

a. advogadu amerikanu ida	‘um advogado americano’
b. fábrica amerikana ida	‘uma fábrica americana’
c. mapa polítiku Ázia nian	‘o mapa político da Ásia’
d. nia karreira política	‘a sua carreira política’
e. ezersísiu proveitozu ida	‘um exercício proveitoso’
f. atividade proveitoza ida	‘uma atividade proveitosa’
g. konsuladu portugés	‘o consulado português’
h. embaixada portugeza	‘a embaixada portuguesa’

(Hull e Eccles, 2005, p. 156, adaptado por Choupina, 2011, p. 76)

Nos dados acima, com exceção de (33a) em que o substantivo *advogadu* indica um referente com o traço [+humano], o contraste de gênero pode ser visto nos adjetivos oriundos de empréstimo, ainda que os substantivos modificados sejam inanimados.

Embora a abordagem de Choupina (2011) seja mais descritiva do que analítica, a autora afirma que, sendo uma língua sem flexão nominal nem verbal, as categorias como número e gênero, nos nomes ou tempo, aspeto, pessoa e número, nos verbos, são majoritariamente por composição no TD, alinhando-se à abordagem de Costa (2015).

Também a partir de uma abordagem comparada com os dados do português, o trabalho de Belo e Oliveira (2021) traz o foco mais especificamente para diferentes tipos de composição no TD, tratando entre tais dados as formas com *mane*, *feto*, *aman* e *inan*. Na análise dos autores, no entanto, as formas de gênero do TD são mais bem analisadas como derivação do que como composição, alinhando-se, portanto, à Hull e Eccles (2005). O centro da argumentação dos autores se baseia na definição de composição fornecida em Silva e Rio-Torto (2013, p. 466):

Os compostos são unidades impermeáveis a qualquer alteração/inserção no seu interior, sendo marcados pela

impossibilidade de alterar a ordem dos elementos compositivos ou de os substituir por outros.

A questão trazida em Belo e Oliveira (2021) é que os dados com os marcadores de gênero no TD não podem ser tratados como compostos, uma vez que eles violam a condição de impermeabilidade que está no centro da definição de Silva e Rio-Torto (2013). Embora os autores não tragam dados que ilustrem tal propriedade, é possível pensarmos em dados como (34) no qual o marcador anafórico *ne'e*, que pode funcionar tanto como demonstrativo ('esse/essa; este/esta; aquele/aquela'), quanto como artigo definido ('o/a,') é colocado entre o substantivo e a marca de gênero.

(34)

- |                             |                        |
|-----------------------------|------------------------|
| a. labarik <i>ne'e</i> mane | (‘a criança é menino’) |
| b. labarik <i>ne'e</i> feto | (‘a criança é menina’) |

O que acreditamos estar acontecendo nesse tipo de dado que contam com elementos intervenientes é uma estrutura de sintagma. Isso é possível porque os marcadores em questão também podem atuar são formas livres na língua.

Uma evidência nessa direção são dados em que o conteúdo conceitual desses elementos pode ser recuperado, quando ele são empregados isoladamente.

(35)

- |                                 |  |
|---------------------------------|--|
| a. asu <i>ne'e</i> aman         | (‘o cachorro é macho’/ ‘o cachorro é pai)  |
| b. asu <i>ne'e</i> inan         | (‘o cachorro é fêmea’/ ‘o cachorro é mãe’) |
| c. liurai <i>ne'e</i> feto      | (‘o rei é mulher’)                         |
| d. liurai <i>ne'e</i> mane      | (‘o rei é homem’)                          |
| e. liurai-feto <i>ne'e</i> mane | (‘a rainha é homem’)                       |

Em (35a-b) os substantivos *aman* e *inan* podem ser interpretados como ‘pai’ e ‘mãe’, respectivamente. Essa leitura não pode ser veiculada dentro da estrutura de combinação mais local, o que parece apontar para a ideia de que estamos diante de um sintagma. O mesmo raciocínio pode ser estendido para os dados em (35c-e), em que as leituras de *feto* e *mane* recuperam o conteúdo conceitual de ‘mulher’ e ‘homem’, respectivamente. Em (35e) vemos que é possível inclusive que as formas *feto* e *mane* coexistam na mesma formação, o que evidencia que se trata de

elementos com estatuto distinto.

A ideia de que um mesmo elemento na língua possa ora se comportar como elemento lexical, ora como elemento funcional corrobora a ideia da MD de que estruturas complexas, seja no nível da palavra ou do sintagma, são formadas no mesmo componente da gramática, a sintaxe. Essa é a perspectiva que será explorada no desenvolvimento da nossa análise no capítulo 5.

### 3.3. GÊNERO EM TÉTUM-DÍLI: OUTRAS PROPRIEDADES EMPÍRICAS DETECTADAS NA PESQUISA

Nesta seção, buscamos contribuir para ampliar a descrição do funcionamento das marcas relacionadas a gênero no TD, a partir de propriedades detectadas no curso da própria pesquisa e que ainda não foram abordadas na literatura. Tais propriedades, somadas àquelas já descritas na literatura, deverão ser capturadas pela análise formal desenvolvida no capítulo 5.

Como vimos anteriormente, nos substantivos não animados, majoritariamente, não há marcas de gênero no TD, uma vez que o gênero na língua é, em geral, semântico. Em Costa (2015), no entanto, vimos a possibilidade de que os elementos *mane* e *feto* acompanhem nomes de vegetais, distinguindo aqueles que dão somente flor daqueles que dão flor e fruto, como em (25). Além disso, é importante apontarmos que a anexação desses elementos não está restrita a nomes não animados de vegetais, como ilustrado em (36a-d) a seguir:

(36)

- |              |                                    |
|--------------|------------------------------------|
| a. tais-mane | ‘vestimenta tradicional masculina’ |
| b. tais-feto | ‘vestimenta tradicional feminina’  |
| c. tasi-mane | ‘maremoto (mar revolto)’           |
| d. tasi-feto | ‘marola (mar calmo)’               |

Nos exemplos acima, os substantivos inanimados *táis*, que significa ‘roupa’, e *tasi*, que significa ‘mar’, estão acompanhados das marcas gênero da língua, indicando no primeiro caso um traje tradicionalmente masculino ou feminino, a depender do elemento que é anexado ao substantivo de base e, no segundo caso, um mar revolto ou calmo também a depender das marcas de gênero.

É interessante ressaltar ainda que, a depender do contexto em que são

empregados, tais substantivos inanimados associados a marcas de gênero podem ainda veicular outros significados, como a semântica de ‘sul’ e ‘nordeste’, trazidas nas sentenças abaixo:

(37)

- a. *nia mai hosi tasi-mane* Timor ‘ele/ela é do sul do Timor’
- b. *nia mai hosi tasi-feto* Timor ‘ele/ela é do nordeste do Timor’

Ainda sobre a associação entre as marcas de gênero e substantivos inanimados, o TD apresenta algumas formas em que tais substantivos são marcados apenas para gênero feminino ou para gênero masculino, sem a possibilidade de pareamento dessas duas formas.

(38)

- a. *ai-inan* ‘lenha com tronco grande’  
árvore-mãe
- b. *lia-inan* ‘língua materna’ ou ‘lei’  
língua-mãe
- c. *uma-mane/umane* ‘tios do lado paterno/materno’  
casa-homem

Os substantivos em (38a-c) significam isoladamente ‘árvore’, ‘língua’ e ‘casa’, respectivamente. Quando associados a formas de gênero, tais substantivos licenciam outras leituras: *ai-inan* em (38a) possui um significado idiomático de ‘lenha com tronco grande’, *lia-inan* em (38b) pode ter duas leituras, uma delas é a interpretação de ‘língua mãe/materna’ e a outra é a interpretação de ‘lei’. Finalmente, *uma-mane* em (38c) é ambíguo entre a interpretação de ‘tios do lado paterno ou materno’, com a semântica definida pelo contexto.

Além disso, é interessante apontar que os substantivos inanimados podem ser transformados em substantivos animados, se tais elementos forem atrelados às marcas relacionadas a gênero em um mundo criado em que o contexto seja apropriado para essa interpretação, como ilustrado a seguir:

- (39) a. **Odamatan-mane** no **odamatan-feto** laran-monu ba malu  
porta<sub>masc</sub> e porta<sub>fem</sub> se apaixonam

b. **Odamatan-aman** no **odamatan-inan** laran-monu ba malu  
 porta<sub>masc</sub> e porta<sub>fem</sub> se apaixonam

Em (39), o substantivo inanimado *odamatan* ('porta') foi associado às marcas *mane* e *feto* em (39a) e *aman* e *inan* em (39b) forçando uma leitura animada que é corroborada pelo verbo ('apaixonar'), que também atua nessa coerção de animacidade.

Outro ponto relevante a ser ressaltado é que, em TD, o pronome pessoal de terceira pessoa, tanto no singular, como no plural, também apresenta marcação de gênero. Na tabela abaixo, sistematizamos os pronomes pessoais do TD, e exploramos apenas os pontos que são relevantes ao fenômeno em questão.

Tabela 2 – Pronomes pessoais do TD

	<b>Tétum</b>	<b>Português</b>
<b>Singular</b>	<i>ha'u</i>	eu
	Ó	tu
	<i>ita-boot</i>	você
	<i>nia- mane nia- feto</i>	ele ela
<b>Plural</b>	<i>Ita</i>	nós (inclusivo)
	<i>Ami</i>	nós (exclusivo)
	<i>Imi</i>	vocês
	<i>sira- mane sira-feto</i>	eles elas

Fonte: elaborada pela autora.

Na tabela acima podemos ver que os pronomes pessoais de terceira pessoa podem se associar às marcas de gênero para especificar o seu referente. É interessante ressaltar que as formas *nia* e *sira*, respectivamente, de terceira pessoa

do singular e do plural, são gramaticais mesmo se empregadas sem as marcas de gênero. Nesse caso, o gênero do referente não é especificado, como em (40c) e (40f) abaixo:

(40)

- |                   |                        |
|-------------------|------------------------|
| a. nia-mane mate  | (‘ele morreu’)         |
| b. nia-feto mate  | (‘ela morreu’)         |
| c. nia mate       | (‘ele/ela morreu’)     |
| d. sira-mane mate | (‘eles morreram’)      |
| e. sira-feto mate | (‘elas morreram’)      |
| f. sira mate      | (‘eles/elas morreram’) |

Ainda em relação ao sistema pronominal, é interessante ressaltar que os elementos *aman* e *inan* em sua interpretação conceitual de ‘pai’ e ‘mãe’, respectivamente, podem se associar aos pronomes pessoais do TD, desencadeando a interpretação de possessivo em todas as pessoas do sistema pronominal, tanto no plural, como no singular:

(41)

- |   |  |
|---|--|
| a. ha’u-aman<br>ha’u-inan                       | meu pai<br>minha mãe                         |
| b. ó-aman<br>ó-inan                             | teu pai<br>tua mãe                           |
| c. ita-boot-aman<br>ita-boot-inan               | seu pai<br>sua mãe                           |
| d. nia-aman (mane/feto)<br>nia-inan (mane/feto) | seu pai (pai dele/a)<br>sua mãe (mãe dele/a) |
| e. ita-aman<br>ita-inan                         | nosso pai<br>nossa mãe                       |
| f. ami-aman<br>ami-inan                         | nosso pai<br>nossa mãe                       |

- |                          |                        |
|--------------------------|------------------------|
| g. imi-aman              | vosso pai              |
| imi-inan                 | vossa mãe              |
| h. sira-aman (mane/feto) | seu pai (pai deles/as) |
| sira-inan (mane/feto)    | sua mãe (mãe deles/as) |

Para completar o panorama empírico explorado até aqui, trazemos também alguns pontos relevantes sobre o funcionamento da marcação de gênero no TD a partir de elementos emprestados do português.

As formas emprestadas que não apresentam uma marca explícita de gênero no português, como aquelas em *-ista*, são incorporadas ao sistema de gênero do TD com a explicitação dessa informação acontecendo via anexação das formas *mane* e *feto*:

(42)

- |                     |                               |
|---------------------|-------------------------------|
| a. jornalista       | 'o jornalista'/ a jornalista' |
| b. jornalista- mane | 'o jornalista'                |
| c. jornalista-feto  | 'a jornalista'                |

Como a forma morfofonológica do substantivo *jornalista* é idêntica tanto no feminino, como no masculino em português, no TD há a necessidade do emprego das marcas *mane* e *feto* para especificação do gênero do referente.

Alguns empréstimos, no entanto, entram no TD já com a marcação de gênero da língua de origem, sendo o contraste de morfofonológico realizado no português suficiente para marcar o gênero no próprio TD. Em alguns desses casos, como ilustrado a seguir, as formas emprestadas não aceitam *mane* e *feto*, mantendo obrigatoriamente sua marcação de gênero da língua origem:

(43)

- |                 |          |
|-----------------|----------|
| a. meninu       | 'menino' |
| b. *meninu-mane |          |
| c. menina       | 'menina' |
| d. *menina-feto |          |

Em outros casos de empréstimos, por sua vez, apesar de haver formas que

entram na língua com as marcas de gênero do português, ainda assim é possível especificar o gênero da formação através das formas nativas *mane* e *feto* a partir da realização morfofonológica da base masculina, como acontece nas formas em *-dor*:

(44)

a. treballador	‘trabalhador’
b. treballadora	‘trabalhadora’
c. treballador-mane	‘trabalhador’
d. treballador-feto	‘trabalhadora’

Por outro lado, verificamos ainda a existência de empréstimos que, embora apresentem contraste de gênero na forma de origem do português, esse contraste não é absorvido no sistema da língua:

(45)

a. avó	‘avô’/ ‘avó’
b. avó-mane	‘avô’
c. avó-feto	‘avó’

Nos exemplos em (45), apesar de haver um contraste ortográfico e morfofonológico nas formas “avô” e “avó” do português, no TD a grafia passa a conservar somente o acento agudo. Essa alteração de grafia é um padrão no sistema da língua: as palavras emprestadas que têm acento circunflexo, passam a ter acento agudo. Dada a similaridade tanto da forma gráfica, como da realização fonológica da forma no TD, para que haja explicitação do gênero é necessário juntar as marcas *-mane* (‘homem’) e *-feto* (‘mulher’) ao substantivo.

Tomando como base o panorama empírico delineado neste capítulo e buscando ampliá-lo de modo a abarcar as controvérsias de análise que permeiam as formações relacionadas à gênero no TD, no próximo capítulo, revisitamos as fronteiras entre composição e derivação.

### 3.4. CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo procuramos estabelecer as propriedades empíricas das formações de gênero no TD, com vistas a recortar as características que

necessariamente precisam ser capturadas pela análise a ser desenvolvida no capítulo 5 deste trabalho.

A partir da literatura selecionada para tanto, foi possível perceber que há um debate a respeito do estatuto formal desses elementos, se eles devem ser tratados como morfema derivacional (Hull e Eccles, 2005; Belo e Oliveira, 2021) ou como um substantivo independente que é parte de um processo de composição quando expressa gênero na língua (Costa, 2015; Choupina, 2011).

Para nos inserir de maneira mais apropriada nesse debate, no próximo capítulo, procuramos revisitar as fronteiras entre derivação e composição para, então, situar de maneira mais apropriada os dados do TD.

## **CAPÍTULO 4: GÊNERO NO TÉTUM-DÍLI E A FRONTEIRA ENTRE DERIVAÇÃO E COMPOSIÇÃO**

Neste capítulo nos aprofundamos nas discussões empíricas sobre as propriedades que caracterizam os processos de derivação e composição nas línguas naturais. Nosso objetivo a partir daí é situar os formadores de gênero do TD nesse panorama com vistas a nos colocar de maneira mais ampla no debate que cerca tais elementos.

De um modo geral, dois aspectos aparecem comumente na literatura como definidores da composição e da derivação: o estatuto dos elementos que se combinam nessas formações e a posição desses elementos. Dessa forma, a composição é definida como um processo que combina palavras ou radicais, enquanto a derivação requer a presença de apenas uma palavra ou radical acrescido de um afixo. No mesmo sentido, os afixos são formas presas com posições pré-definidas, como os prefixos e sufixos, por exemplo, enquanto, nos compostos, se os elementos envolvidos forem livres, em princípios eles poderiam se concatenar em qualquer posição.

No entanto, a distinção entre esses dois processos nem sempre é clara, sendo que as análises disponíveis na literatura podem variar conforme a perspectiva teórica adotada. Como aponta Gonçalves (2011), as visões a respeito das fronteiras e similaridades entre composição e derivação se dividem nas seguintes linhas:

- Compostos e formas derivadas são formados em diferentes componentes da gramática;
- Compostos e formas derivadas não são distinções categóricas, mas parte de um continuum;
- Compostos e formas derivadas são formados no mesmo componente da gramática;

Dentro da primeira perspectiva, há propostas como a de Anderson (1992) em que as formas derivadas são geradas no léxico, enquanto os compostos, por sua vez, são objeto de estudo da sintaxe, aproximando-os dos sintagmas. Por outro lado,

dentro da perspectiva da fonologia lexical, por sua vez, alguns autores (KIPARSKY, 1982; BOOIJ & RUBACH, 1984) propuseram que tanto as formas derivadas, como os compostos seriam ambos formados no mesmo componente, mais especificamente, dentro léxico, sem a necessidade de que tais formas sejam acessadas pela sintaxe.

Relativizando as distinções entre a derivação e a composição por conta de seus limites serem difusos, algumas propostas na linha da morfologia construcional (Kastovsky, 2009; Bauer, 2005; Gonçalves, 2011), por sua vez, assumem que tais noções não são categóricas, mas que haveria nas línguas uma escala de prototipicidade, com a derivação e a composição sendo polos

Por sua vez, em uma abordagem sintática de formação de palavras, como a MD (Halle; Marantz, 1993; Marantz, 1997), os objetos linguísticos complexos são entendidos como elementos construídos pelo único componente gerativo da gramática, a sintaxe. Assim, as diferenças empíricas que se revelam nos diferentes processos de formação de palavras, como a derivação e a composição, deixam de ser entendidas como fruto de pressupostos, para ser analisada como resultados das relações estabelecidas pelos morfemas na estrutura sintática.

Buscando, de um ponto de vista empírico, nos colocar nesse debate, recorreremos ao panorama mais geral das características associadas aos processos derivacionais e de composição com base na sistematização fornecida em Gonçalves e Andrade (2016), que comparam tais processos a partir de suas propriedades mais recorrentes da literatura. Com base nessas características, promovemos uma reflexão sobre o funcionamento das marcas interpretadas como gênero no TD. Essa reflexão parece evidenciar que as formas relacionadas à gênero da língua apresentam um comportamento não estável, se comportando ora como derivação, ora como composição a depender da propriedade empírica que está sendo considerada. Isso explica o debate existente na literatura entre a classificação dos marcadores relacionados a gênero no TD como representante do processo de derivação (Hull & Eccles, 2005; Belo & Oliveira, 2021) ou composição (Costa 2015; Choupina, 2011). Isso também parece ser compatível com a ideia de que *mane*, *feto*, *aman* e *inan* estão em processo de gramaticalização na língua, como propomos no próximo capítulo.

Para desenvolver as discussões deste capítulo, ele está organizado da

seguinte maneira: na seção 4.1, fazemos uma revisão das fronteiras entre composição e derivação com base nos critérios sistematizados por Gonçalves e Andrade (2016) que trazem um conjunto de propriedades comumente associadas a cada um desses processos na literatura; já na seção 4.2, apresentamos uma descrição da marcação sugestiva de gênero no TD a partir das formas *-mane*, *-feto*, *-aman* e *-inan*, tomando como base os critérios elencados por Gonçalves e Andrade (2016). A nossa proposta é que, empiricamente, a marcação relacionada à gênero no TD apresenta propriedades difusas, se aproximando ora do processo de composição, ora do processo de derivação. Por fim, a seção 4.3, encerra o capítulo com as considerações finais.

#### 4.1. COMPOSIÇÃO E DERIVAÇÃO: PROPRIEDADES EMPÍRICAS

Para nos aprofundarmos no debate a respeito da distinção entre derivação e composição de um ponto de vista empírico, nos baseamos, nesta seção, nos critérios sistematizados por Gonçalves e Andrade (2016), que trazem um conjunto de propriedades comumente associadas a cada um desses processos na literatura. Tais propriedades serão, então, colocadas em discussão a partir do funcionamento das formas *mane*, *feto*, *aman* e *inan* no TD para que possamos compreender de maneira mais clara suas propriedades. Essa descrição empírica é relevante para o desenvolvimento da pesquisa na medida em que a proposta de uma análise formal precisa naturalmente capturar o comportamento dessas formas.

Na tabela a seguir, estão representados os critérios, tal como organizados por Gonçalves e Andrade (2016), com as principais diferenças entre os processos de composição e derivação.

Tabela 3 – Principais diferenças entre composição e derivação

		<b>Composição</b>	<b>Derivação</b>
Unidades	A	Palavras ou radicais	Afixos
	B	Formas livres ou presas que correspondem a palavras	Formas presas que não correspondem a palavras de conteúdo

	C	Unidades com posição não necessariamente fixa na estrutura da palavra	Unidades definidas por uma posição pré-determinada numa palavra complexa (à esquerda ou à direita)
	D	As unidades combinam com uma grande variedade de tipos morfológicos	Sufixos combinam predominantemente com radicais; prefixos combinam exclusivamente com palavras
Propriedades estruturais	E	A cabeça lexical fica à esquerda, predominantemente	Cabeça lexical fica à direita
	F	Possibilidade de coordenação entre os constituintes	Não há possibilidade de coordenação
	G	Por expressar ideias mais específicas, há um grande contingente de unidades linguísticas	Por expressar ideias mais gerais, há um número relativamente pequeno de unidades linguísticas
	H	Caracterizam um inventário aberto	Caracterizam um inventário fechado
	I	Possibilidade de flexão entre os constituintes	A flexão é sempre periférica
Propriedades fonológicas <sup>14</sup>	J	Unidades com acento próprio	Partículas que recebem acento apenas na combinação com a base

<sup>14</sup> No original, as propriedades fonológicas começam apenas em L. Em função de J já abordar uma propriedade dessa natureza, fizemos a alteração para fins de discussão neste capítulo.

	L	Ausência de isomorfismo entre palavra morfológica e palavra fonológica	Isomorfismo entre palavra morfológica e palavra fonológica
	M	Manutenção de propriedades segmentais e prosódicas das bases	Mudança na base pela aplicação de regras fonológicas cujo domínio é a palavra fonológica
Propriedades semânticas <sup>15</sup>	N	As unidades expressam um significado lexical	As unidades atualizam conteúdos semânticos mais gerais, capazes de combinação com um número maior de formas linguísticas
	O	Interpretação frequentemente holística	Interpretação quase sempre composicional
	P	Pode ser endocêntrica ou exocêntrica	Massivamente endocêntrica
	Q	Menos estável porque o significado dos elementos geralmente muda por extensões metafóricas ou metonímicas	Mais estável, apresentando funções sintáticas e semânticas predeterminadas, definindo os possíveis usos e significados das palavras derivadas
Produtividade e produção	R	Constrói conjuntos mais fechados de palavras (ad hoc)	Constrói conjuntos mais completos de palavras (mais regular)
	S	Apresenta muitas formas manufaturadas	Cria séries de palavras mais naturalmente

Fonte: Gonçalves e Andrade (2016, p. 265).

<sup>15</sup> No original, as propriedades semânticas começam apenas em O. Em função de N já abordar uma propriedade dessa natureza, fizemos a alteração para fins de discussão neste capítulo.

As propriedades elencadas em Gonçalves e Andrade (2016) estão divididas a partir de critérios de natureza diversa:

- (i) o estatuto das unidades envolvidas no processo de formação de palavras;
- (ii) as propriedades estruturais das formações analisadas;
- (iii) as propriedades fonológicas dos elementos linguísticos gerados;
- (iv) as propriedades semânticas dessas formações;
- (v) as questões relativas à produtividade e à produção das formas analisadas.

Em relação ao estatuto dos elementos envolvidos, que englobam os critérios de A a D, podemos ver que a composição é caracterizada pela combinação de palavras (formas livres) ou radicais (formam presas) que não apresentam necessariamente uma posição fixa. Em contraste, as formações derivadas são caracterizadas por conterem afixos, entendidos como formas presas com posição definida na estrutura da palavra. Em relação aos tipos morfológicos, há a generalização de que os sufixos geralmente se combinam com radicais, enquanto os prefixos se combinam exclusivamente com palavras.

Já o segundo conjunto de propriedades, diz respeito aos aspectos de natureza estrutural e agrupam os critérios de E a I. Mais especificamente, a generalização proposta pelos autores é a de que os compostos são caracterizados predominantemente pelo núcleo à esquerda, enquanto a derivação apresentaria esse núcleo à direita, ou seja, nos próprios sufixos. Além disso, nos compostos há a possibilidade de coordenação entre os elementos formadores desse tipo de estrutura, enquanto isso não acontece na derivação. Da mesma forma, a presença de marcas morfológicas flexionais entre os membros da formação seria uma característica do processo de composição, mas não de derivação. Finalmente, em termos de seu inventário de formadores, a composição seria caracterizada por unidades linguísticas de classe aberta, enquanto os afixos se constituem como um conjunto mais fechado na língua.

Por sua vez, o terceiro conjunto de propriedades, que agrupa as linhas de J a M, engloba, as questões de natureza fonológica. Assim, as unidades do composto, são caracterizadas por manterem seus respectivos acentos no interior da formação.

Assim os compostos acabam por conter duas palavras fonológicas distintas, sendo, portanto, um domínio que não está sujeito à aplicação de regras fonológicas restritas à fronteira de palavra fonológica. Com isso, os compostos tendem a manter as propriedades segmentais das bases. Já a derivação, seria caracterizada por um único domínio de palavra fonológica, contendo apenas um acento primário e estando mais suscetível a mudanças fonológicas na base a depender da natureza do afixo que participa da formação.

As propriedades relativas às características semânticas desses processos de formação estão abarcadas de N a Q na tabela. Semanticamente, então, a generalização é que os compostos veiculam significado de natureza lexical, enquanto a derivação, por sua vez, pode manifestar também um significado mais gramatical ou funcional. Da mesma forma, os autores apontam que, enquanto os compostos podem ser semanticamente endo ou exocêntricos, as formas derivadas por sua vez seriam majoritariamente endocêntricas. Ainda, ao envolver unidades lexicais de classe aberta, a semântica dos compostos tende a ser menos estável do que aquela veiculada pelas formas derivacionais que constituem uma classe mais fechada de elementos.

Finalmente, no que diz respeito à produtividade e produção desses dois tipos de processos, com as propriedades dispostas em R e S na tabela, Gonçalves e Andrade (2016) apontam que a produtividade de compostos é mais *ad hoc*, estando geralmente ligada a uma motivação expressiva e, portanto, apresentando menos regularidade. Por outro lado, a derivação apresentaria maior regularidade em termos de formação, tendo menos lacunas quando comparada à composição.

Como apontado pelos próprios autores, é interessante ressaltar que a distinção entre composição e derivação não é absoluta. Isso significa que os critérios sistematizados na tabela podem ser questionados ou mesmo levar a classificações contraditórias para uma mesma formação, a depender da propriedade analisada.

Sem promover, no entanto, uma discussão aprofundada de cada um desses critérios, podemos problematizar, à título de exemplificação, pelo menos um critério em cada uma das categorias trazidas na tabela. Assim, dentre as propriedades que englobam o estatuto das unidades combinadas, a característica em D aponta que sufixos se combinam com radicais e não com palavras. No entanto, alguns sufixos do português, como o formador de advérbio *-mente* e o diminutivo *-zinho*, se combinam com palavras. Isso dialoga também diretamente com a propriedade de

natureza estrutural em I, que aponta que flexão é sempre periférica nas formas derivadas. Nas formações com *-mente* e *-zinho* do português, no entanto, podemos notar uma flexão de gênero antecedendo o afixo:

(46)

- a. lindazinha
- b. lindamente

Ainda em termos de propriedades estruturais, podemos também problematizar a propriedade F, que aponta não haver a possibilidade de coordenação na derivação. No português, no entanto, alguns afixos, como *pré-*, *bi-*, *tri-* e *-mente*, licenciam a coordenação, como exemplificado em (47):

(47)

- a. pré e pós-lexical
- b. bi e tri-campeão
- c. calma e tranquilamente

Dentre as propriedades fonológicas, por sua vez, Gonçalves e Andrade (2016) apontam que os afixos acentuados, ou seja, que apresentam tonicidade, terão, naturalmente, um comportamento fonológico semelhante ao previsto para a composição, como mapeado nas propriedades J e L:

(48)

- a. {[vice Af] PrWd [prefeit Rad o] PrWd} MWd
- b. {[pré Af] PrWd [test Rad e] PrWd} MWd

(GONÇALVES e ANDRADE, 2016, p. 266)

Na presença de duas palavras fonológicas, também o critério M apresenta questionamentos interessantes na medida em que as formações com esses afixos acentuados não estão sujeitas à processos licenciados no domínio da palavra fonológica, como, por exemplo, a neutralização das pré-tônicas:

(49)

- a. {[bɛl Rad a] PrWd [mente Af] PrWd} MWd      (b[e]lamente)
- b. {[nɔv Rad a] PrWd [mente Af] PrWd} MWd      (\*n[o]vamente)

- c. {[cɛu Rad] PrWd [zinho Af] PrWd} MWd (\*c[e]uzinho)  
 d. {[dɔlar Rad] PrWd [zinho Af] PrWd} MWd (\*d[o]larzinho)

(GONÇALVES e ANDRADE, 2016, p. 267)

Já em relação às propriedades semânticas é possível problematizar o critério em O, que aponta haver nas formações derivadas uma interpretação quase sempre composicional. A questão relevante aqui é que não são raros os casos nos quais formações derivadas podem licenciar leitura não composicional:

(50)

- |               |   |
|---------------|---|
| a. beijinho   | ‘beijo pequeno’ ou ‘tipo de doce’                           |
| b. açougueiro | ‘aquele que trabalha no açougue’ ou ‘médico cirurgião ruim’ |
| b. lavagem    | ‘processo de lavar’ ou ‘comida de porco’                    |

Em (50), encontramos diferentes sufixos do português (*-inho*, *-eiro* e *-agem*) desencadeando a possibilidade de uma leitura não composicional das formações.

Finalmente, em relação aos critérios de produtividade e produção, a derivação é associada à construção de conjuntos mais completos de palavras, sendo, portanto, mais regular, como sistematizado em R. No entanto, também é bastante comum encontrarmos afixos derivacionais com uma produtividade muito restrita, como acontece com *-ebre*, que parece ser licenciado somente em *casebre*, por exemplo, sem potencial para outras formações. Além disso, mesmo em afixos considerados produtivos como nos exemplos abaixo, não é difícil encontrarmos lacunas de formação:

(51)

- |            |            |
|------------|------------|
| a. laranja | laranjada  |
| b. limão   | limonada   |
| b. morango | ?morangada |
| c. uva     | ?uvada     |

Nos dados em (51), na interpretação de ‘suco de fruta X’, algumas bases, como “laranja” e “limão” são licenciadas, enquanto outras, como “morango” e “uva”, não são gerando lacunas nessa regra derivacional.

Apesar desses importantes questionamentos, na próxima seção, procuramos

explorar as propriedades empíricas das formas interpretadas como gênero no TD para verificar se tais formações, empiricamente se alinham mais propriamente às características da composição, como previsto por Costa (2015) e Choupina (2011) ou da derivação, conforme proposto em Hull e Eccles (2005) e Belo e Oliveira (2021).

#### 4.2. FORMAÇÕES DE GÊNERO NO TD: COMPOSIÇÃO OU DERIVAÇÃO?

Nesta seção, apresentamos uma descrição da marcação de gênero no TD a partir das formas *mane*, *feto*, *aman* e *inan*, tomando como base os critérios elencados por Gonçalves e Andrade (2016). Para tanto, atribuímos uma classificação – como composição ou derivação – para cada uma das propriedades sistematizadas pelos autores, com as suas respectivas ilustrações exemplificadas com dados do TD na sequência da tabela.

A nossa visão a partir dessas propriedades é que, empiricamente, as formações interpretadas como gênero no TD apresentam algumas características compatíveis com a composição e outras mais próximas do processo de derivação. Essa ausência de uma sistematicidade clara é compatível com o debate frequentemente encontrado na literatura que se debruçou sobre o tema.

Tabela 4 – Comportamento das marcas sugestivas de gênero do TD

		<b>Tétum-Díli</b>	<b>Justificativa</b>
Unidades	A	Composição	os marcadores de gênero são palavras independentes
	B	Composição	os marcadores de gênero podem ser usados isoladamente
	C	Composição	os marcadores de gênero podem aparecer antes ou depois da base
	D	Derivação	os marcadores de gênero se combinam apenas a substantivos e adjetivos

Propriedades estruturais	E	Composição	o núcleo dessas formações é a própria base e não o marcador de gênero
	F	Composição	a marca de gênero tem escopo sobre os elementos da coordenação
	G	Derivação	número relativo pequeno de itens que atuam como marcador de gênero na língua
	H	Derivação	inventário fechado de marcadores
	I	----	não se aplica – ausência de marcas flexionais na língua
Propriedade fonológicas	J	Composição	unidades com acento próprio
	L	Composição	constituem-se como duas palavras fonológicas
	M	----	Não se aplica – necessidade de estudos fonológicos no sistema do TD
Propriedades semânticas	N	Derivação	em combinação com as bases expressam significado geral
	O	Derivação	interpretação semântica composicional estável
	P	Derivação	massivamente endocêntrica
	Q	Derivação	função semântica relativamente estável na marcação de gênero
Produtividade e produção	R	Derivação	paradigmas mais regulares de palavras
	S	Derivação	formas produtivas na marcação de gênero

Fonte: Elaborada pela autora (2024) a partir de Gonçalves e Andrade (2016, p. 265).

Para compreender melhor a classificação proposta na tabela 4, vamos explorar com mais detalhes os critérios nela elencado, fornecendo exemplos do TD que empiricamente ilustram cada um deles.

Em referência às unidades envolvidas na formação, especialmente englobando os critérios A e B, é importante apontar que as formas *mane*, *feto*, *aman* e *inan* são formas livres, que podem ser empregadas de modo isolado na língua, sem, portanto, a necessidade de se anexar a uma base. Esse comportamento, como vimos em Gonçalves e Andrade (2016), é característico de elementos que participam do processo de composição, já que afixos são formas presas:

(52)

- a. **Mane** ne'e servisu-na'in  
homem ANF trabalhador  
'Esse homem é trabalhador'
- b. **Feto** ne'e oin-na'in  
mulher ANF simpática  
'Essa mulher é simpática'
- c. Ha'u-nia **inan** laran-  
di'ak minha mãe  
bondosa 'Minha mãe  
é bondosa'
- d. Ha'u-nia **aman** mestri  
meu pai  
professor  
'meu pai é professor'

Considerando o critério C, por sua vez, na formação de compostos, as formas de gênero geralmente são realizadas linearmente após os substantivos, como apresentado em (53). No entanto, é interessante ressaltar que as formas *mane*, *feto*, *aman* e *inan* podem ser realizadas antes dos substantivos, gerando construções predicativas como em (54):

(53)

- a. liurai-mane ('rei')
- b. liurai-feto ('rainha')
- c. asu-aman ('cachorro')
- d. asu-inan ('cachorra')

(54)

- |    |             |                     |
|----|-------------|---------------------|
| a. | mane liurai | (‘homem é rei’)     |
| b. | feto liurai | (‘mulher é rainha’) |
| c. | aman asu    | (‘pai é cachorro’)  |
| d. | inan asu    | (‘mãe é cachorra’)  |

É interessante ressaltar, no entanto, que, na posição pré-nominal, a semântica da forma isolada é mantida, o que parece nos sugerir que, nesse contexto, estamos diante de um sintagma [nome + nome]. Ainda assim, consideramos na tabela que esse comportamento aponta para uma classificação das marcas de gênero como formadores de compostos, uma vez que essa mobilidade de linearização não é encontrada em afixos que apresentam posição fixa.

Já em relação à propriedade D, as unidades marcadoras de gênero no TD não se combinam uma variedade muito grande de tipos morfológicos, restringindo-se a adjetivos e substantivos, o que parece se assemelhar mais ao comportamento dos afixos derivacionais que dos compostos.

No que concerne às propriedades estruturais, que agrupam os critérios de E a I<sup>16</sup> na tabela 3, propomos que o núcleo lexical das formações de gênero no TD seja a própria base e não o marcador sugestivo de gênero. Isso porque a semântica do referente da base é mantida, sendo o marcador relacionado ao gênero um elemento que atua como uma espécie de modificador. Não há também evidência de alterações propriamente formais decorrentes da anexação desse marcador. Dessa forma, as formações interpretadas como gênero no TD apresentam núcleo à esquerda, em linha com o comportamento predominante nos compostos na sistematização de Gonçalves e Andrade (2016) no critério E. Já em relação à propriedade F, que diz respeito à possibilidade de coordenação, as marcas sugestivas de gênero *mane*, *feto*, *aman* e *inan* podem se anexar ao primeiro elemento de uma coordenação, mas ter escopo sobre ambos os elementos nela presente, como ilustram os exemplos a seguir:

(55)

- |    |                                  |                               |
|----|----------------------------------|-------------------------------|
| a. | Nia liurai-mane no servisu-na’in | ‘ele é rei e trabalhador’     |
| b. | Nia liurai-feto no servisu-na’in | ‘ela é rainha e trabalhadora’ |

<sup>16</sup> Excluímos da discussão a propriedade em I devido à ausência de marcas flexionais nominais no sistema da língua.

- c. Nia asu-aman no bulak                    ‘ele é cachorro e doido’  
 d. Nia asu-inan no bulak                    ‘ela é cachorra e doida’

Já em relação às propriedades G e H, os marcadores de gênero parecem se alinhar à derivação nesses dois critérios. Isso porque há um número relativamente pequeno de unidades linguísticas no TD com essa função, caracterizando-se, portanto, como uma classe fechada, o que é mais característico dos afixos derivacionais.

Passando para as propriedades de natureza fonológica, que englobam os critérios de J a M<sup>17</sup>, as formações de gênero do TD se alinham à formação de compostos, uma vez que as unidades combinadas mantêm seu acento próprio, constituindo-se como duas palavras fonológicas:

(56)

- a. Labarik-mane    {[labarik] PrWd [mane] PrWd} MWd    ‘menino’  
 b. Busa-inan        {[asu] PrWd [inan] PrWd} MWd            ‘gata’

Em relação às propriedades semânticas, que englobam os critérios de N a Q, podemos dizer que o significado lexical da forma isolada é perdida quando os substantivos *mane*, *feto*, *aman* e *inan* se associam às bases e veiculam a semântica de gênero. Esse comportamento é, segundo a propriedade N, compatível com aquele das formas derivadas, já que nos compostos as unidades expressam um significado lexical.

(57)

	significado lexical	interpretação
a. labarik-mane	criança + homem	menino
b. labarik-feto	criança + mulher	menina

Em geral, esse processo gera uma interpretação semântica composicional estável, em que o significado da forma gerada é previsível a partir dos significados

<sup>17</sup> Excluímos da discussão a propriedade M, uma vez que não conseguimos encontrar uma literatura a respeito dos processos fonológicos envolvidos em processos derivacionais na língua.

das partes que a compõem. Além disso, essa estrutura é massivamente endocêntrica, centrada na base, que define a categoria e o significado do composto. Isso faz essas formações se comportarem de maneira mais próxima ao comportamento das formas derivadas em relação aos critérios de O a Q, segundo a sistematização de Gonçalves e Andrade (2016).

Por fim, no que diz respeito às propriedades R e S, que trazem critérios relacionados à produtividade, os marcadores *mane*, *feto*, *aman* e *inan* também se alinham aos critérios de derivação. Isso porque os marcadores de gênero do TD licenciam conjuntos mais completos de palavras, apresentam paradigmas regulares e consistentes, além de serem formas produtivas na criação de novas palavras que marcam o gênero. Essa regularidade pode ser atestada, por exemplo, na possibilidade da criação de formas marcadas em gênero a partir de substantivos inanimados, como mostramos no exemplo do capítulo anterior, repetido aqui por comodidade:

(58)

- a. **Odamatan-mane** no **odamatan-feto** laran-monu ba  
 malu porta<sub>masc</sub> e porta<sub>fem</sub> se apaixonam
- b. **Odamatan-aman** no **odamatan-inan** laran-monu ba  
 malu  
 porta<sub>masc</sub> e porta<sub>fem</sub> se apaixonam

Outro argumento em favor da regularidade dessas formas é a possibilidade de anexação das marcas de gênero inclusive a empréstimos do português que, na língua de origem, apresentam contraste de gênero marcado de outra forma, como ilustramos também no capítulo 2, repetido em (59) abaixo:

(59)

- a. treballadór-mane ‘trabalhador’  
 b. treballadór-feto ‘trabalhadora’

Em suma, a partir das propriedades sistematizada em Gonçalves e Andrade (2016), se tirarmos aquelas que não conseguimos aplicar – critério I e M – colocamos em discussão um total de 16 critérios. Desses, em 7 critérios detectamos um

comportamento mais próximo da composição e em 9 deles encontramos propriedades que os aproximam da derivação. Essa dualidade de comportamento parece levantar questões importantes. Uma delas é a própria fragilidade desses critérios empíricos, o que mostra que a distinção entre esses dois processos de formação de palavras não é clara. Outro fator interessante de se apontar é que essa mistura de propriedades explica a existência do próprio debate que se coloca sobre as formações de gênero na literatura a respeito do TD. Provavelmente, os diferentes autores estão olhando para diferentes critérios ao atribuírem as suas respectivas classificações. Finalmente, o fato empírico de as formas do TD apresentarem esse comportamento ambíguo entre composição e derivação pode ser interpretado a partir da ideia de que tais elementos apresentam estatuto duplo no sistema – atuando ora como unidade lexical, como elemento gramatical – pelo fato de estar passando por um processo de gramaticalização na língua. Além disso, de um ponto de vista teórico, essa ambiguidade empírica parece apontar para a ideia de que os diferentes processos de formação de palavras devem ser derivados pelos mesmos mecanismos e no mesmo componente da gramática, como propõe o modelo da MD (Halle; Marantz, 1993; Marantz, 1997). É a partir dessas ideias que, no próximo capítulo, desenvolvemos nossa proposta de análise para as formações de gênero do TD.

#### 4.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo buscamos nos aprofundar no debate empírico a respeito das propriedades que mais tradicionalmente distinguem os processos de derivação e composição. Para tanto, recorreremos à sistematização fornecida em Gonçalves e Andrade (2016), que comparam tais processos a partir de suas propriedades mais recorrentes da literatura.

Com o objetivo de verificar, empiricamente, em que medida os dados de gênero no TD se aproximam ou se afastam desses processos, aplicamos as propriedades detectadas em Gonçalves e Andrade (2016) aos dados do TD. Nessa aplicação, verificamos que o comportamento das formações relacionadas a gênero do TD apresenta propriedades tanto de composição, como de derivação.

A partir desse comportamento ambíguo, buscamos uma perspectiva de análise em que tais elementos não são entendidos, nem como parte de composição,

nem como parte de derivação, mas como um dispositivo de classificação nominal que resulta de um processo de gramaticalização a partir do substantivo livre de origem.

## **CAPÍTULO 5: UMA PROPOSTA DE ANÁLISE PARA *MANE*, *FETO*, *AMAN* E *INAN* COMO CLASSIFICADORES NOMINAIS**

Neste capítulo desenvolvemos a proposta de que as propriedades anteriormente descritas dos formadores *mane*, *feto*, *aman* e *inan* do TD são mais bem compreendidas diante da hipótese de que tais elementos atuam como classificadores nominais. Mais especificamente, com base em Aikhenvald (2000), entendemos os classificadores nominais como unidades linguísticas que categorizam o nome ao denotar alguma característica saliente da entidade a que um substantivo associado se refere.

Dessa forma, localizamos os formadores *mane*, *feto*, *aman* e *inan* em um contínuo de gramaticalização em que elementos com uma origem lexical passam a integrar o sistema gramatical da língua, funcionando como um dispositivo de classificação nominal. Nos dados do TD, a origem lexical de *mane*, *feto*, *aman* e *inan* é observada a partir do fato de que tais elementos podem ser empregados isoladamente com conteúdo conceitual distinto da interpretação de gênero. Por outro lado, o estatuto gramaticalizado dessas marcas pode ser identificado em algumas funções morfossintáticas que tais elementos assumem, como no licenciamento de correferência, retomada anafórica e leitura definida, por exemplo.

Além disso, a partir das propriedades elencadas em Aikhenvald (2000) e Grinevald (2000, 2002, 2015), propomos que o estatuto de classificador nominal de *mane*, *feto*, *aman* e *inan* é compatível com o fato de seu emprego não ser obrigatório, de modo que tais elementos não se aplicam a todos nomes animados/humanos da língua.

Finalmente, para colocar em debate a estrutura sintática dessas formações, tomamos como base o modelo da MD (Halle; Marantz, 1993; Marantz, 1997) e propomos uma formalização inicial dos classificadores nominais do TD, tratando-os sintaticamente como a realização de núcleos funcionais adjuntos, que não projetam seu rótulo na estrutura.

Para desenvolver tais discussões, este capítulo está organizado da seguinte maneira: na seção 5.1 revisitamos a literatura que sistematiza as propriedades empíricas dos classificadores nominais e localizamos os marcadores do TD nessa discussão. Na seção 5.2, por sua vez, buscamos propor uma formalização dos

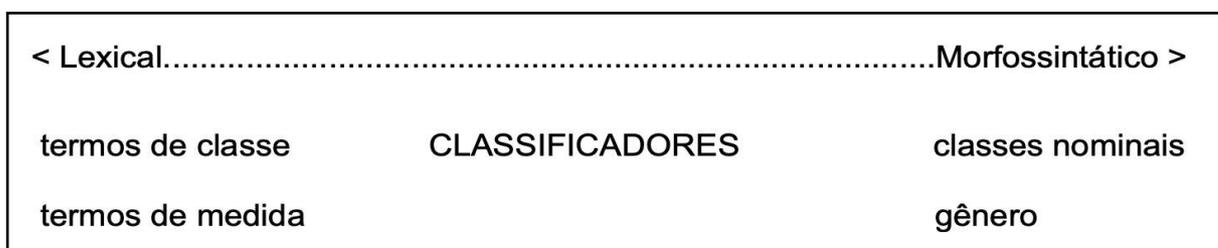
classificadores nominais em uma estrutura sintática, colocando em contraste raízes e núcleos funcionais em termos de estatuto formal. A seção 5.3 encerra o capítulo com as considerações finais.

### 5.1. AS PROPRIEDADES DOS SISTEMAS DE CLASSIFICADORES E OS DADOS DO TÉTUM-DÍLI

De maneira geral, as línguas naturais apresentam uma diversidade de ferramentas que desempenham a função de classificação dos seus nominais. Segundo Grinevald (2002, 2015), os sistemas de classificação linguística podem ser identificados a partir de uma tipologia que considera o estatuto desses elementos no âmbito da distinção lexical vs. gramatical.

Translinguisticamente, os sistemas de classificação englobam elementos que se distribuem em um *continuum* de gramaticalização, tal como esquematizado a seguir:

Figura 3 – Visão geral dos sistemas de classificação no *continuum* de gramaticalização



Fonte: Baseado em Grinevald, 2002, p. 260.

No extremo lexical do esquema acima, estão o que Grinevald (2002, 2015) chama de termos de classe e termos de medida. Os termos de medida aparecem em construções expressando quantidades e tipos de arranjos, tal como exemplificado a seguir:

- (60) Exemplos de termos de medida do inglês
- a. a glass of water, a pound of sugar, a slice of bread
  - b. a pile of books, a group of children, a line of cars

(Grinevald, 2002, p. 260)

Os termos de classe, por sua vez, são itens lexicais que participam de processos de composição na formação de palavras, sendo funcionalmente, segundo a autora, equivalentes a processos derivacionais, tais como nos exemplos a seguir:

- (61) Exemplos de termos de classe
- a. Strawberry, blueberry, raspberry, boysenberry, gooseberry
  - b. apple tree, banana tree, orange tree, cherry tree
  - c. mailman, policeman, garbage man

(Grinevald, 2002, p. 261)

No domínio morfossintático, os classificadores podem ser diferenciados de sistemas de concordância, como o gênero e os sistemas de classe nominal, como aqueles encontrados nas línguas bantu. Isso porque elementos mais gramaticalizados, como gênero e classe nominal, se caracterizam por terem realização fora do próprio substantivo, como em modificadores adjetivais, determinantes e possessivos, por exemplo.

Os classificadores nominais, por sua vez, são intermediários nesse *continuum*, tendo uma origem lexical, mas funcionando, em alguns aspectos, como um elemento funcional na língua. Propomos que esse seja exatamente o percurso pelo qual as formas *mane*, *feto*, *aman* e *inan* estejam passando no TD. Em processo de gramaticalização, tais elementos podem ser empregados isoladamente com um conteúdo conceitual próprio ou se associar a substantivos da língua, licenciando a explicitação do gênero ao atuar como classificadores nominais.

Propomos que uma das diferenças entre o comportamento dessas peças como classificadores ou palavras independentes é a sua posição de realização. Mais especificamente, como classificadores tais marcas aparecem à direita do elemento a que se anexam. No entanto, quando atuam como substantivos independentes, tais elementos aparecem à esquerda.

- |      |                     |                      |
|------|---------------------|----------------------|
| (62) | a. Mane oin-furak   | ‘homem bonito’       |
|      | b. Feto matenek     | ‘mulher inteligente’ |
|      | c. Aman laran-luak  | ‘pai generoso’       |
|      | d. Inan terus-na'in | ‘mãe sofredora’      |

Nos dados acima, observamos que a semântica conceitual dos substantivos *mane* (‘homem’), *feto* (‘mulher’), *aman* (‘pai’) e *inan* (‘mãe’) está presente na formação.

Além disso, quando tais elementos aparecem à esquerda do substantivo, é possível separá-los do elemento a que eles se relacionam por outros itens linguísticos, como pelo marcador anafórico *ne'e*:

- (63) a. Feto ne'e matenek  
'essa/a mulher é inteligente'
- b. Mane ne'e oin-furak  
'esse/o homem é bonito'
- c. Inan ne'e terus-na'in  
'essa/a mãe é sofredora'
- d. Aman ne'e laran-luak  
'esse/o pai generoso'
- e. Asu metan ne'e aman  
'O cachorro é preto'
- f. Asu metan ne'e inan  
'A cachorra é preta'

Além do marcador anafórico, é possível adicionar ainda mais material sintático entre *mane*, *feto*, *aman* e *inan* e o substantivo relevante. Nos dados em (64), por exemplo, há a presença de adjetivos intervenientes:

- (64) a. Labarik tanis-te'en ne'e mane  
'a criança chorona é homem'
- b. Labarik oin-furak ne'e feto  
'a criança bonita é mulher'

Essa flexibilidade de distribuição é uma característica típica de elementos lexicais, o que sugere que, nessas construções, *mane*, *feto*, *aman* e *inan* atuam como nomes independentes e não como classificadores. Como classificadores, no entanto, em que as marcas de gênero se realizam à direita da base, o classificador parece estabelecer com a base uma relação de modificação, em que *mane*, *feto*, *aman*, e *inan* especificando o gênero do substantivo, como nos dados abaixo repetidos do capítulo 3 para facilitar a visualização:

(65)

a.	labarik	'criança'	labarik-mane	'menino'	labarik-feto	'menina'
b.	oan	'filho/a'	oan-mane	'filho'	oan-feto	'filha'
c.	banin	'sogro/a'	banin-mane	'sogro'	banin-feto	'sogra'

(66)

a.	fahi	'porco/a'	fahi-aman	'porco'	fahi-inan	'porca'
b.	asu	'cachorro/a'	asu-aman	'cachorro'	asu-inan	'cachorra'
c.	manu	'galo/inha'	manu-aman	'galo'	manu-inan	'galinha'

(Hull e Eccles, 2005, p.23-24)

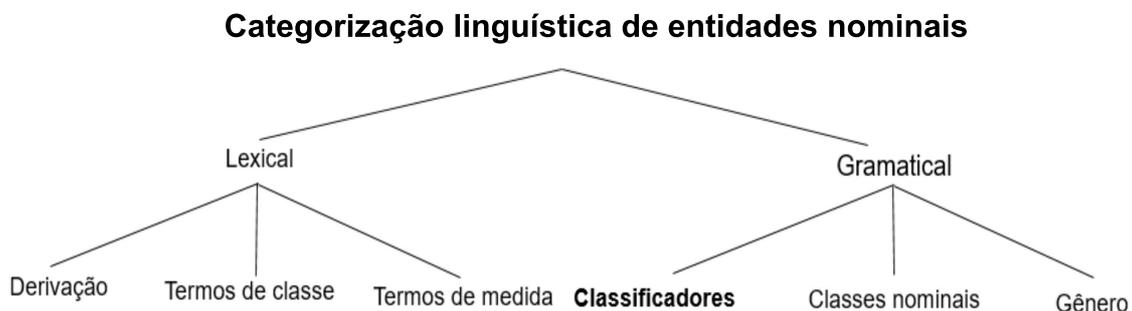
O caráter ora lexical, ora gramatical desses elementos no TD pode ser evidenciado também pela possibilidade de coocorrência entre esses dois elementos, ou seja, o nome, como forma livre, pode ser diretamente concatenado a um classificador nominal, como exemplificado nas sentenças a seguir:

(67)

- a. Feto-inan                      bulak    ne'e    hamnasa la    para  
 MULHER-CL:FÊMEA    doida    esta    rir            não parar  
 'A mulher doida não para de rir'.
- b. Mane-aman                      boot    ne'e    foin    to'o  
 HOMEM-CL:MACHO    grande    este    acabou chegar  
 'O homem grande acabou de chegar'

Nos dados em (67), as formas livres *feto* ('mulher') e *mane* ('homem') aparecem associadas aos classificadores nominais *inan* ('fêmea') e *aman* ('macho'). Essa combinação parece trazer uma contribuição semântica na forma de uma espécie de ênfase ao estado denotado pelo adjetivo.

classificadores nominais refletem, então, a atuação de elementos com origem lexical que se gramaticalizaram na língua, passando a atuar no domínio funcional. Nesse cenário, tais classificadores dividem espaço com a atuação de outros mecanismos de classificação, como as classes nominais/sistemas de gênero das línguas:

Figura 4: Visão geral dos sistemas em um *continuum* de gramaticalização

Fonte: Baseado em Grinevald (2015, p. 812).

Buscando, distinguir as propriedades dos elementos gramaticais de classificação, ou seja, das classes nominais/gênero de um lado e classificadores nominais de outro, Grinevald (2002) apresenta o seguinte conjunto de propriedades:

Tabela 5 – Classificadores vs. sistemas mais gramaticalizados de classificação

	<b>CLASSES NOMINAIS</b>	<b>CLASSIFICADORES NOMINAIS</b>
<b>a.</b>	classificam todos os nomes	não classificam todos os nomes
<b>b.</b>	em um pequeno número de classes	em número maior
<b>c.</b>	sistema fechado	sistema aberto
<b>d.</b>	fundido com outras categorias gramaticais (número, caso...)	não fundido com outras categorias gramaticais
<b>e.</b>	Pode ser marcada no N	não marcada no N propriamente
<b>f.</b>	envolvido em padrões de concordância	não é parte dos sistemas de concordância
<b>g.</b>	Nome atribuído a uma classe	Nome pode ser atribuído a várias classes
<b>h.</b>	Não está sujeito à variação pelos faltantes	Pode estar sujeito à variação pelos falantes
<b>i.</b>	Sem variação de registro	Possível variação de registro formal vs. informal

Fonte: Grinevald (2002, p.260)

Para colocar em discussão as propriedades contrastadas na tabela acima, vamos selecionar algumas características que consideramos cruciais para a compreensão dos dados do TD. Entre elas, destacamos inicialmente as propriedades em (a), (f), (h) e (i), sendo as demais consideradas secundárias a partir do próprio sistema da língua, como veremos mais adiante.

Em relação à propriedade em (a), ou seja, a opcionalidade dos classificadores nominais, no TD, a classificação nominal é semanticamente motivada, já que se restringe a substantivo humanos e animados. No entanto, é importante salientar que nem todo substantivo com esses traços necessita ser acompanhado de classificador. Como vimos no capítulo 4, em diversos contextos essa marcação é opcional.

- |      |                                  |                            |
|------|----------------------------------|----------------------------|
| (68) | a. liurai katuas laran luak      | ‘o rei idoso é bondoso’    |
|      | b. liurai mane katuas laran luak | ‘o rei idoso é bondoso’    |
|      | c. liurai ferik laran luak       | ‘a rainha idosa é bondosa’ |
|      | d. liurai feto ferik laran luak  | ‘a rainha idosa é bondosa’ |

(baseado em Choupina, 2011, p. 73)

Outra propriedade relevante a ser retomada nessa discussão é que adjetivos do TD permanecem invariáveis, mesmo quando modificam substantivos humanos ou animados atrelados a uma classificação nominal de gênero, tal como ilustrado nos seguintes exemplos também retomados do capítulo 4:

- |      |                               |                              |
|------|-------------------------------|------------------------------|
| (69) | a. labarik mane aas bá iskola | ‘o menino alto vai à escola’ |
|      | b. labarik feto aas bá iskola | ‘a menina alto vai à escola’ |

Nesses exemplos, o adjetivo *aas* (‘alto/alta’) é invariável, mantendo, portanto, uma mesma forma, independentemente de se referir ao masculino ou ao feminino. Isso mostra que, diferentemente das classes nominais, o sistema de classificadores do TD não está envolvido em um processo de concordância, como previsto na tabela 5 em (f).

Além disso, quando olhamos para a associação entre empréstimos do português que entram na língua já marcados para gênero, é possível verificar no emprego de *mane*, *feto*, *aman*, e *inan* uma variação, relacionada a um efeito sociolinguístico. Isso porque falantes mais escolarizados e com maior domínio do

português tendem a evitar o emprego das marcas nativas de gênero associadas a formas já marcadas pelo contraste fonológico do português, enquanto falantes que apresentam menos contato com o português associam tais formas aos classificadores de gênero da língua mais livremente, como nas formas em (70b) e (70d). Esse cenário está em consonância com as propriedades em (h) e (i) relacionadas por Grinevald (2002) aos classificadores nominais.

- (70) a. *traballadora*  
 b. *traballadora-feto*
- c. *kantora*  
 d. *kantora-feto*

Retomando agora as propriedades que consideramos secundárias, no que diz respeito ao tamanho do inventário de classificadores nominais, identificado em (b) e (c), salientamos que esse não é um ponto crucial para a distinção entre classificadores nominais e gênero, uma vez que, como aponta Aikhenvald (2000), trata-se de um aspecto que pode variar translinguisticamente. Assim, há desde línguas que apresentam um conjunto fechado de classificadores, até um conjunto aberto relativamente grande desses elementos, o que por sua vez pode estar relacionado com o grau de gramaticalização dessas peças na língua.

O conjunto pequeno de classificadores nominais do TD também torna a discussão sobre a propriedade em (g) menos central. Nessa propriedade prevê-se que, no sistema de classificação nominal, um mesmo nome pode ser atribuído a várias classes. Embora o inventário do TD seja restrito, o padrão é que esse seja o caso. Assim, para além da distribuição dos traços de [humano] e [animado], a classificação não é efetivamente imposta ou selecionada pelo nome de base, com um mesmo nome aparecendo ora com *mane*, ora com *feto*, por exemplo.

Já em relação à propriedade em (d), que diz respeito à ideia de que gênero, diferentemente das classes nominais, poderia aparecer fundido com outras categorias gramaticais (como número e caso, por exemplo), é importante salientar que como o TD é uma língua isolante, não há muitos contextos de marcação gramatical em que essa fusão poderia acontecer na língua. De qualquer forma, há evidência de que a classificação nominal baseada nos traços [animado] e [humano] é, por exemplo, distinta da marcação de plural realizada via *sira*. Isso porque, enquanto as classes

nominais são semanticamente restritas, a marcação de plural não segue essa mesma restrição.

Outro ponto bastante relevante no funcionamento dos classificadores nominais translinguisticamente é que seu comportamento gramaticalizado pode ser detectado no papel característicos de elementos funcionais que ele vai assumindo dentro do sistema. No TD, por exemplo, é possível verificar contextos, como o de coordenação, em que um classificador nominal consegue licenciar escopo sobre uma base que é anexada a apenas ao primeiro classificador, como ilustrado a seguir:

- (71) a. *Labarik-feto mane sira hananu*  
 criança-CL:FEM CL:MASC PL cantam  
 ‘as meninas e os meninos cantam’

No exemplo em (71a) a base *labarik* (‘criança’) está diretamente anexada ao classificador *feto*. No entanto, o classificador *mane*, apesar de não ser acompanhado de uma base realizada, é semanticamente associado a essa mesma base nominal, gerando a interpretação coordenada de ‘meninas’ e ‘meninos’.

Além das questões de escopo, os classificadores nominais do TD podem desempenhar um papel funcional no licenciamento de relações de correferência. Isso evidencia o caráter mais funcional e menos lexical desses elementos no sistema da língua.

- (72) a. *Nia oan-feto tanis lós!*  
 sua filha-CL:FEM chora muito  
 ‘Sua filha chora muito’
- b. *Feto boot ona jeitu aat lós.*  
 CL:FEM grande já jeito ruim muito.  
 ‘Ela já é adulta ainda se comporta mal’
- (73) a. *Nia alin-mane gosta toba de’it!*  
 seu irmão-CL:HOMEM gosta dormir só!  
 ‘seu irmão mais novo só gosta de dormir’
- b. *Mane boot ona baruk-teen lós.*  
 CL:HOMEM grande já preguiçoso muito  
 ‘ele já é adulto é muito preguiçoso’

No diálogo em (72) e em (73), os referentes relevantes no discurso *oan-feto* (‘filha’) e *alin-mane* (‘irmão’), respectivamente, são estabelecidos na primeira

sentença. Na segunda sentença, por sua vez, esses referentes são retomados apenas pelo classificador nominal, sem a necessidade de se repetir a base, licenciando a correferencialidade a partir da interpretação estabelecida pelo referente introduzido na primeira sentença.

Assim, tais elementos que originalmente funcionavam como substantivos independentes parecem ter adquirido um papel mais funcional no sistema, desempenhando um papel relevante no licenciamento na correferencialidade. Essa retomada correferencial pode ser vista também em contextos de pergunta e resposta, em que o referente é introduzido na pergunta pela associação entre substantivo e classificador, mas retomado na resposta apenas pelo classificador nominal:

(74) Pessoa A: *João haree labarik-feto iha shopping?*  
 João viu criança-CL:FEM no shopping?  
 'João viu a menina no shopping?'

Pessoa B: *Sin, feto sosa hela roupa*  
 Sim, CL:FEM comprar ainda roupa  
 Sim, ela estava comprando roupa'

(75) Pessoa A: *Professor hirus labarik-mane ne'e?*  
 Professor zangou criança-CL:MASC este?  
 'O Professor ficou bravo com esse menino?'

Pessoa B: *Sin, mane ulun-toos liu*  
 Sim, CL:MASC teimoso muito  
 'Sim, ele é muito teimoso'

Esse mesmo comportamento gramaticalizado nas retomadas correferenciais pode ser desempenhada pelas formas *aman* e *inan*, como exemplificado a seguir:

(76) Pessoa A: *Asu-aman ne'e kekas lós*  
 cachorro-CL:MACHO este magro muito  
 'Este cachorro está muito magro'

Pessoa B: *Aman ida de'it tan*  
 CL:MACHO um só mais  
 'Só tenho ele'

(77) Pessoa A: *Asu-inan ne'e fa'an tiha ona*  
 Cachorra-CL:FÊMEA este vender.passado já  
 'A cachorra já foi vendida'

Pessoa B: Inan            si'ak        liu  
 CL:FÊMEA    agressiva    muito  
 'Ela é muito agressiva'

Nesta seção argumentamos em favor do estatuto de classificador nominal dos elementos *mane*, *feto*, *aman* e *inan* do TD a partir de seu comportamento empírico. Com isso em mente, na próxima seção, propomos uma formalização inicial dessas formações tomando como base o modelo da Morfologia Distribuída (Halle; Marantz, 1993; Marantz, 1997).

## 5.2. UMA PROPOSTA DE FORMALIZAÇÃO PARA OS CATEGORIZADORES NOMINAIS DO TÉTUM-DÍLI

A formalização de elementos como os classificadores nominais é uma questão ainda controversa na teoria linguística. Isso porque, como vimos ser o caso do TD, tais elementos apresentam um estatuto intermediário entre propriedades lexicais e funcionais.

Em um modelo decomposicional como a MD, em que o estatuto das peças é definido por sua posição na estrutura sintática, podemos levantar a hipótese de que *mane*, *feto*, *aman* e *inan* são licenciados em diferentes lugares na estrutura, de modo a ter um estatuto duplo no sistema. Assim, tais elementos podem ocupar uma posição de raiz que, quando categorizada, será associada a um conteúdo conceitual na Lista 3. Da mesma forma, tais elementos podem ocupar uma posição de núcleo funcional, atuando como uma peça gramatical da língua. A nossa ideia, portanto, é que *mane*, *feto*, *aman* e *inan* estão em um processo de gramaticalização, passando de raízes para núcleos funcionais.

Para desenvolver esse raciocínio, passamos inicialmente por uma visão geral de raiz para, em seguida, colocar em discussão as peças envolvidas no domínio funcional dos nomes.

### 5.2.1. RAÍZES E CATEGORIZAÇÃO NA MORFOLOGIA DISTRIBUÍDA

O modelo da MD toma como primitivos do sistema dois tipos de elementos: os traços morfossintáticos e as raízes. Nesta seção, buscando discutir o estatuto dos formadores *mane*, *feto*, *aman* e *inan* do TD, nos focamos na natureza das raízes, em

especial buscando as evidências fornecidas no âmbito do modelo para desatrelar as noções de raiz e categoria.

As raízes são necessariamente assumidas como primitivos da MD a partir da assunção de decomposição plena, ou seja, de que todos os objetos linguísticos formados por mais de um morfema, são derivados sintaticamente toda vez que são empregados. Como consequência, não há palavras prontas armazenadas no léxico, sendo tais elementos entendidos como a combinação de primitivos (raízes e traços) na sintaxe. A respeito da natureza das raízes há ainda muito debate acontecendo no modelo, mas em relação à categoria, há um consenso de que as raízes são destituídas dessa informação.

De um ponto de vista empírico, uma evidência da ausência de categoria nas raízes é o fato de que é comum encontrarmos nas línguas naturais raízes que podem aparecer em mais de uma categoria.

(78)	a. √CERT	acertar	certeza	certamente
	b. √SIMPL	simplificar	simplicidade	simplesmente

Os dados acima parecem mostrar que as raízes têm estatuto categorial indefinido, podendo se tornar verbos, nomes ou advérbios. Além disso, quanto mais informação morfossintática está presente, mais clara fica a categoria da formação. Tais dados parecem apontar para o fato de que é na relação que a raiz estabelece com outros elementos da estrutura sintática que a categoria é estabelecida.

Nesse mesmo sentido, muitas vezes uma mesma realização morfológica de um elemento linguístico pode licenciar a sua inserção em diferentes categorias, como nos dados a seguir:

(79)

- a. A Maria gosta de cantar.
- b. O cantar dos pássaros.
  
- c. O menino bonito chegou.
- b. O bonito chegou.
  
- c. O pedreiro é rápido.
- d. O pedreiro terminou rápido a obra.

Nos dados em (79) podemos ver que a categoria dos itens sublinhados em cada uma das sentenças é definida distribucionalmente nas relações sintáticas que são estabelecidas. Isso mais uma vez parece evidenciar que a categoria é atribuída sintaticamente, não sendo uma propriedade da raiz.

As línguas naturais frequentemente apresentam em seu inventário de peças morfológicas que determinam a categoria da formação e que não são propriamente parte da raiz, como *-mento*, *-vel* e *-mente* do português formadores, respectivamente, de substantivos, adjetivos e advérbios na língua.

- (80) a. casar[V]    casamento[N]  
 b. lavar[V]    lavável[A]  
 c. simples[A]    simplesmente[ADV]

Tais elementos definidores de categoria são conhecidos no modelo da MD como núcleos categorizadores. Os categorizadores se associam à raiz na sintaxe para definição de um domínio categorial. Essa ideia é formalizada na hipótese de categorização:

Hipótese de Categorização: Raízes não podem aparecer (não podem ser pronunciadas ou interpretadas) sem serem categorizadas; elas são categorizadas por meio de concatenação sintática com núcleos funcionais definidores de categoria.<sup>18</sup>

(Embick; Marantz, 2008, p. 6)

A assunção de decomposição plena e a hipótese de categorização tomadas em conjunto apresentam como consequência a ideia de que todo item linguístico que apresenta uma categoria é necessariamente complexo, ou seja, formado pela concatenação entre uma raiz acategorial e um núcleo categorizador.

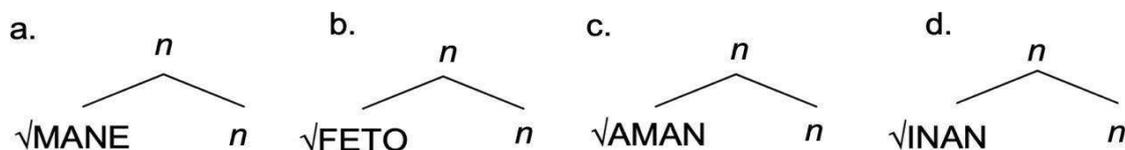
A partir desse panorama, propomos que, quando empregados como substantivos livres na língua, as formas *mane*, *feto*, *aman* e *inan* do TD atuam como

---

<sup>18</sup> Categorization assumption: Roots cannot appear (cannot be pronounced or interpreted) without being categorized; they are categorized by merging syntactically with category defining functional heads." (Embick; Marantz, 2008, p. 6 – tradução nossa).

raízes que, por sua vez, são concatenadas a um núcleo categorizador nominal, como ilustrado nas representações a seguir:

(81)



No entanto, quando esses elementos atuam como classificador nominal, eles perdem seu conteúdo conceitual e passam a licenciar relações estabelecidas por elementos funcionais, como correferencialidade e relações de escopo. Perseguindo esse raciocínio, na próxima seção, buscamos delinear de modo mais preciso quais são os núcleos funcionais assumidos no domínio nominal, com o intuito de localizar os classificadores nessa discussão.

### 5.2.2 NÚCLEOS FUNCIONAIS NO DOMÍNIO NOMINAL

Para sustentar a discussão sobre os núcleos funcionais no domínio nominal, tomamos como ponto de partida a visão desenvolvida por Wiltschko (2014) de que, embora o inventário de núcleos funcionais ativos nos sistemas linguísticos esteja sujeitos à variação de uma língua para a outra, as categoriais funcionais estariam distribuídas em domínios sintáticos comuns que são universais. Esse panorama é interessante para a discussão do estatuto formal dos classificadores nominais já que tais elementos não estão presentes em todas as línguas.

A proposta de Wiltschko (2014), conhecida como “Hipótese da Coluna Universal”<sup>19</sup> (HCU), estabelece uma estrutura hierárquica de categorias universais ( $\kappa$ ), que servem como base para a organização das categorias funcionais (c) nas línguas naturais, tal como sistematizado a seguir:

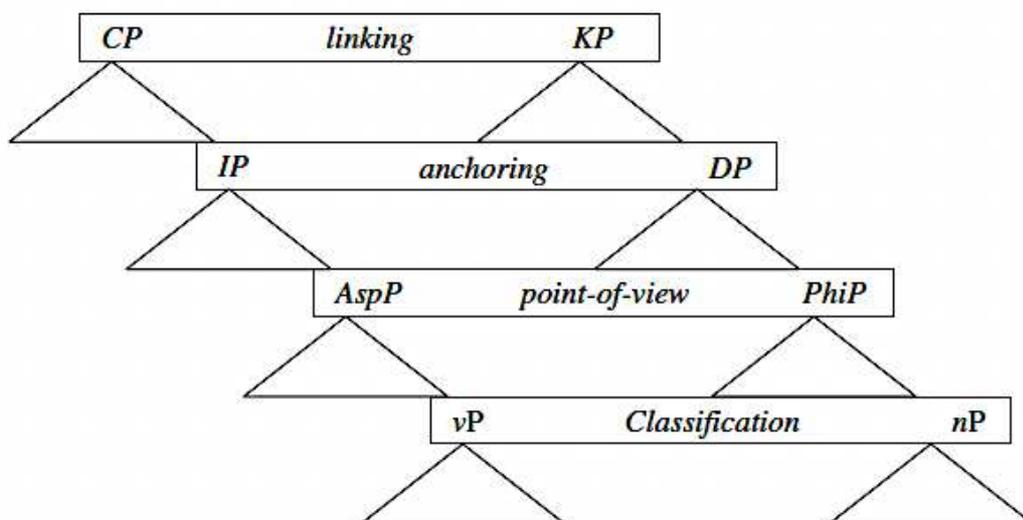
<sup>19</sup> No inglês, *The Universal Spine Hypothesis*. Estamos adotando a tradução proposta em Nóbrega (2018), referência que remetemos ao leitor para uma discussão ampliada dessa proposta em oposição a outras abordagens consagradas na literatura a respeito da organização dos núcleos funcionais nas línguas naturais.

(i) Categorias específicas de cada língua (c) são construídas a partir de um pequeno conjunto de categorias universais  $\kappa$  [...]; (ii) o conjunto de categorias universais  $\kappa$  é organizado hierarquicamente, sendo que cada camada de  $\kappa$  é definida por uma função única.

Wiltschko (2014, p. 24<sup>20</sup>)

Mais especificamente, a proposta de Wiltschko (2014) é que a organização estrutural da HCU é baseada em um conjunto de categorias universais e hierárquicas sistematizadas a seguir, com seus respectivos paralelismos entre os domínios verbais e nominais:

(82) linking (ligação discursiva) > anchoring (ancoragem) > point-of-view (ponto de vista) > classification (classificação)



Wiltschko (2014, p. 24, p. 78)

Os domínios funcionais universais representados em (22) são considerados neutros em relação à categoria, ou seja, a mesma organização de domínios serve como base para a construção da coluna funcional associada tanto a verbos, como a

<sup>20</sup> "(i) Language-specific categories (c) are constructed from a small set of universal categories  $\kappa$  and language-specific UoLs [Units of Language]; (ii) the set of universal categories  $\kappa$  is hierarchically organized where each layer of  $\kappa$  is defined by a unique function". Wiltschko (2014, p. 24 – tradução nossa).

nomes. A função de cada uma dessas camadas tal como sistematizada em Nóbrega (2018) segue nos pontos abaixo:

- ***κ:classification (classificação)*** – tem a função de determinar se a estrutura sintática a ser derivada irá dar origem a um evento ou a um indivíduo. Em outras palavras, essa categoria “classifica” uma raiz lexical como verbo ou nome. *κ:classification* serve, portanto, como base para a formação dos núcleos categorizadores, *v* e *n*.
- ***κ:point-of-view (ponto de vista)*** – introduz uma perspectiva acerca do evento ou do indivíduo. No domínio verbal, por exemplo, *κ: ponto de vista* é normalmente realizada por uma categoria aspectual, que adiciona um tempo de referência ao evento. Esse tempo de referência agrega um ponto de vista com relação ao evento descrito pelo verbo. Já no domínio nominal, essa mesma categoria promove a formação dos traços  $\varphi$ , tais como gênero e número.
- ***κ:anchoring (ancoragem)*** – insere o evento ou o indivíduo no espaço e/ou no tempo de enunciação. No domínio verbal, ela é instanciada comumente pela categoria tempo, enquanto no domínio nominal ela é realizada por um núcleo determinante, instanciado por um artigo definido ou por um demonstrativo.
- ***κ:linking (vinculação discursiva)*** – vinculação estabelece uma relação entre a proposição ou o referente e o discurso em andamento. A autora atrela essa categoria universal ao núcleo complementizador C, no domínio verbal, e a um núcleo de Caso, K, no domínio nominal.

Nóbrega (2018, p.157-158)

Focando, mais especificamente no domínio nominal – lugar de interesse para a localização dos classificadores nominais – apontamos como relevante para a discussão uma revisão proposta em Nóbrega (2018) especificamente na camada *κ: ponto de vista*. Na proposta de Wiltschko (2014) essa categoria se relaciona aos traços  $\varphi$ , como gênero e número, por exemplo. Nóbrega (2018), no entanto,

estabelece uma dissociação entre os traços que expressam c:NÚMERO e c:GÊNERO. Nessa distinção proposta pelo autor, os valores atrelados à c:GÊNERO, bem como às demais formas gramaticais empregadas para a classificação dos nomes, como as classes nominais, são determinados através de κ:classificação. Por sua vez, a categoria κ:ponto de vista está atrelada, na perspectiva do autor, a c:NÚMERO.

A combinação de gênero e classe nominal sob uma mesma camada κ:classificação se justifica uma vez que gênero e classes nominais apresentam uma distribuição formal semelhante (Nóbrega, 2018, p. 178):

#### (83) Gênero e classes nominais

- São categorias funcionais regulares;
- São categorias obrigatórias, ou seja, todo nome em uma língua com sistemas de classificação deve ser atribuído a um gênero ou a uma classe;
- Apresentam relações de dependência para fora do nome e do sintagma nominal.

Essa similaridade de comportamento entre gênero e classe nominal vai ao encontro das nossas discussões anteriores de que os sistemas de classificação linguística podem ser identificados a partir de uma tipologia que considera o estatuto desses elementos em um *continuum* de gramaticalização (Grinevald, 2000, 2015). Enquanto gênero e classe nominal ocupam o polo morfossintático dessa escala, os classificadores nominais, por sua vez, são intermediários nesse *continuum*, tendo uma origem lexical, mas funcionando, em alguns aspectos, como um elemento funcional na língua.

O contraste entre as propriedades de gênero/classe nominal e o funcionamento dos classificadores nominais no TD apontam para a ideia de que a camada de classificação, que abrange gênero e classe nominal, não pode ser diretamente aplicada aos classificadores nominais. Mais especificamente, *mane*, *feto*, *aman* e *inan* apresentam o seguinte conjunto de propriedades em contraste a (83):

#### (84) Classificadores nominais

- Apresentam origem lexical, podendo funcionar como substantivos livres na língua;

- Quando atuam como classificadores, não são elementos obrigatórios, ou seja, nem todo nome na língua se associa a esses formadores;
- Não desencadeiam relações de concordância.

Outra propriedade relevante para a nossa discussão é que os classificadores nominais não parecem alterar a categoria da formação. Esse ponto está relacionado ao fato de que os nomes sem classificadores são perfeitamente gramaticais na língua.

Considerando que os elementos podem ser anexados na estrutura sintática de duas maneiras distintas: como núcleos ou como adjuntos, em linha com Nóbrega (2018), propomos que classificadores nominais, como *mane*, *feto*, *aman* e *inan* do TD, são sintaticamente adjuntos. Assim, esses elementos são concatenados à estrutura, mas não por requerimentos de traços morfossintáticos, atuando como modificadores da estrutura a que se anexam. Dessa forma, os adjuntos não projetam seu rótulo na estrutura sintática. Tais propriedades são sistematizadas no contraste abaixo:

Tabela 6 – Núcleos e Adjuntos

Propriedade	Núcleos	Adjuntos
Rotulação	projetam seu rótulo na estrutura	Não projetam seu rótulo na estrutura
Propriedades Formais	definem as propriedades formais da estrutura	Não alteram as propriedades formais da estrutura
Motivação	São requeridos por traços morfossintáticos	Não satisfazem traços morfossintáticos

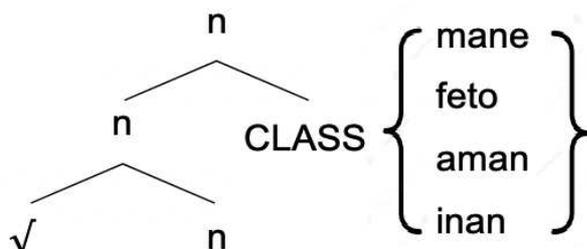
Fonte: baseado em Nóbrega, Bassani e Armelin (2022, p. 239)

Com a anexação de *mane*, *feto*, *aman* ou *inan* à estrutura a semântica do referente da base é mantida, sendo o classificador um elemento que atua como uma espécie de modificador. Não há também evidência de alterações propriamente formais decorrentes da anexação desse marcador.

Em relação ao mecanismo de concatenação, dois subtipos de Merge foram propostos por Chomsky (2000, 2004) para capturar as assimetrias entre argumento e adjunto: *Set-Merge* e *Pair-Merge*. Mais especificamente, ao tomar dois objetos sintáticos, o *Set-Merge* gera conjuntos simples e não ordenados, enquanto o *Pair-Merge* gera pares ordenados. A partir do raciocínio de que os classificadores

nominais são inseridos na estrutura via *Pair-Merge*, ilustramos as formações do TD a partir da seguinte estrutura:

(85)



Na estrutura em (85), a entrada do classificador nominal, sob o rótulo *CLASS*, não altera a categoria da formação que havia sido estabelecida pelo categorizador nominal *n* na concatenação com a raiz. Um último ponto a ser ressaltado é que, embora os classificadores *mane*, *feto*, *aman* e *inan* no TD não atuem como núcleos categorizadores, é importante ressaltar que eles apresentam alguma correlação com o processo de categorização, uma vez que eles parecem definir um domínio de interpretação categorial para o elemento com que eles se concatenam. Esse raciocínio abre a perspectiva de compreendermos o processo de gramaticalização como um cenário em que um elemento que originalmente é uma raiz no sistema passa a conseguir funcionar como um elemento capaz de definir uma categoria para a estrutura com a qual ela se concatena.

### 5.3 CONSIDERAÇÕES FINAIS

Neste capítulo desenvolvemos a proposta de que os formadores *mane*, *feto*, *aman* e *inan* do TD apresentam o estatuto de classificadores nominais. Mais especificamente, tais elementos, elementos apresentam uma origem lexical na língua, mas estão passando também a integrar o sistema gramatical, funcionando como um dispositivo de classificação nominal.

Para implementar essa ideia, retomamos as propriedades empíricas dos classificadores nominais em contraste com as classes nominais/gênero a partir da proposta de Grinevald (2002). Mais especificamente, os formadores *mane*, *feto*, *aman* e *inan* do TD, como previsto para os classificadores nominais, não são

elementos obrigatórios, ou seja, nem todo nome na língua se associa a esses formadores e não estão envolvidos em relações de concordância.

Buscando contribuir para o debate sobre a formalização desses classificadores nominais, tomamos como ponto de partida a visão desenvolvida por Wiltschko (2014) de que as categorias funcionais estariam distribuídas em domínios sintáticos comuns que são universais. A partir das reflexões de Nóbrega (2018), entendemos a camada de k:classificação como o locus das informações de gênero e classes nominais. Em contraste a tais elementos, por sua vez, propomos que os classificadores nominais *mane*, *feto*, *aman* e *inan* do TD são sintaticamente adjuntos. Dessa forma, esses elementos são concatenados à estrutura, mas não por requerimentos de traços morfossintáticos, atuando como modificadores da estrutura a que se anexam.

## CAPÍTULO 6: CONSIDERAÇÕES FINAIS

Nesta pesquisa buscamos descrever e analisar as marcas de gênero (*mane*, *feto*, *aman* e *inan*) do TD, com o objetivo de investigar seu estatuto formal e propor uma estrutura sintática que seja capaz de explicar o comportamento dessas formações.

Para tanto, tomamos como ponto de partida uma controvérsia existente na literatura em relação ao estatuto das formas *feto*, *mane*, *aman* e *inan* e do processo de formação de palavras responsável por gerar os elementos linguísticos complexos construídos com essas marcas de gênero. Mais especificamente, alguns autores defendem que tais elementos são formalmente sufixos que participam de um processo de derivação (Hull e Eccles, 2005; Belo e Oliveira, 2021), sendo as formas de gênero consideradas, então, morfologicamente derivadas. Por sua vez, outros autores (Costa, 2015; Choupina, 2011) defendem que os marcadores de gênero são palavras independentes e que, portanto, as formações de gênero são geradas via processo de composição.

Inserindo-nos nesse debate, de um ponto de vista empírico recorreremos às características associadas aos processos derivacionais e de composição com base na sistematização fornecida em Gonçalves e Andrade (2016). Com base nessas propriedades, fizemos uma descrição do funcionamento das marcas de gênero no TD que evidenciou que essas formas apresentam um comportamento não estável, se comportando ora como derivação, ora como composição a depender da propriedade empírica que está sendo considerada. A partir dessa percepção, levantamos a hipótese de que tais elementos não atuam nem como parte de uma derivação, nem como membro de uma composição, mas sim como classificadores nominais.

Dessa forma, localizamos os formadores *mane*, *feto*, *aman* e *inan* em um contínuo de gramaticalização em que elementos com uma origem lexical passam a integrar o sistema gramatical da língua, funcionando como um dispositivo de classificação nominal. Para implementar essa ideia, retomamos as propriedades empíricas dos classificadores nominais em contraste com as classes nominais/gênero a partir da proposta de Grinevald (2002). Mais especificamente, os formadores *mane*, *feto*, *aman* e *inan* do TD, como previsto para os classificadores

nominais, não são elementos obrigatórios, ou seja, nem todo nome na língua se associa a esses formadores e não estão envolvidos em relações de concordância.

De um ponto de vista teórico, então, adotamos a abordagem sintática de formação de palavras proposta pela MD (Halle e Marantz, 1993; Marantz, 1997 e muitos trabalhos subsequentes), que se justifica como um quadro interessante ao propor que a distinção entre as expressões linguísticas é derivada do estatuto formal das peças envolvidas e do modo como tais peças se organizam no interior da estrutura sintática. Um modelo decomposicional como a MD, nos fornecesse ferramentas para a compreensão de que *mane*, *feto*, *aman* e *inan* podem ser licenciados em diferentes lugares na estrutura, de modo a ter um estatuto duplo no sistema. Assim, tais elementos podem ocupar uma posição de raiz que, quando categorizada, será associada a um conteúdo conceitual na Lista 3. Da mesma forma, tais elementos podem ocupar uma posição de núcleo funcional, atuando como uma peça gramatical da língua. A nossa ideia, portanto, é que *mane*, *feto*, *aman* e *inan* estão em um processo de gramaticalização, passando de raízes para núcleos funcionais.

Para contribuir com a formalização dos classificadores nominais, nos baseamos na visão de Wiltschko (2014), que distribui categorias funcionais em domínios sintáticos universais. Seguimos a modificação sugerida em Nóbrega (2018) na identificação da camada de *k*:classificação como *locus* das informações de gênero e classes nominais, e propomos que *mane*, *feto*, *aman* e *inan* são incompatíveis com essa camada. Mais especificamente, propomos que tais elementos funcionam como adjuntos sintáticos, sendo adicionados à estrutura como modificadores, sem atenderem a requerimentos morfosintáticos.

Uma vez que o estatuto e formalização dos classificadores nominais ainda é amplo tema de debate na literatura, acreditamos que esta pesquisa abre perspectivas interessantes para investigações futuras. Uma delas é a investigação a respeito das formações com classificadores que parecem licenciar uma leitura não composicional. Além disso, o comportamento de retomada anafórica desses formadores também precisa ser mais bem investigado considerando a proposta de que tais elementos são adjuntos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXIADOU, ARTEMIS. Roots don't take complements. *Theoretical Linguistics*, v. 40, n.3/4, p. 287-297, 2014.
- ACQUAVIVA, P. Roots and lexicality in Distributed Morphology. In: GALANI, Alexandra; REDINGER, Daniel; YEO, Norman (ed.). *York Papers in Linguistics 2*, York: University of York, p. 1-21, 2009.
- AIKHENVALD, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford Studies in Typology and Linguistic Theory.
- ARAD, MAYA. Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language & Linguistic Theory*, v. 21, p. 737-778, 2003.
- BASSANI, INDAIÁ DE SANTANA; MINUSSI, RAFAEL DIAS. Contra a seleção de argumentos pelas raízes: nominalizações e verbos complexos. *ReVEL*, v. 13, n. 24, p. 139-173, 2015.
- BELO, J. L.; OLIVEIRA, B. L. R. Processo de formação de palavras por composição do português e do tétum - Descrição e Confronto. In Filipe *et al.* Livro de Atas das V Jornadas Pedagógicas - Língua, Literatura e Cultura em (Trans) Formação: os desafios da atualidade. Centro de Língua Portuguesa da UNTL. Editora: Imprensa Nacional de Timor-Leste, I.P., p. 90-108, 2021.
- BISETTO, A.; SCALISE, S. The classification of compounds. *Lingue e Linguaggio*, v. 4, n. 2, p. 319-332, 2005.
- BORER, H. Exo-skeletal vs. Endo-skeletal Explanations: Syntactic Projections and the Lexicon. In: POLINSKY, M.; MOORE, J. (Eds.) *The Nature of Explanation*. Chicago: Chicago University Press (distributed by CSLI), 2003. p. 1-35.
- BORER, H. *Structuring Sense Volume III: Taking Form*. United Kingdom: Oxford University Press, 2013.
- CARNEIRO, A. S. R. As políticas linguísticas e de ensino de línguas em Timor Leste: desafios de um contexto multilíngue. *Filol. linguist. port.*, n. 12(1), p. 9-25, 2010.
- CARVALHO, D. O traço de gênero na morfossintaxe do português. *D.E.L.T.A.*, 34.2, p. 635-660, 2018. <http://dx.doi.org/10.1590/0102-445008104720040323>
- CARVALHO, M. J. A. Panorama linguístico de Timor. Identidade regional, nacional e pessoal. *Camões: Revista de Letras e Culturas Lusófonas*, 14, 65-89. 2001.
- CHOMSKY, N. Remarks on Nominalization. In: JACOBS, R. A.; P. ROSENBAUM

- (Eds). *Readings in English Transformational Grammar*, Waltham, Massachusetts: Ginn and Company, 1970.
- CHOMSKY, N. Minimalist inquiries: the framework. In: ROGER, Martin; MICHAELS, David; URIAGEREKA, Juan (org.). *Step by step: essays on minimalist syntax in honor of Howard Lasnik*. Cambridge, MA: The MIT Press, 2000, p. 89-155.
- CHOMSKY, N. Beyond explanatory adequacy. In A. Belletti. (Ed.) *Structure and Beyond* (pp. 104-131). Oxford: Oxford University Press, 2004.
- CHOUPINA, C. M. Reflexões sobre o género em português europeu e em tétum. Pág. 64-77, 2011. eLingUp [Centro de Linguística da Universidade do Porto] Volume 3, Número 1, 2011 ISSN 1647-4058.
- CORBETT, G.G. *Gender*. Cambridge University Press, Cambridge, 1991.
- CORREIA, A. J. G. Maklokek ba SOCIOLINGUÍSTIKA. Edisaun dahuluk. Instituto Nacional de Linguística - UNTL, Avenida Cidade de Lisboa - Díli, Timor-Leste, 2022.
- COSTA, L. *Língua Tétum – Contributos Para Uma Gramática*. Lisboa: Colibri. 2015.
- DE ALBUQUERQUE, D. B. e TAYLOR-LEECH, K. Política linguística para as línguas oficiais em Timor-Leste: o português e o Tétum-Praça. GRAGOATÁ - Revista dos programas de pós-graduação do Instituto de Letras da UFF, Niterói, n. 32, p. 153-169, 1. sem. 2012. DOI: <https://doi.org/10.22409/gragoata.v17i32.33037>
- DE ALBUQUERQUE, D. B. As línguas de Timor Leste: perspectivas. *Língua e Literatura*, n. 27, p. 313-335, 2003.
- Disionáriu Tetun-Portugés-Indonéziu: Ho Matadalan Gramátika nian. BUKA HATENE. Edisaun uluk. INL: Díli. 2005.
- EMBICK, D.; MARANTZ, A. Architecture and blocking. *Linguistic Inquiry*, Cambridge, MA, v. 39, n. 1, p. 1–53, 2008.
- EMBICK, D. *The morpheme: a theoretical introduction*. Berlin, München, Boston: De Gruyter Mouton, 2015.
- GONÇALVES, C. A. V.; ANDRADE, K. A instabilidade categorial dos constituintes morfológicos: evidência a favor do continuum composição-derivação. *DELTA*, São Paulo, 32 (2), 261-294, 2016.
- GREENBERG, J. H. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements, In: Joseph H. Greenberg (ed.). *Universals of Language*. London, 1963.
- GUEVARA, E.; SCALISE, S. *Searching for Universals in Compounding*. 2009. HAJEK, J. Language Contact and Convergence in East Timor: The Case of Tetun Dili. In:

- Aikhenvald and Dixon. *Grammars in Contact: A Cross-Linguistic Typology*. Research Centre for Linguistic Typology La Trobe University. Cap. 6. p. 163-178, 2017.
- GRINEVALD, C. (2015). *Classifier systems in the context of a typology of nominal classification*. Em: *Linguistic Typology*, p. 811-818.
- GRINEVALD, C. (2000). *SYSTEMS OF NOMINAL CLASSIFICATION: A morphosyntactic typology of classifiers*. Em: *Linguistic Typology*, p. 50-93.
- GRINEVALD, C. (2002). *Making Sense of Nominal Classification Systems*. Em: *Linguistic Typology*, p. 258-275.
- HAJEK, J.; KLINKEN, C. W. *Language Contact and Gender in Tetun Dili*. *Oceanic Linguistics*, v. 58, n. 1, p. 59-91, 2019.
- HALLE, M.; MARANTZ, A. *Distributed morphology and the Pieces of Inflection*. In: HALLE, K.; KEYSER, S. J. (eds.). *The view from building 20: essays in linguistics in honor of Sylvian Bromberger*. Cambridge, Mass.: MIT Press, p. 111-176, 1993.
- HALLE, M. *Distributed Morphology: impoverishment and fission*. In: BRUENING, Benjamin; KANG, Yoonjung; MCGINNIS, Martha (ed.). *MIT Working Papers in Linguistics: PF: papers at the Interface*, n. 30, p. 425–449, Cambridge, MA: The MIT Press, 1997.
- HARLEY, HEIDI. *On the identity of roots*. *Theoretical Linguistics: An Open Peer Review Journal*, v. 40, n. 3-4, p. 225-276, 2014.
- HULL, G. e ECCLES, L. *Gramática da Língua Tétum*. Edição portuguesa. Lisboa: LIDEL. 2005.
- KLINKEN, C. W. et al. *Tetun Dili: a grammar of an East Timorese language*. Canberra: Pacific Linguistics, 520, 2002.
- KRAMER, R. T. *The morphosyntax of gender*. Oxford University Press, 2015.
- KRAMER, R. *The location of gender in the syntax*. *Language and Linguistics Compass* 10, n.11: 661-677, 2016. <http://dx.doi.org/10.1111/lnc3.12226>.
- MARANTZ, A. *No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon*. In: DIMITRADIS, A.; SIEGEL, L. SUREK-CLARK, C.; WILLIAMS (eds.). *Proceedings of the 21st Penn Linguistics Colloquium*. Working Papers in Linguistics, Philadelphia, p. 201- 225, 1997.
- MARANTZ, A. *Locality domains for contextual allomorphy across the interfaces*. In: MATUSHANSKY, Ora; MARANTZ, Alec (org.). *Distributed Morphology today: morphemes for Morris Halle*. Cambridge, MA: The MIT Press, 2013, p. 95-115.
- Matadalan Ortográfiku ba Lia-Tetun*. INL: Díli. Instituto Nacional de Linguística 2003.

- MINUSSI, R. D.; BASSANI, I. S. Em favor do conteúdo semântico das raízes. *Revistas Letras*, v. 96, 2017.
- NÓBREGA, V. A., & PANAGIOTIDIS, P. *Headedness and Exocentric Compounding*. 2020.
- NÓBREGA, V. A. *No Escape from Categorization: An Insider's View of Compounds*. 2020.
- NÓBREGA, V. A. *Abordagem Isomórfica: A Articulação entre o Léxico e a Sintaxe na Emergência da Linguagem*. São Paulo: Universidade de São Paulo, 2018.
- NÓBREGA, V. A.; BASSANI, I. S.; ARMELIN, P. R. G. Flexão, Derivação e Composição em Morfologia Distribuída. In: *Manual de Morfologia Distribuída*. São Paulo: Editora da Abralín, 2022, pp. 222-264.
- PAULINO, V. Multilinguismo e pluralidade linguística em Timor- -Leste. *Todas as Letras – Revista de Língua e Literatura*, São Paulo, v. 25, n. 2, p. 1-17, maio/ago. 2023. DOI 10.5935/1980-6914/eLETD016136
- PFAU, ROLAND. *Grammar as process: A Distributed Morphology account of spontaneous speech errors*. Amsterdam: John Benjamins Publishing Company, 2009.
- RAMOS, R. e CARMO, J. A língua portuguesa entre as línguas de Timor-Leste: um estudo de caso numa escola secundária timorense. *Articles • Rev. bras. linguist. apl.* 20 (3) • Jul-Sep 2020 • <https://doi.org/10.1590/1984-6398202016719>.
- ROCHA, SONIA R. A ocorrência de “coisar” em Língua Portuguesa como contribuição para a hipótese das raízes abstratas. *Dissertação (Mestrado em Letras) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo*. São Paulo, 2008.
- RIO-TORTO, G. Composição. In Rio-Torto *et al.* *Gramática Derivacional do português*. Imprensa da Universidade de Coimbra, Cap. 7, pp.385-431, 2013.
- SCHER, A. P.; BASSANI, I. S.; ARMELIN, P. R. G. *Manual de Morfologia Distribuída*. Editora da Abralín. ISBN 978-85-68990-22-3, 2022.