

Rui Manuel Ferreira Pinto

# ANÁLISE DOS DESVIOS ORTOGRÁFICOS NA REPRESENTAÇÃO DAS SIBILANTES POR APRENDENTES DO ENSINO SUPERIOR TIMORENSE

Dissertação de Mestrado em Português como Língua Estrangeira e Língua Segunda, orientada pelas Professoras Doutoradas Cristina dos Santos Pereira Martins e Isabel Almeida Santos, apresentada à Faculdade de Letras da Universidade de Coimbra

2017

• U • C •



UNIVERSIDADE DE COIMBRA



Faculdade de Letras

ANÁLISE DOS DESVIOS  
ORTOGRÁFICOS NA REPRESENTAÇÃO  
DAS SIBILANTES POR  
APRENDENTES DO ENSINO  
SUPERIOR TIMORENSE

**Ficha Técnica:**

<b>Tipo de trabalho</b>	<b>Dissertação de Mestrado</b>
<b>Título</b>	<b>Análise dos desvios ortográficos na representação das sibilantes por aprendentes do Ensino Superior timorense</b>
<b>Autor/a</b>	<b>Rui Manuel Ferreira Pinto</b>
<b>Orientador/a</b>	<b>Cristina dos Santos Pereira Martins</b>
<b>Coorientador/a</b>	<b>Isabel Almeida Santos</b>
<b>Identificação do Curso</b>	<b>2º Ciclo em Português como Língua Estrangeira e Língua Segunda</b>
<b>Área científica</b>	<b>Língua e Literatura Materna</b>
<b>Especialidade/Ramo</b>	<b>Linguística Aplicada</b>
<b>Data</b>	<b>2017</b>

---

## Índice de Conteúdos

---

Índice de Tabelas .....	v
Índice de Gráficos.....	vi
Índice de Imagens.....	vii
Índice de Anexos.....	viii
Abreviaturas e Siglas.....	ix
Agradecimentos .....	xi
Resumo.....	xii
Abstract.....	xiii
Rezumu .....	xiv
<b>Introdução.....</b>	<b>2</b>
<b>1. Enquadramento teórico-descritivo.....</b>	<b>5</b>
1.1. Panorama linguístico de Timor-Leste .....	6
1.2. Sistemas de escrita: alguns aspetos.....	14
1.3. Descrição do sistema fonológico (as sibilantes) e a sua representação gráfica no português e no tétum .....	16
1.3.1. O sistema das sibilantes em português e em tétum.....	20
1.3.2. Relações fonema-grafema: o caso das sibilantes no português e no tétum .....	21
<b>2. Metodologia .....</b>	<b>25</b>
2.1. Instrumentos de recolha de dados .....	26
2.2. Perfil dos informantes.....	29
2.2.1. Apresentação dos dados do questionário .....	31
2.3. Procedimentos de tratamento dos dados .....	40
<b>3. Apresentação e discussão dos resultados.....</b>	<b>43</b>
3.1. Representação gráfica das consoantes sibilantes no <i>corpus</i> recolhido: ocorrências convergentes e divergentes.....	44
3.1.1. Análise dos dados por município .....	47
3.1.2. Análise dos dados por LM.....	48
3.2. Distribuição dos desvios por possibilidade descritiva/interpretativa .....	48
3.3. Distribuição dos desvios por posição silábica e por segmentos afetados .....	50
3.3.1. Ataque .....	51
3.3.1.1. Desvios na representação gráfica do segmento fonológico /s/ .....	52

3.3.1.2. Desvios na representação gráfica do segmento fonológico /z/ .....	55
3.3.1.3. Desvios na representação gráfica do segmento fonológico /ʃ/ .....	57
3.3.1.4. Desvios na representação gráfica do segmento fonológico /ʒ/ .....	58
3.3.2. Coda .....	59
3.3.3. Sequências .....	61
<b>Conclusão .....</b>	<b>63</b>
<b>Referências.....</b>	<b>66</b>
<b>Anexos .....</b>	<b>71</b>

---

## Índice de Tabelas

---

- Tabela 1: Línguas de Timor-Leste - Fonte: Hull (2002)
- Tabela 2: Alfabetos do PE e do tétum
- Tabela 3: Matriz fonológica das consoantes sibilantes do português e do tétum
- Tabela 4: Representação gráfica das sibilantes do sistema ortográfico do PE e do tétum
- Tabela 5: Valores fonético-fonológicos dos grafemas que representam as sibilantes do PE e do tétum
- Tabela 6: Instituições acreditadas pela ANAAA, com o respetivo n.º de alunos matriculados em 2015
- Tabela 7: Distribuição dos informantes por município de origem
- Tabela 8: Síntese dos dados do *corpus* por município
- Tabela 9: Síntese dos dados do *corpus* por LM
- Tabela 10: Distribuição dos desvios por possibilidade descritiva/interpretativa
- Tabela 11: Distribuição das ocorrências convergentes e divergentes em posição de ataque silábico.
- Tabela 12: Distribuição das representações gráficas desviantes de [s]
- Tabela 13: Distribuição das representações gráficas desviantes de [z]
- Tabela 14: Distribuição das representações gráficas desviantes de [ʃ]
- Tabela 15: Distribuição das representações gráficas desviantes de [ʒ]
- Tabela 16: Distribuição das ocorrências convergentes e divergentes em posição de coda silábica
- Tabela 17: Distribuição das representações gráficas desviantes de [ʃ], [ʒ] e [z], em coda silábica
- Tabela 18: Distribuição das ocorrências convergentes e divergentes de /s.s/
- Tabela 19: Distribuição das representações gráficas desviantes de [ʃs]

---

## Índice de Gráficos

---

- Gráfico 1: Distribuição das produções com e sem desvios na representação das sibilantes
- Gráfico 2: Percentagem de produções com e sem desvios por distrito de origem
- Gráfico 3: Comparação da percentagem de produções com e sem desvios na representação gráfica das sibilantes, por LM
- Gráfico 4: Distribuição da totalidade das ocorrências - convergentes e divergentes
- Gráfico 5: Distribuição dos desvios por posição silábica
- Gráfico 6: Distribuição dos desvios em função do segmento alvo
- Gráfico 7: Distribuição dos desvios em posição de ataque silábico em função da unidade alvo

---

## Índice de Imagens

---

Imagem 1: Distribuição geográfica das principais línguas de Timor-Leste

Imagem 2: Tratamento, em Excel, dos dados recolhidos (convergentes e divergentes)

Imagem 3: Tratamento, em Excel, das ocorrências desviantes



---

## Índice de Anexos

---

Anexo 1: Questionário e exercício de produção escrita

Anexo 2: Orientações para os professores aplicadores

Anexo 3: Autorização dos informantes

Anexo 4: Apresentação dos dados do questionário

Anexo 5: Principais línguas de Timor-Leste

Anexo 6: Distribuição das ocorrências convergentes e divergentes

Anexo 7: Tratamento dos dados em *Excel*

Anexo 8: Ocorrências desviantes excluídas da análise

---

## Abreviaturas e Siglas

---

ANAAA	Agência Nacional para a Avaliação e Acreditação Académica
AFI	Alfabético Fónico Internacional
CPLP	Comunidade dos Países de Língua Portuguesa
FRETILIN	Frente Revolucionária de Timor-Leste Independente
EB	Ensino Básico
ESG	Ensino Secundário Geral
ESUP	Ensino Superior
ICFP	Instituto Católico de Formação de Professores
INFORDEPE	Instituto Nacional de Formação de Docentes e Profissionais da Educação
LA	Língua Alvo
LE	Língua de Ensino
LI	Língua de Instrução
LN	Língua Nacional
LO	Língua Oficial
LP	Língua Portuguesa
LS	Língua Segunda
PCLP	Projeto de Consolidação da Língua Portuguesa
PE	Português Europeu
PFICP	Projeto de Formação Inicial e Contínua de Professores

PRLP	Projeto de Reintrodução da Língua Portuguesa
PLNM	Português Língua Não Materna
QECRL	Quadro Europeu Comum de Referência para as Línguas
RDTL	República Democrática de Timor-Leste
RTPI	Rádio Televisão Portuguesa Internacional
RTTL	Rádio Televisão de Timor-Leste
UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>

---

## Agradecimentos

---

À Professora Doutora Cristina Martins e à Professora Doutora Isabel Santos, pela orientação científica, pelas palavras de estímulo e pela disponibilidade absoluta, em todos os momentos.

Aos alunos da Licenciatura em Formação de Professores do INFORDEPE, pelo valioso contributo que viabilizou esta dissertação.

Aos professores do Projeto de Formação Inicial e Contínua de Professores, pelo apoio na aplicação dos questionários e dos exercícios de produção escrita.

À Ana Caetano, pela paciência, pelo apoio incondicional e constantes incentivos nesta caminhada.

À Ana Remelgado, por partilhar parte desta caminhada.

A todos os que, de alguma forma, contribuíram para que esta investigação se tornasse possível.

À Cristina e à Miriam, pelo amor, companheirismo, pela força e otimismo nos momentos mais difíceis, fundamentais para a realização desta dissertação.

---

## Resumo

---

### **Análise dos desvios ortográficos na representação das sibilantes por aprendentes do Ensino Superior timorense**

Tendo por base o contexto multilinguístico e de contacto de línguas em Timor-Leste, onde o português assume um papel de língua cooficial e de ensino, a par do tétum, mas se afigura, ainda assim, como língua não materna para a maioria da população, este estudo pretende proceder à análise das dificuldades sentidas por aprendentes do Ensino Superior timorense na representação gráfica das sibilantes.

Além disso, este trabalho pretende, ainda, indicar as possibilidades descritivas e interpretativas que poderão estar na base dos desvios apresentados, tendo-se para esse efeito considerado também as diferentes posições silábicas em que as consoantes sibilantes ocorrem.

Partindo da descrição das relações entre fonemas e grafemas sibilantes quer na língua alvo, quer no tétum, procede-se, então, à análise dos desvios recolhidos de um conjunto de textos produzidos por um grupo de estudantes do Ensino Superior timorense que frequentam a Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do Ensino Básico no INFORDEPE.

**Palavras-chave:** português língua não materna; ortografia; sibilantes; fonologia; português de Timor-Leste

---

## Abstract

---

### **Analysis of errors in the orthographic representation of sibilants by Timorese Higher Education Students**

Bearing in mind language contact and the multilingual setting in East Timor, where Portuguese, a non native language of the majority of the population, is one of the co-official languages of the country, alongside with Tetun, the present study aims at analysing the difficulties encountered by Timorese Higher Education students regarding the orthographic representations of sibilants.

This study also aims at identifying descriptive and interpretative possibilities that might explain the orthographic errors presented by learners, considering, to this effect, the different syllabic positions in which the sibilants occur.

Starting with the description of the relationships between sibilant phonemes and graphemes both in the target language and in Tetun, we then analyse the errors collected from a corpus of texts written by a group of Higher Education students who attend the Degree in Teacher Education of the 1<sup>st</sup> and 2<sup>nd</sup> Cycles of Basic Education at the INFORDEPE.

**Keywords:** Portuguese as a non native language; orthography; sibilants; phonology; East Timor portuguese

---

## Rezumu

---

### **Análize kona-ba dezviu ortográfiku sira iha representasaun son sibilante sira ba timoroan sira mak estuda iha universidade**

Iha ninian abut/baze kontestu dalen barabarak no kontaktu dalen oioin iha Timor-Leste, ne'ebé dalen portugéz iha ninian knaar hanesan dalen koofisiál iha lia-hanorin hamutuk ho tetun, maske dalen ida ne'e, la'os lia-inan ba ema barak, maibé, estudu ida ne'e hakarak halo análise ida kona-ba difikuldade prátika, ne'ebé estudante sira ne'e sente iha sira nian estudu iha Ensinu Superior.

Alende ne'e, traballu ida ne'e hakarak buka hatudu possibilidade interpretativa sira ne'ebé bele sai hanesan abut/baze ba dezviu hosi prátika son sira ne'e,, ho efeitu ida ne'e, ita Konsidera mós pozisaun silábica sira mak la hanesan ne'ebé Konsoante sibilante sira mosu.

Hahú ho deskrisaun kona-ba relasaun entre fonema no grafema ba representaun son sira ne'e, nune'e iha dalen ne'ebé temi, hanesan mós iha tetun, nune'e, lori ba halo análise ida kona-ba tamba saída mak hetan dezviu ka hasees-an son sira ne'e, ne'ebé rekolle ka tau hamutuk hosi testu sira mak produs hosi grupu estudante universitáriu timoroan sira mak frekuenta dadaun kursu lisensiatuura iha Kursu Formasaun ba manorin sira iha 1<sup>o</sup> no 2<sup>o</sup> Siklu Ensinu Báziku iha INFORDEPE.

**Liafuan xave sira:** Portugéz la'os lia-inan; ortografia; sibilante sira; fonolojia; portugéz husi Timor-Leste

*“Mesmo que o currículo comece com o português, nos bancos da escola,  
nós temos segmentos em que uns falam português,  
outros inglês, outros tétum, outros bahasa.*

*Esse é o nosso problema.*

*Há uma geração toda em Timor que  
nasceu no tempo indonésio e só sabe bahasa.*

*Como resolver esse problema?”*

**Xanana Gusmão, 2007**



---

## Introdução

---

A presente dissertação surge enquadrada no contexto poliglóstico de Timor-Leste e pretende constituir-se como uma reflexão sobre os desvios ortográficos na representação das sibilantes por aprendentes do Ensino Superior timorense.

Quem chega a Timor-Leste encontra um país onde várias línguas estão em contacto e onde o português influencia e é influenciado por várias outras línguas. Nesta amostra de informantes – 279 alunos do Curso de Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do Ensino Básico –, o português, consagrado como língua oficial (LO) e língua de instrução (LI) é aprendido como língua segunda (LS). O português surge, assim, para os informantes em causa, como objeto de estudo e, paralelamente, como veículo de transmissão de conhecimentos das outras unidades curriculares que compõem o Curso de Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do Ensino Básico.

Considerando a complexidade das relações entre unidades fónicas e gráficas, que inevitavelmente se refletem, também, no desempenho ortográfico de aprendentes cuja língua materna (LM) é o português, procura-se verificar quais as principais dificuldades evidenciadas na representação gráfica das sibilantes, por estes aprendentes de português língua não materna (PLNM). A escolha recaiu sobre o tratamento das sibilantes por esta ser uma área do sistema consonântico que revela uma grande complexidade e opacidade, pois não existe sempre, entre fonemas e grafemas, uma relação de biunivocidade, ou seja, de “um para um”. Esta complexidade, bem visível nos desvios que se vão encontrando nos textos produzidos por alunos e professores timorenses, despertou uma vontade crescente de estudar esta questão. Além disso, as conclusões retiradas no trabalho final da unidade curricular de Língua Portuguesa - Fonologia<sup>1</sup> demonstraram que a representação gráfica das sibilantes foi a que causou maiores dificuldades aos estudantes timorenses.

---

<sup>1</sup> Unidade curricular relativa ao primeiro ano do Curso de Mestrado em Língua e Linguística Portuguesa, na Universidade Nacional de Timor Lorosa'e, ministrado em parceria com a Universidade de Coimbra.

Definida a problemática de investigação, e tendo em vista traçar padrões relevantes de representação gráfica desviante das sibilantes, bem como identificar os fatores (linguísticos e extralinguísticos) que mais os promovem, delinear-se-ão como objetivos deste estudo:

- (i) identificar a sibilante cuja representação gráfica se afigura mais problemática;
- (ii) identificar a posição silábica em que ocorrem mais desvios;
- (iii) apontar o grupo de aprendentes, em função da LM, que apresenta o comportamento mais desviante;
- (iv) identificar o grupo de aprendentes, em função do município de origem, que apresenta maiores dificuldades no registo ortográfico das quatro sibilantes do português;
- (v) interpretar/descrever as razões que poderão estar na base dos desvios apresentados.

Por forma a dar cumprimento a estes objetivos, o presente estudo foi estruturado em 3 capítulos: o primeiro foi concebido com o intuito de: i) contextualizar o papel que a língua portuguesa (LP) desempenha neste espaço multilingue (1.1); ii) apresentar algumas reflexões sobre a escrita e a oralidade com o objetivo de salientar a complexidade existente na relação entre estes dois sistemas (1.2) e, por último, iii) descrever os aspetos relevantes nas relações fonema-grafema da LP e do tétum, nomeadamente no que concerne aos fonemas sibilantes em análise /s, z, ʃ, ʒ/ (1.3).

O segundo capítulo deste estudo é dedicado à descrição e justificação da metodologia adotada, apresentando-se o perfil sociolinguístico dos informantes e os procedimentos de recolha do *corpus* onde são contabilizadas, não só as ocorrências desviantes, mas também as convergentes.

No terceiro e último capítulo deste trabalho, são apresentados e analisados os resultados do presente estudo. Assim, analisam-se os desvios na representação gráfica das sibilantes tendo em consideração, por um lado, a posição silábica e o segmento afetado e, por outro, o município de origem e a LM do informante. Discutem-se, também, a

expressividade e a natureza dos desvios procurando descrever/interpretar as razões que poderão estar na base dos mesmos. Para esse efeito, são considerados fatores como: a “opacidade” da ortografia do português; o desconhecimento das regras contextuais de uso dos grafemas; e o desconhecimento da identidade do segmento fonológico.

Finalmente, considerando os resultados obtidos, apresentam-se as principais conclusões deste estudo, na tentativa de esboçar um perfil dos aprendentes do Ensino Superior timorense, no que respeita à representação grafemática das sibilantes do português.

---

# **1. Enquadramento teórico-descritivo**

---

### 1.1. Panorama linguístico de Timor-Leste

Apesar de corresponder a um pequeno território, Timor-Leste<sup>2</sup>, à semelhança de muitos países do sudeste asiático<sup>3</sup>, possui uma enorme e complexa diversidade linguística fruto da sua história, marcada por invasões e domínio por parte de outros países. Na verdade, Timor-Leste foi uma colónia portuguesa entre o século XVI e a segunda metade do século XX; sofreu a ocupação do Japão durante 3 anos, no período que antecedeu a Segunda Guerra Mundial, foi invadido pela Austrália, em 1942, e ocupado pela Indonésia, em 1975, até à chegada das Forças de Paz das Nações Unidas, em 1999.

Durante a ocupação portuguesa, a situação linguística envolvia três dimensões: (i) a das línguas locais<sup>4</sup>, veículos de comunicação nas diversas localidades; (ii) a da língua veicular – o tétum, funcionando como elemento de integração e conhecido como tétum praça, variante do tétum terik, gramaticalmente simplificado e mesclado com elementos do português<sup>5</sup>; e (iii) a da língua administrativa – o português, a única língua normalmente escrita, que também exercia uma função integradora entre a camada dirigente e o ambiente

---

<sup>2</sup> A ilha de Timor tem cerca de 480 km de comprimento, 100 km de largura e uma área de cerca de 19.000 km<sup>2</sup>. A República Democrática de Timor-Leste (RDTL) é constituída pela parte oriental da ilha (correspondendo a cerca de dois terços), pelo enclave de Oe-Cusse Ambeno (na costa norte da parte ocidental), pela ilha de Ataúro (a 23 km a norte da capital, Díli) e, ainda, pelo ilhéu de Jaco (na ponta leste, no distrito de Lautém) (cf. Brito, 2010: 179).

<sup>3</sup> Caetano (2014: 29-30), com base em dados da UNESCO, numa tentativa de desmistificar a ideia de que Timor-Leste é um caso único de diversidade linguística, refere, por exemplo, que a Indonésia possui cerca de 700 línguas e que apenas 10% da população tem o indonésio como LM. Afirma igualmente que existem cerca de 140 línguas diferentes na Malásia, sendo a língua malaia usada na maioria das escolas. No entanto, na Malásia, também o mandarim e outras línguas indígenas (tâmul, telugu, panjabi, tai, entre outras) são usadas como LI em algumas escolas primárias.

<sup>4</sup> No caso das designações das LM dos habitantes de Timor-Leste (também designadas na literatura por línguas locais ou vernaculares), e dado não existirem, para cada uma, grafias estabilizadas, optou-se, neste trabalho, pela grafia adotada por Simons & Fennig, tal como se observa na página de internet <https://www.ethnologue.com/country/TL> (consultado a 12 de novembro de 2016). Esta página aponta para a existência de 19 línguas autóctones em Timor-Leste (adabe; baikeno; bunak; fataluku; galolen; habun; idaté; kairui-midiki; kemak; lakalei; makalero; makasae; makuva; mambae; nauete; tetun; tetun Dili; tukude; waima'a) e de um *pidgin* já extinto.

<sup>5</sup> Hull & Eccles (2005: viii) referem, por exemplo, que a fonte mais comum de empréstimos para designar novas realidades é o português e que os estudantes de tétum com conhecimento de português gozam de uma vantagem considerável porque grande parte do vocabulário e numerosos morfemas e construções do tétum são de origem portuguesa.

letrado (cf. Thomaz, 2002: 140-4).

Durante o período da invasão Indonésia, e ainda que a língua indonésia tenha sido declarada língua nacional (LN) e LO, Timor-Leste manteve a vontade e a determinação em preservar a sua riqueza linguística e cultural.

Depois da independência do país, em 2002, a Constituição passou a consagrar, no artigo 13.º, o tétum e o português como línguas oficiais. Além disso, o artigo 159.º determina que as línguas indonésia e inglesa figurem como línguas de trabalho, em uso na administração pública, a par das línguas oficiais, enquanto tal se mostrar necessário. É de salientar, no entanto, que, no presente, nenhuma das quatro línguas supracitadas é falada por toda a população. Numerosas outras línguas e dialetos são utilizados localmente, nomeadamente nas relações familiares e, inclusive, em contexto escolar.

Sobre o número de línguas locais não existe, na literatura, consenso, pelas razões apresentadas por Esperança (2001: 98):

*“Timor tem sido descrito frequentemente como uma Babel, devido à sua diversidade linguística. O número de línguas e dialetos varia conforme os autores, principalmente pelos critérios (ou a falta deles) que usam para fazer a distinção entre uns e outros.”*

Carvalho (2001: 65), por exemplo, considera que existem 18 LL em Timor:

*“A jovem República situa-se numa ilha dividida em 18 línguas nacionais [segundo] a seguinte classificação provisória: (i) um grupo A, integrado no ‘continuum’ de Roti a Wetar, no que corresponde à parte ocidental, compõe-se do Dawan, com o seu dialecto Baiqueno; no sector central, da ilha, acrescenta-se o Tétum, com os seus dialectos Térik, Belu, Bekais, Praça ou Díli e o Habu; a norte inclui-se o Raklungu ao lado do Rasuk e do Raklungy, assim como o Galoli, muito aparentado com certos dialectos de Wetar; e, para finalizar, na região oriental apresentam-se o Kairui, o Waimata, o Midiki e o dialecto Nauete;(ii) um grupo B compõe-se das seguintes regiões: ocidental, com o Kemak (e o seu dialecto Nogo), o Tokodede (e o seu*

*dialecto Keta); central, com o Mambae (e o seu dialecto Lolein) e oriental, com o Idaté e o Lakalei. Há ainda cinco línguas – Bunak, com o dialecto Maraé, Makasai, Makalero, Fataluku e Lovaia, com o dialecto Maku’a – que, não constituindo um grupo, partilham características com A e com B.”*

Hull<sup>6</sup> (2002: 9), linguista australiano e autor de alguns estudos sobre as origens das línguas da ilha de Timor-Leste, elencava, em 2002, 16 línguas diferentes, tendo em consideração a sua origem.

LÍNGUAS DE TIMOR-LESTE	LÍNGUAS PAPUAS		Bunak	
			Fataluku	
			Makasae	
			Makalero	
	LÍNGUAS AUSTRONÉSICAS		LÍNGUAS FABRÓNICAS	Tétum
				Habun
				Kawaimina (Kairui, Waima’a, Nauete)
				Idalaka (Lakalei, Idaté, Isni)
			LÍNGUAS RAMELAICAS	Galolen
				Wetarese (Rahesuk, Resuk, Rklungu)
				Bekais
				Dawan (Baikeno)
	LÍNGUAS AUSTRONÉSICAS		LÍNGUAS RAMELAICAS	Mambae
				Kemak
				Tokodede
				Lovaia (Maku'a)

Tabela 1: Línguas de Timor-Leste - Fonte: Hull (2002)

Por seu lado, Thomaz, historiador e especialista em história do sudeste asiático, aponta para a existência de “19 a 31 línguas locais, segundo as contagens – provindo a discrepância do critério dotado na destriça entre as línguas autónomas e variantes dialetais da mesma língua” (2002: 141).

Em termos territoriais, verifica-se que, à exceção do tétum<sup>7</sup>, que se encontra

<sup>6</sup> Além de defender o estatuto do português como LO em Timor-Leste, Geoffrey Hull é um dos maiores especialistas do tétum e das línguas nativas presentes no país.

difundido numa área mais vasta, embora descontínua, as línguas de Timor possuem uma expressão geográfica bem definida na metade oriental da ilha (cf. Mapa 1).

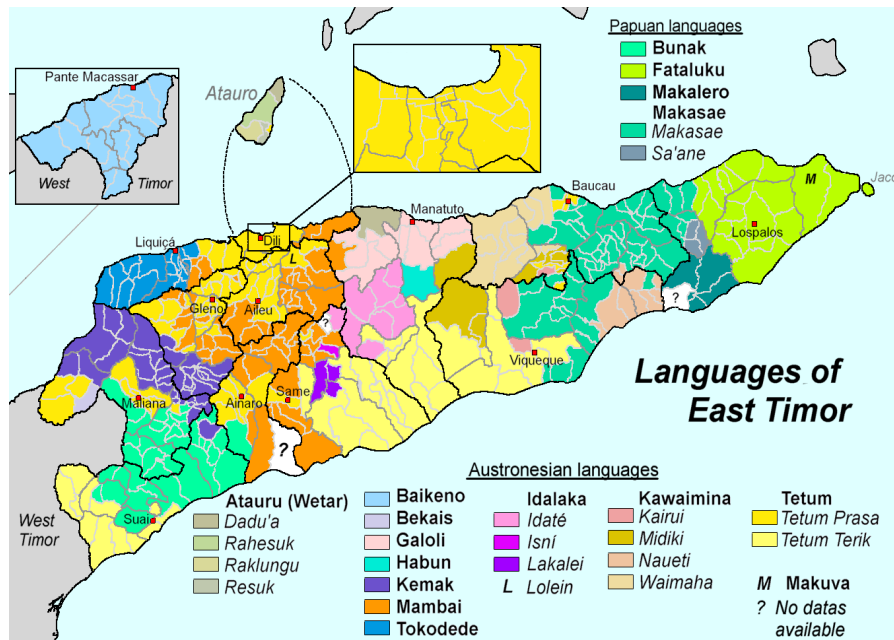


Imagem 1: Distribuição geográfica das principais línguas de Timor-Leste<sup>8</sup>

Verifica-se alguma heterogeneidade junto à fronteira com a Indonésia, com o tétum<sup>9</sup> a predominar na zona de Covalima, o kemak, idioma mais usado na parte a sul de Bobonaro e, a norte, o bunak como língua mais utilizada. Verifica-se também que o tétum praça assume um papel de destaque em Díli, o tukodede é língua dominante no município de

<sup>7</sup> O artigo 2.º do Decreto do Governo n.º 1/2004 da RDTL clarifica que “a variedade do tétum afirmada como língua oficial e nacional é o tétum oficial, uma forma literária moderna do vernáculo mais comum no país, baseado no Tétum-Praça”, sendo a ortografia adotada definida e desenvolvida pelo Instituto Nacional de Linguística (INL). O tétum assume-se, portanto, como LM para alguns falantes e como língua veicular na globalidade do território.

<sup>8</sup> Fonte: [http://www.wikiwand.com/pt/L%C3%ADnguas\\_de\\_Timor-Leste](http://www.wikiwand.com/pt/L%C3%ADnguas_de_Timor-Leste) (consultado a 18 de novembro de 2016).

<sup>9</sup> Acerca das variedades históricas do tétum, Hull & Eccles (2005) salientam que existem, até ao momento, seis variedades principais em Timor-Leste, com muita variação interna: (i) a **variedade acroletal do tétum praça de Díli, falada pelos habitantes de Díli** também fluentes em português; (ii) a **variedade acroletal do tétum praça de Díli, falada pelos habitantes fora de Díli** também fluentes em português; (iii) a **variedade mesoletal do tétum praça**, falada em todo o país pelas pessoas instruídas não fluentes em português; (iv) **as variedades basiletais do tétum praça** utilizadas por pessoas iletradas que não sabem falar português e, muitas vezes, não completamente fluentes em tétum; (v) **os dialetos tétum térrik** falados nas zonas de Balibó, Suai, Alas, Lacluta e Viqueque e (iv) **o tétum litúrgico**, como meio de comunicação exclusivamente escrito cultivado pelos escritores católicos e que, basicamente, tem por base o tétum praça, mas recorrendo ao fonetismo, morfologia e léxico do tétum térrik.



Liquiçá e, no interior montanhoso, na zona de Aileu, Ainaro e Manufahi, o mambae afirma-se como a língua principal. Na zona central do país (município de Manatuto) dominam o galolen, o habun e o tétum; na parte leste da ilha, o makasae, nos municípios de Baucau e Viqueque, e o fataluku, na zona de Lautém. Por fim, no enclave de Oe-cusse<sup>10</sup> e na ilha de Ataúro, as línguas privilegiadas para a comunicação são o baikeno e o wetar, respetivamente.

É importante salientar que algumas destas línguas se destacam em termos de percentagem do número de falantes (cf. anexo 5). Segundo os dados dos Censos de 2015, o tétum praça é LM de 361.027 cidadãos timorenses (30,6% da população)<sup>11</sup>, o mambae, LM de 195.778 (16,6%) e o makasae, LM de 123.840 habitantes (10,5%). Com valores acima dos 5% encontramos, ainda, o tétum terik (6,05%), o baikeno (5,87%), o kemak (5,85%) e o bunak (5,48%)<sup>12</sup>.

Ora, como será facilmente perceptível, esta complexidade linguística comporta dificuldades acrescidas em termos de comunicação, conforme atesta Costa (2005: 614-615):

*“O estrangeiro que queira percorrer o território e tentar comunicar com os seus habitantes vê-se perante uma babélica imagem de Timor e tem de recorrer a um intérprete, já que não pode estabelecer um verdadeiro entendimento entre indivíduos de grupos diferentes, cada um falando a sua língua materna. Assim, há que recorrer ao tétum, língua comum à maioria dos grupos.”*

A diversidade etnolinguística que caracteriza o espaço social timorense e a diferente funcionalidade de cada idioma resultam numa enorme complexidade, fazendo com que se tenha sempre de considerar os diferentes papéis sociais das línguas (português, tétum e

---

<sup>10</sup> Para os propósitos deste trabalho, e no sentido de simplificar o tratamento dos dados, a Zona Económica Especial de Mercado Social de Oe-cusse será contabilizada como sendo um dos municípios de Timor-Leste.

<sup>11</sup> O estatuto de língua com maior percentagem de falantes em Timor-Leste assumido pelo tétum é compreensível, uma vez que esse idioma beneficia do estatuto de LO e de ensino (LE), sendo difundido através do sistema de ensino e dos meios de comunicação social do país.

<sup>12</sup> Fonte: dados apresentados em *Timor-Leste Population and Housing Census 2015*, disponível em <http://www.statistics.gov.tl/category/publications/census-publications> (consultado a 18 de novembro de 2016).

línguas estrangeiras, especialmente o inglês e o indonésio) nas variadas situações quotidianas de interação comunicativa.

É neste mosaico linguístico<sup>13</sup>, caracterizado por uma situação de bi- e, mesmo, plurilinguismo social com diglossia<sup>14</sup>, que a língua portuguesa tenta assumir um lugar de destaque. Na sua alocução<sup>15</sup> proferida em Brasília, no dia 1 de agosto de 2002, durante a IV Conferência de Chefes de Estado e de Governo da Comunidade dos Países de Língua Portuguesa (CPLP), o Presidente da República de então, Kay Rala Xanana Gusmão, referiu que:

*“A opção política de natureza estratégica que Timor-Leste concretizou com a consagração constitucional do Português como língua oficial, a par da língua nacional, o Tétum, reflecte a afirmação da nossa identidade pela diferença que se impôs ao mundo e, em particular, na nossa região onde, deve-se dizer, existem também similares e vínculos de carácter étnico e cultural, com os vizinhos mais próximos. Manter esta identidade é vital para consolidar a soberania nacional.”*

Além desta óbvia alusão à afirmação da identidade do país, consolidando, assim, a soberania nacional, a escolha do português e do tétum reflete também a ligação secular existente entre as duas línguas. Esta ligação terá estado na base da elevação do estatuto do tétum de língua indígena a língua franca e nacional<sup>16</sup>, tornando, assim, a escolha do

---

<sup>13</sup> Numa alusão à combinação de cores que se pode encontrar nos panos tradicionais de Timor, Baltazar (2013: 7) identifica este panorama linguístico como sendo um *“tais linguístico”*.

<sup>14</sup> Esse fenómeno *“traduz uma distribuição funcional e complementar de idiomas por domínios de uso, permitindo, assim, que os falantes venham a estabelecer uma associação convencional entre determinados usos linguísticos e situações de interação particulares”*. (Martins, 1994: 99)

<sup>15</sup> Disponível em [www.cplp.org/noticias/ccegc/di7.htm](http://www.cplp.org/noticias/ccegc/di7.htm) (consultado a 16 de novembro de 2016).

<sup>16</sup> A respeito da valorização e desenvolvimento do tétum, Batoréo (2009: 4), citando Moura (2007: 2), aponta para as palavras do antigo presidente de Timor-Leste, José Ramos-Horta: *“A língua portuguesa é fundamental para a nossa identidade. O próprio tétum, para se desenvolver, precisa do português. Alimenta-se dele.”* Também a esse respeito, Corte-Real & Brito (2006: 129) referem que se *“[a]ssiste (...) a uma naturalidade na escolha do português, pela parceria secular com o tétum – que lhe valeu a elevação estatutária – que resulta numa interpenetração mútua entre as duas línguas, em que se tipifica o português falado por timorenses e em que o tétum absorve do português influências nos níveis fonológico, morfológico, sintático-semântico e*

português como língua cooficial perfeitamente natural, embora, para a grande maioria da população, este idioma figure como uma terceira ou quarta língua no seu repertório linguístico. Além de refletir esta ligação secular e a necessidade natural de suportar o tétum no seu desenvolvimento contínuo (o tétum continua a assimilar neologismos, designadamente do português) outros aspetos terão estado, certamente, na origem da escolha do português como LO. De entre estes, podem destacar-se: (i) o facto de a língua portuguesa ter sido a língua da resistência<sup>17</sup>; (ii) fatores político-estratégicos, que se prendem com a afirmação da identidade nacional face aos países da região<sup>18</sup>; (iii) o valor simbólico e afetivo<sup>19</sup> da LP; e (iv) as semelhanças entre o português e o tétum<sup>20</sup>.

Se considerarmos os diferentes domínios de uso na comunidade, o português é a língua privilegiada para o acesso ao conhecimento, a língua da literacia e da instrução formal. No entanto, e apesar da obrigatoriedade do seu uso nas escolas, muitos professores recorrem ao tétum ou ao indonésio na lecionação das suas aulas, pois possuem uma proficiência muito elementar na língua portuguesa<sup>21</sup>. Além do contexto escolar, há também

---

*pragmático. Se tentássemos eliminar todos os termos de origem portuguesa numa conversa dentro do contexto dos preparativos para a realização de uma actividade sociocultural – por exemplo, o casamento católico timorense – diríamos que os noivos não chegariam a casar-se, pois, o casamento tornar-se-ia inviável.*“

<sup>17</sup> Como é sobejamente conhecido, “o Português sobreviveu como língua de resistência, sendo utilizada pela FRETILIN e pelas outras organizações da oposição ao ocupante nas suas comunicações internas e no contacto para o exterior.” (Batoréo, 2010: 59).

<sup>18</sup> Batoréo (2010: 59) cita, a este propósito, as palavras de um político timorense, Francisco Guterres, presidente da Frente Revolucionária de Timor-Leste Independente (FRETILIN), referindo que se trata de “uma opção política de natureza estratégica para afirmação da nossa identidade pela diferença que se impôs ao mundo.”

<sup>19</sup> Batoréo (2010: 60) refere que “o português goza de uma forte carga simbólica e de uma grande afetividade, rara noutros países em relação a uma antiga língua colonial.”

<sup>20</sup> Segundo Hull (2001: 42-43), “[o] português tem também a vantagem de que o tétum (língua franca) não seja formalmente muito afastado do português na sua pronúncia, gramática e vocabulário. O português não é um idioma demasiado difícil para os timorenses pois estes já possuem um relativo conhecimento passivo do português devido ao facto de que já falam o tétum-Díli. O mesmo já não se poderia dizer da relação entre o tétum e o inglês.”

<sup>21</sup> No sentido de mapear o estado geral das competências comunicativas dos cerca de quinze mil professores que compõem o quadro docente timorense, o Instituto Nacional de Formação de Docentes e Profissionais da Educação (INFORDEPE) realizou, em finais de 2015, um teste diagnóstico de proficiência em língua portuguesa a todos os professores do território. Apesar de ainda não terem sido divulgados oficialmente, os resultados apontam para cerca de seis mil professores com competência linguística de nível A1 no Quadro Europeu Comum de Referência para as Línguas (QERCRL).

uma tentativa de usar o português em algumas instituições do estado, nomeadamente na administração pública (ministérios e tribunais, por exemplo), mas apenas quando a abordagem se faz com interlocutores nativos do português.

A este respeito, Albuquerque (2011: 69) refere que:

*“a língua portuguesa em Timor-Leste está limitada somente a poucas situações formais e é utilizada apenas por uma minoria sociolinguística: a população mais idosa, pelo fato de terem nascido [sic] antes da invasão indonésia em 1974 e, conseqüentemente, terem frequentado [sic] o sistema educacional português.”*

Em relação à evolução da língua portuguesa em Timor-Leste, Brito & Martins (2005: 646), entre outros<sup>22</sup>, defendem que, tal como nos outros países da CPLP, esta língua está a construir a sua própria história, potenciando, no futuro, a existência de uma variedade própria, com as suas especificidades:

*“Em Timor-Leste, como em Angola, Brasil, Cabo Verde, Guiné-Bissau, Moçambique, Portugal ou São Tomé e Príncipe, a língua portuguesa conhece e constrói a sua própria história – e, por isso, está muito longe de poder ser tratada como um idioma uniforme. Devemos encarar o desafio da Língua Portuguesa nesta perspectiva, com a certeza de que, seja em que contexto lusófono for, estaremos diante de mais uma variedade do português. A nossa tarefa será então a de procurar descrever a língua portuguesa nos seus contextos específicos e entender as idiossincrasias que a caracterizam, respeitando-lhe as experiências particulares, os valores diferentes, a especificidade cultural e a sua peculiar visão do mundo. Em cada variedade do português exprime-se uma comunidade que, se por um lado constrói e define a sua própria identidade, por outro lado, é parte do imenso mosaico que constitui o sistema linguístico português.”*

---

<sup>22</sup> Cf., por exemplo, Albuquerque (2011: 66): “esta variedade foge à norma europeia, ou seja, há estruturas típicas do Português de Timor-Leste.”

Em suma, apesar de ser uma das LO do país, o português assume o estatuto de LM de um número muito reduzido de elementos da população<sup>23</sup>. Na realidade, é utilizado - quer como LM quer como língua não materna (LNM) - por uma pequena fração da população e, dependendo das perspetivas, a percentagem de falantes varia entre os 5% e os 37%<sup>24</sup>. O português assume, assim, o estatuto de língua segunda (LS) para a maioria dos timorenses que o falam<sup>25</sup>.

Poder-se-á então dizer, de algum modo, que o tétum é a língua veicular, o português é, em grande parte, a língua do estado, a língua indonésia é, essencialmente, utilizada nas transações comerciais e o inglês é uma língua de trabalho, nomeadamente nos contactos internacionais.

## 1.2. Sistemas de escrita: alguns aspetos

Sabemos que a escrita é uma invenção que é precedida, em muito, do desenvolvimento da linguagem oral. A linguagem oral terá surgido centenas de milhares de anos antes da escrita, forma de comunicação que constitui *“um produto cultural e não universal da espécie, na medida em que (...) existem indivíduos e comunidades humanas que não fazem uso dela”* (Veloso, 2007: 138).

Franco *et al.* (2003: 22-23) também apontam esta distinção, quando afirmam que *“a linguagem escrita é um sistema simbólico que surge na sequência do desenvolvimento da*

---

<sup>23</sup> O recenseamento de 2015 refere que o português assume o estatuto de LM para 1.384 timorenses (cf. nota 12).

<sup>24</sup> Esta questão é levantada por Albuquerque (2012: 2), quando refere que, *“[de] acordo com a fonte Timor-Leste Census of Population and Housing 2004 (NATIONAL BOARD OF STATISTICS, 2006), 37% da população é falante de língua portuguesa enquanto que o Relatório de Desenvolvimento Humano de Timor-Leste (PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO, 2002) afirma que apenas 5% da população é fluente na língua.”*

<sup>25</sup> Note-se que o estatuto de LS que o português assume em Timor-Leste vai ao encontro da opinião de Stern (1983: 16), que define a LS como sendo uma língua não-nativa aprendida e usada dentro de fronteiras territoriais em que ela tem uma função reconhecida: *“...a non-native language learnt and used with reference to a speech community outside national or territorial boundaries to which the term foreign language is commonly given. A second language usually has official status or a recognized function within a country which a foreign language has not.”*

*linguagem oral. (...) Ao contrário da linguagem oral, a linguagem escrita não se desenvolve de forma espontânea e universal, necessitando do recurso ao ensino formal.”*

No dizer de Barroso (2011: 177-179), “há diversos tipos e/ ou sistemas de escrita, com diferentes organizações e com uma quantidade variável de sinais para representar as respectivas línguas orais que lhes servem, naturalmente, de referência.” O autor afirma que existem muitos sistemas de escrita no mundo e que estes podem classificar-se **de ideográficos (ou logográficos)**, quando têm por referência unidades do conteúdo, e **de fonográficos**, quando têm, ao invés, por referência, unidades do plano da expressão, isto é, os símbolos deste tipo de escrita representam sílabas ou fonemas (e, indiretamente, também fones). Em relação a estes últimos, se a referência for a sílaba, estamos na presença do tipo silábico de escrita. Se a referência for o fonema, temos o tipo alfabético de escrita.

Os sistemas de escrita alfabética, como é o caso dos utilizados nas línguas europeias, são os mais económicos e assentam numa correspondência entre unidades fónicas segmentais e grafemas. No entanto, dada a arbitrariedade na seleção dos grafemas na representação dos segmentos fonológicos, tornou-se necessária a criação de normas que orientem esta relação e que regulem a correspondência entre o plano fónico e o plano gráfico. Ao conjunto dessas normas corresponde a ortografia. A ortografia surge, assim, para garantir a existência de uma forma única de representação gráfica de uma língua, independentemente das pronúncias diferentes – regionais ou sociais- que caracterizem as suas variedades orais em distintos momentos da sua história.

O sistema de escrita alfabético assenta num princípio geral de correspondência biunívoca entre unidade fónica e grafema. No entanto, esta correspondência não é sistematicamente cumprida, criando-se, portanto, áreas de maior complexidade ou opacidade do sistema ortográfico. Cria-se, por exemplo, uma ambiguidade fonémica quando determinado grafema pode representar vários fonemas (ex: o grafema <x> pode, por exemplo, representar os segmentos /ʃ/ (*xa*rope), /s/ (pró*x*imo), /z/ (ex*a*me) e a sequência /ks/ (pró*x*imo)) ou uma ambiguidade grafémica quando um fonema pode ser representado por vários grafemas (ex: o fonema /s/ pode ser representado graficamente por <s> (sabão), <c> (centro), <ç> (licença), <x> (próximo) e <ss> (assegura)).

Em virtude da diferente complexidade dos respetivos sistemas ortográficos, existem

línguas com sistemas de escrita fonemicamente transparentes e línguas com sistemas de escrita fonemicamente opacos. Veloso (2005: 55), por exemplo, refere que:

*“as línguas com escrita fonemicamente transparente são aquelas que melhor correspondem àquele (...) sistema, (em que) existiria uma correspondência maximamente regular, sistemática e biunívoca entre os segmentos fonológicos e os símbolos gráficos discretos. Já as línguas com escrita fonemicamente opaca consubstanciam um afastamento em relação a esse sistema ideal, pois na representação gráfica acabamos por encontrar símbolos que (...) escapam a uma correspondência directa, biunívoca e isomórfica com as cadeias segmentais da forma fonético-fonológica das palavras.”*

### **1.3. Descrição do sistema fonológico (as sibilantes) e a sua representação gráfica no português e no tétum**

Antes de nos debruçarmos sobre as questões que muito especificamente respeitam à descrição das sibilantes do português e do tétum, proceder-se-á, neste subcapítulo, a uma breve descrição dos sistemas ortográficos do português e do tétum.

O sistema ortográfico da LP caracteriza-se por ser dotado de alguma complexidade<sup>26</sup> pois afasta-se substancialmente da correspondência biunívoca entre as unidades fónicas e as unidades gráficas. A respeito dessa complexidade, Fernandes (2008: 94) afirma que:

---

<sup>26</sup> Numa referência à escala elaborada por Seymor, que tem em conta o grau de transparência e de regularidade das relações grafema-fonema, Barbeiro (2008: 5) afirma que, relativamente a outras línguas, o português apresenta um nível de complexidade médio. Numa escala decrescente, encontramos:

1 – finlandês, italiano, espanhol;

2 – grego, alemão;

3 – português, holandês;

4 – islandês, norueguês;

5 – sueco;

6 – francês, dinamarquês;

7 – inglês.

*“sendo a língua portuguesa portadora de alguma inconsistência na correspondência fonema grafema, poderá ser incluída nos sistemas intermédios em que, no dizer de I. Horta & M. Martins (2004: 214), «a correspondência fonema/grafema é irregular, uma vez que um fonema pode ser representado por uma ou mais letras (...), a mesma letra pode representar fonemas diferentes (...) ou, como no caso do h, não ter contrapartida fonológica.» «Deste modo, a ortografia do português apresenta um nível de complexidade que pode ser considerado médio face a outras línguas ocidentais.» (Barbeiro, 2006)”*.

Assim, podemos verificar que se estabelecem relações de diferentes tipos entre as unidades fónicas (fonemas materializados através de fones) e os grafemas, a saber:

- a) **um grafema (ou dígrafo) representa apenas um fonema e este é sempre representado pelo mesmo símbolo gráfico (relação biunívoca)** – na variedade padrão do português são biunívocas as relações entre os grafemas: <b, d, f, p, t, v, l> e os fonemas que estes representam. Por exemplo, não existe nenhum caso em que o fonema /b/ não seja representado pelo grafema <b>. O mesmo acontece com os fonemas /ɲ/ e /λ/ que apenas admitem os dígrafos <nh> e <lh>, sendo que estes também só representam aquelas unidades fonológicas;
- b) **um grafema representa univocamente um fonema:** o grafema <j>, por exemplo, representa apenas o fonema /ʒ/ (*janela, loja*). No entanto, este segmento fonológico pode também ser representado graficamente por <g> (*gelo, girafa*);
- c) **um fonema é representado univocamente por um grafema:** por exemplo, o fonema /r/ só pode ser representado pelo grafema <r> (*coro, preto, amor*). No entanto, este grafema pode representar também o fonema /R/ (*rádio*);
- d) **um fonema é representado por vários grafemas:** na variedade padrão do português, o caso mais evidente é o de /s/, que tem 6 representações gráficas possíveis - <ç>, <ss>, <s>, <c>, <x> e <z> (*licença, assegura, sabão, centro, próximo e produz*), sendo que a última apenas é possível em final de palavra morfológica;
- e) **um grafema representa vários fonemas:** por exemplo, o grafema <c> é utilizado para grafar os segmentos fonológicos /k/ e /s/ (*casaco e cedo*);



- f) **um grafema não representa qualquer segmento fonológico:** é o caso do grafema <h> que, quando ocorre em início de palavra, não representa qualquer segmento fonológico (*herança*);
- g) **um fonema é representado por um dígrafo:** o segmento fonológico /ʃ/ pode ser representado por <ch> (*chave*);
- h) **um grafema representa uma sequência de 2 fonemas:** o grafema <x> pode representar a sequência /ks/ (*sexo*).

Além dos distintos tipos de desfasamento entre os segmentos fonológicos e os símbolos gráficos que os representam levando, em muitos casos, a que a ortografia específica de certas palavras tenha de ser aprendida por si e memorizada, Barbeiro (2008) salienta um conjunto de fatores de outra natureza que contribuem para a complexidade da ortografia do português e que colocam entraves à aprendizagem da escrita. Assim, e para além do conhecimento sobre as relações entre fonema-grafema e grafema-fonema já enumeradas, o domínio da ortografia obriga à ativação de outras informações tais como:

- (i) fatores contextuais: por exemplo, antes de <p> e <b> a nasalidade da vogal é sempre representada por <m><sup>27</sup>;
- (ii) o facto de diferentes possibilidades de representação dos sons poderem depender também de informações de natureza morfossintática: por exemplo, as formas verbais seguidas do pronome clítico *se* (*lê-se*) escrevem-se com hífen, não devendo ser confundidas com as formas do pretérito imperfeito do conjuntivo (*lesse*), apesar da semelhança da pronúncia;
- (iii) o valor dos diferentes sinais auxiliares como o acento gráfico em *café*; o til em *romã*, o hífen em *fim-de-semana* e o apóstrofe em *borda-d'água*.

Já no que diz respeito à ortografia do tétum moderno, Hull (1999: 4) caracteriza-a como sendo “fonémica e corresponde[nte] em grande parte à pronúncia real das palavras”.

---

<sup>27</sup> Esta situação diminui, de certo modo, a opacidade da ortografia, já que restringe a determinados contextos determinadas possibilidades gráficas.

O tétum usa, tal como o português, o alfabeto latino e integra 24 letras: 18 consoantes, 5 vogais e o diacrítico <'><sup>28</sup>.

PE		TÉTUM		
Letra	Letra	Fone	Exemplo	OBSERVAÇÕES
<b>Vogais</b>				
a	a	[a]	<i>matak</i> (verde)	
		[ɨ]	<i>midar</i> (doce)	quando ocorre no segmento <ar> em final de palavra <sup>29</sup>
e	e	[e] <sup>30</sup>	<i>etu</i> (arroz)	
i	i	[i]	<i>hili</i> (apanhar)	
o	o	[o]	<i>motín</i> (motim)	
		[ɔ]	<i>kópia</i> (cópia)	Em alguns contextos em que a vogal ocorre como
u	u	[u]	<i>sunu</i> (fogo)	
<b>Consoantes</b>				
b	b	[b]	<i>bosok</i> (mentir)	
c				
d	d	[d]	<i>disionáriu</i>	
f	f	[f]	<i>foer</i> (lixo)	
g	g	[g]	<i>konsege</i> (consegue)	
h	h	[h]	<i>halo</i> (fazer)	não silencioso, ao invés do português <sup>31</sup>
j	j	[ʒ]	<i>viajem</i> (viagem)	ou [z] na pronúncia popular ( “ <i>zanela</i> ” [ze'nele])
k	k	[k]	<i>maromak</i> (Deus)	
l	l	[l]	<i>nasionál</i> (nacional)	nunca se pronuncia como [ɫ]
m	m	[m]	<i>matan</i> (olho)	em posição de coda silábica confere nasalidade à vogal anterior ( “ <i>komprimidu</i> ” [kõpri'midu])
n	n	[n]	<i>nia</i> (ela)	em posição de coda silábica confere nasalidade à vogal anterior ( “ <i>komún</i> ” [ko'mũ])
p	p	[p]	<i>paun</i> (pão)	
q				
r	r	[r]	<i>labarik</i> (criança)	no meio ou no fim de uma palavra
		[R]	<i>rai</i> (terra)	no início de uma palavra
s	s	[s]	<i>fasil</i> (fácil)	sempre como [s], independentemente do contexto
t	t	[t]	<i>tinan</i> (ano)	
v	v	[v]	<i>viajen</i> (viagem)	ou [b] na pronúncia popular ( “ <i>biajen</i> ” [bi'azẽ])
w	w	[w]	<i>wainhira</i> (quando)	apenas no tétum rural e arcaico
x	x	[ʃ]	<i>xá</i> (chá)	ou [s] na pronúncia popular ( “ <i>sá</i> ” – [‘sa])
y				
z	z	[z]	<i>kazadu</i> (casado)	ou [ʒ] na pronúncia popular ( “ <i>mezmu</i> ” [‘mezmu])
	'	[ʔ]	<i>to'os</i> (duro)	em algumas zonas, não se pronuncia

Tabela 2: Alfabetos do PE e do tétum<sup>32</sup>

<sup>28</sup> A oclusiva glotal é representada por <'>, um sinal diacrítico, um sinal de aspiração.

<sup>29</sup> Hull & Eccles (2005: 225) afirmam que “o fonema /a/ pode ser interpretado como [ɨ] quando ocorre no segmento /ar/ em final de palavra: /'midar/, /'susar/, /ha'futar/.”

<sup>30</sup> Curiosamente, Hull & Eccles (2005: 227) apenas admitem a realização fonética [e] como ocorre em “[e]tu” (arroz). No entanto, existem inúmeros casos em que o grafema <e> admite a realização fonética [ɛ], como atestam os vocábulos “h[ɛ]tan” (encontrar) e “b[ɛ]l[e]” (poder).

<sup>31</sup> Cf. Hull (1999: 6).

Se atentarmos na tabela 2, verifica-se que são apenas 3 as letras do alfabeto do PE que não se usam na ortografia do tétum: <c, q, y<sup>33</sup>>. Verifica-se, também, que o tétum possui o diacrítico <ˆ>, que não se encontra no alfabeto do PE, e que alguns grafemas não possuem, sempre e necessariamente, os mesmos valores que têm em português, como é o caso de <h> que, em tétum, é sempre aspirado.

Assim, e pese embora as respetivas especificidades, podemos facilmente concluir que as duas línguas partilham um elevado número de convenções ortográficas, nomeadamente no que concerne ao valor fónico de um conjunto alargado de grafemas.

### 1.3.1. O sistema das sibilantes em português e em tétum

Para os propósitos deste trabalho, foi considerado unicamente um domínio do sistema consonântico da língua portuguesa, mais concretamente, o conjunto das unidades sibilantes /s, z, ʃ, ʒ/. Este sistema de consoantes sibilantes é exatamente igual ao que se encontra na língua tétum.

Em relação ao vozeamento, as referidas consoantes são surdas ou [-vozeadas] (/s, ʃ/) e sonoras ou [+vozeadas] (/z, ʒ/). No que diz respeito ao ponto de articulação, as consoantes sibilantes classificam-se como dentais ([+anteriores] e [+coronais], nos casos de /s, z/), e como palatais ([-anteriores] e [+coronais], nos casos de /ʃ, ʒ/). Portanto, fonologicamente, estes segmentos consonânticos comuns ao PE e ao tétum distinguem-se entre si de acordo com a matriz representada na tabela que se segue<sup>34</sup>:

---

<sup>32</sup> Tabela elaborada com base em Hull (1999) e Hull & Eccles (2005).

<sup>33</sup> O acordo ortográfico de 1990 incluiu as letras K, W e Y no alfabeto. A Base I do Acordo Ortográfico de 2008 refere que estas letras usam-se nos seguintes casos especiais: a) em antropónimos originários de outras línguas e seus derivados: *Franklin, frankliano; Kant, kantismo; Darwin, darwinismo* [...]; b) em topónimos originários de outras línguas e seus derivados: *Kwanza; Kwuait, Kuwaitiano* [...] e c), em siglas, símbolos e mesmo em palavras dotadas como unidades de medida de curso internacional: *TWA, [...], W – oeste (West); Kg quilograma* [...]. Disponível em <http://www.portaldalinguaportuguesa.org/acordo.php?acordo&version=1990> (consultado a 3 de junho de 2017).

<sup>34</sup> Tabela elaborada com base em Mateus *et al.* (2005).

Ponto de articulação	Vozeamento	
	Não vozeadas	Vozeadas
Dentais	/s/	/z/
Palatais	/ʃ/	/ʒ/

Tabela 3: Matriz fonológica das consoantes sibilantes do português e do tétum

Tanto em português como em tétum, estes 4 segmentos fonológicos (/s, z, ʃ, ʒ/) podem ocorrer em ataque de sílaba (posição inicial ou interior de palavra), pelo que estabelecem, aí, oposições distintivas, enquanto que em posição de coda silábica, apenas o segmento /s/ tem realização fonética.

### 1.3.2. Relações fonema-grafema: o caso das sibilantes no português e no tétum

Apesar da sobreposição dos sistemas de sibilantes do português e do tétum, neste idioma<sup>35</sup> as relações que se estabelecem entre as atualizações fonéticas e as representações gráficas são bem menos opacas do que no português. A tabela que se segue sintetiza essas relações, apresentando alguns exemplos que ilustram a atualização fonética e a representação ortográfica dos segmentos em estudo, no PE e no tétum padronizado, com o objetivo de identificar as semelhanças e as diferenças entre ambos.

Posição silábica	Unidade fonológica	Realização fonética	Representação gráfica		Exemplos	
			PE	Tétum	PE	Tétum
Ataque	/s/	[s]	<ç>	<s>	licença	lisensa
			<ss>		assegura	asegura
			<s>		sabão, ensino	sabaun, ensinu
			<c>		centro, acidente	sentru, asidente
			<x>		próximo	
	/z/	[z]	<z>	<z>	zebra, razão	zebra, razaun
			<s>		abusa	abuza
			<x>		exame	ezame
	/ʒ/	[ʒ]	<g>	<j>	ginásio, religião	jináziu, relijiaun
<j>			jurista, anjo		jurista, anju	

<sup>35</sup> Para o propósito deste trabalho, não foram tidas em conta as variedades mesoletal e basiletal apresentadas por Hull & Eccles (2005: 225-226). As representações fonéticas [ʒ] e [ʃ] da variedade acroletal que aqui se considera são substituídas pelas unidades [z] e [s] na variedade mesoletal ([zɛ'nele] em vez de [ʒɛ'nele] e ['sa] em vez de ['ʃa]) e as representações [z] e [ʒ] são substituídas por [s] e [d] na variedade basiletal ([bi'site] em vez de [vi'zite] e [dĩ'tiu] em vez de [ʒĩ'tiu]) (Hull & Eccles, 2005: 239).

	/ʃ/	[ʃ]	<ch> <x>	<x>	chave, borracha xarope, enxada	xave, borraxa xarope, enxada
Coda	/s/	[ʃ]	<x> <sup>36</sup>	<s>	experiência	esperiência
			<s>		castelo	kastelu
			<z>		produz	prodús
		[ʒ]	<s>	mesmo	mezmu	
			<z>	felizmente	felizmente	

Tabela 4: Representação gráfica das sibilantes do sistema ortográfico do PE e do tétum

Em relação aos valores fonológicos dos grafemas que representam as sibilantes, a tabela 5 sintetiza as situações distintas que se registam no PE e no tétum.

Posição silábica	Representação gráfica	Realização fonética	Unidade fonológica		Exemplos	
			PE	Tétum	PE	Tétum
Ataque	<c>	[k]	/k/		<i>casa, acaso</i>	
		[s]	/s/		<i>centro, acidente</i>	
	<ç>				<i>licença</i>	
	<g>	[g]	/g/	/g/	<i>governo, agora</i>	<i>governu, agora</i>
		[ʒ]	/ʒ/	/ʒ/	<i>ginásio, religião</i>	
	<j>				<i>jurista, anjo</i>	<i>jurista, anju</i>
	<s>	[z]	/z/		<i>abusa</i>	
		[s]	/s/	/s/	<i>sabão, ensino</i>	<i>sabaun, ensinu</i>
	<ss>				<i>assegura</i>	
	<ch>	[ʃ]	/ʃ/		<i>chave, borracha</i>	
	<x>			/ʃ/	<i>xarope, enxada</i>	<i>xarope, enxada</i>
		[z]	/z/		<i>exame</i>	
		[s]	/s/		<i>próximo</i>	
		[ks]	/ks/		<i>prefixo</i>	
	<z>	[z]	/z/	/z/	<i>zebra, razão</i>	<i>zebra, razaun</i>
Coda	<s>	[ʃ]	/s/	/s/	<i>castelo</i>	<i>kastelu, prodús</i>
		[ʒ]			<i>mesmo</i>	
	<z>	[ʃ]			<i>produz</i>	
		[ʒ]		/s/	<i>felizmente</i>	<i>felizmente</i>
	<x>	[ʃ]			<i>experiência</i>	
		[ks]		/ks/	/ks/	<i>toráx</i>

Tabela 5: Valores fonético-fonológicos dos grafemas que representam as sibilantes do PE e do tétum

<sup>36</sup> O grafema <x>, além de poder representar os fonemas /s, z, ʃ/, pode também assumir o valor de /ks/, como atestam os casos de <tórax> e <fixar>.

A análise dos dados apresentados nas tabelas 4 e 5 leva-nos a concluir que, no caso das sibilantes do PE, não há qualquer caso de relação biunívoca entre fonema e grafema, registando-se, por outro lado, diferentes tipos de desajustes entre o plano fónico e o plano gráfico:

1. **um segmento fonológico sibilante pode ser representado por vários grafemas** (o fonema /s/ pode, por exemplo, ser representado, em posição de ataque, por <ç, ss, s, c, x>);
2. **um grafema pode representar vários fonemas** sibilantes (o grafema <x> pode, por exemplo, em posição de ataque, representar os segmentos /ʃ, s, z/ e a sequência /ks/ (cf. também a situação inventariada em 5);
3. **um grafema representa univocamente um fonema** sibilante (o grafema <j> apenas pode representar graficamente o segmento /ʒ/);
4. **um fonema sibilante pode ser representado por um dígrafo** (o segmento /ʃ/ pode ser representado por <ch>);
5. **um grafema pode representar uma sequência de dois fonemas, um dos quais sibilante** (o grafema <x> pode representar graficamente a sequência /ks/<sup>37</sup>).

No caso do tétum, as relações entre o plano fónico e o plano gráfico são bem menos complexas. Em posição de ataque silábico, os grafemas representam biunivocamente um fonema (<s, z, j, x> representam, respetivamente, /s, z, ʒ e ʃ/). Já em posição de coda, verifica-se que um segmento fonológico pode ser representado por vários grafemas (/s/ é representado por <s> quando realiza [ʃ] (ex: *kastelu*) e é representado por <z> (ex: *mezmu*) quando realiza [ʒ]). Ao contrário do que acontece em português, há uma motivação fonética pois, embora o segmento fonológico seja o mesmo, as realizações são diversas, podendo concluir-se, então, que, neste caso, a ortografia do tétum é mais fonética do que fonológica.

Admitamos que a influência do tétum e a transferência do sistema ortográfico desta língua para o registo escrito do português constitua uma das causas que poderão explicar grande parte dos desvios produzidos pelos informantes em estudo no presente trabalho,

---

<sup>37</sup> Em PE, o grafema <x> pode realizar-se foneticamente como [ks] (ex. *prefixo*), correspondendo, na ortografia do tétum padrão, ao dígrafo <ks> (ex. *prefiksu*).

hipótese que se procurará confirmar ou infirmar através da análise dos dados empíricos recolhidos para este estudo.

---

## **2. Metodologia**

---



Depois de apresentar o contexto (socio)linguístico de Timor-Leste e de descrever os sistemas fonológicos de sibilantes do PE e do tétum<sup>38</sup>, com as respetivas representações gráficas<sup>39</sup>, apresentar-se-á, de seguida, a metodologia de recolha e tratamento do *corpus*. Assim, o presente capítulo desenvolver-se-á em três partes: numa primeira, descrever-se-ão os instrumentos de recolha dos dados, apresentando-se o questionário utilizado para traçar o perfil sociolinguístico dos informantes e o exercício de produção escrita; numa segunda parte, apresentar-se-á o perfil sociolinguístico dos informantes - autores dos textos nos quais se recolheram os dados que constituem o *corpus* de descrição e análise deste trabalho. Além de incidir sobre variáveis como a idade e o género, a descrição do perfil dos informantes considerará, essencialmente, aspetos relacionados com a LM, com o conhecimento prévio de outras línguas e com o modo e a frequência do contacto com o português. Numa terceira parte, apresentar-se-ão os procedimentos de recolha dos dados e a metodologia utilizada no seu tratamento.

## 2.1. Instrumentos de recolha de dados<sup>40</sup>

Como já foi referido anteriormente, este estudo tem como principal objetivo descrever e analisar os desvios à norma gráfica do PE na representação das sibilantes por parte de alunos do Ensino Superior (ESUP) leste-timorense. Para esse efeito, recorreu-se a um exercício de produção escrita<sup>41</sup>, desenvolvido em contexto de sala de aula, por jovens oriundos de todos os municípios<sup>42</sup> do país, para os quais o português não é LM e que, na

---

<sup>38</sup> Cf. tabela 3.

<sup>39</sup> Cf. tabelas 4 e 5.

<sup>40</sup> Numa fase inicial, este estudo, enquadrado no curso de Mestrado em Ciências da Linguagem da Universidade do Minho e sob a orientação do Professor Doutor Henrique Barroso, tinha como principal objetivo identificar os desvios mais frequentes à norma ortográfica do PE, observando a produção escrita de uma amostra de alunos do ESUP timorense. O instrumento de recolha de dados é o então desenhado, tendo igualmente sido analisados os dados então recolhidos.

<sup>41</sup> Cf. anexo 1.

<sup>42</sup> Ao longo do presente trabalho, utilizar-se-á a terminologia *distrito* ou *município* de forma indistinta. Apesar de, na página de *internet* do governo, se optar pela divisão administrativa de Timor-Leste em municípios, grande parte dos falantes continua a preferir a designação de *distrito*.

altura da sua aplicação, se encontravam a frequentar o 6.º semestre da 1.ª Licenciatura de Formação de Professores dos 1.º e 2.º Ciclos do Ensino Básico (EB) do INFORDEPE<sup>43</sup>.

Este curso<sup>44</sup> visa a formação de 325 jovens (25 de cada município), no sentido de colmatar as necessidades que irão sentir-se com a aprovação, por parte do governo de Timor-Leste, da Lei que regula a atribuição da pensão de velhice<sup>45</sup>. Segundo dados internos do Gabinete da Formação Académica<sup>46</sup> do INFORDEPE, a lei em causa fará com que 2.168 professores dos 1.º e 2.º Ciclos do EB se aposentem até ao final do ano de 2022<sup>47</sup>. Saliente-se que estes novos docentes deverão integrar o Sistema Educativo timorense na categoria de assistentes<sup>48</sup>, pelo que, finda a formação, deverão, em teoria, possuir as competências mínimas necessárias para exercer a docência em português<sup>49</sup>.

Tendo, entretanto, alguns alunos desistido do curso, encontravam-se a frequentar as atividades letivas do 6.º semestre, na altura da aplicação dos instrumentos de recolha de dados, apenas 313 formandos. O curso decorreu em 3 polos de formação, a saber: (i) em Díli, com 6 turmas, constituídas pelos alunos oriundos dos municípios de Díli, Oe-cusse,

---

<sup>43</sup> O artigo 4.º do Decreto-Lei n.º 04/2011 de 26 de janeiro define o INFORDEPE como um instituto académico, de formação e de investigação, que tem por missão promover a formação académica e profissional de pessoal docente e de profissionais do Sistema Educativo (Decreto-Lei n.º 04/2011 de 26 de janeiro, pp. 4.561-4.569).

<sup>44</sup> Este curso foi criado ao abrigo de um Protocolo de Cooperação entre o Ministério da Educação de Timor-Leste e o Ministério dos Negócios Estrangeiros e Cooperação de Portugal. Esse protocolo estabelece a criação do Projeto de Formação Inicial e Contínua de Professores (PFICP). A elaboração do plano de estudos, bem como a docência de todas as disciplinas, à exceção do tétum, é da responsabilidade dos professores portugueses do PFICP e conta com a supervisão pedagógica da Universidade do Minho, que exerce funções de assessoria científica e pedagógica (Gabinete da Formação Académica do INFORDEPE, Documentos organizativos da Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do EB, Nota Introdutória).

<sup>45</sup> A Lei n.º 06/2012 de 29 de fevereiro que aprova o regime transitório de segurança social define, no artigo 8.º, que a idade mínima para adquirir o direito à reforma é de 60 anos (Lei n.º 06/2012, p. 5.715).

<sup>46</sup> O Gabinete de Formação Académica do INFORDEPE é “... o serviço responsável pela execução das políticas de formação académica do pessoal docente” (Decreto-Lei n.º 04/2011 de 26 de janeiro, p. 4.566).

<sup>47</sup> Segundo dados apresentados pelo Coordenador do Gabinete da Formação Académica, Luís Manuel da Costa Fernandes, 2.703 professores do Sistema Educativo timorense têm uma idade superior a 55 anos, sendo que 2.168 são dos 1.º e 2.º Ciclos do EB (reunião de apresentação do INFORDEPE ao Ministro Coordenador dos Assuntos Sociais e Ministro da Educação, Fernando “La Sama” de Araújo, no dia 17 de fevereiro de 2015).

<sup>48</sup> O Estatuto da Carreira Docente prevê as categorias profissionais hierarquizadas de assistente, professor e professor sénior (Decreto-Lei n.º 23/2010 de 9 de dezembro, artigo 8.º, p. 4.457).

<sup>49</sup> O Quadro de Competências Obrigatórias do Pessoal Docente organiza-se em 4 setores do conhecimento: Domínio das Línguas Oficiais; Conhecimento Técnico-Científico na respetiva área e grau de ensino; Técnicas Pedagógicas e Ética Profissional (Decreto-Lei n.º 23/2010 de 9 de dezembro, artigo 8.º, p. 4.453).

Aileu, Liquiçá, Ainaro e Manufahi, perfazendo um total de 144; (ii) em Baucau, com 4 turmas, constituídas pelos alunos oriundos dos municípios de Baucau, Manatuto, Lautém e Viqueque, correspondendo a um universo de 96; (iii) em Maliana, com 3 turmas, constituídas pelos alunos oriundos dos municípios de Maliana, Ermera e Covalima, contabilizando um total de 73.

A diversidade etnolinguística<sup>50</sup> que caracteriza este grupo exigiu a aplicação de um questionário para determinar o perfil sociolinguístico dos informantes, assim como a frequência de uso das diferentes línguas nas variadas situações de interação comunicativa (português, tétum, LM e LE, especialmente o inglês e o indonésio). Este questionário sociolinguístico, composto por 13 perguntas, revelou-se uma mais-valia, pois forneceu informações relevantes, também referentes à perspetiva dos inquiridos relativamente ao uso e ao domínio da língua portuguesa.

O questionário sociolinguístico<sup>51</sup> foi elaborado seguindo critérios relevantes para o propósito deste trabalho, no sentido de otimizar a análise dos resultados obtidos. Em primeiro lugar, não poderia ser muito extenso, para não causar fadiga nos inquiridos que, assim, poderiam pensar bem nas respostas. Em segundo lugar, por forma a permitir o tratamento quantitativo dos resultados e evitar dispersão na análise, o questionário teria de ser elaborado com respostas fechadas. Em terceiro lugar, dada a presumível proficiência elementar dos informantes em português, considerou-se oportuno fornecer a tradução das questões para tétum.

Relativamente à recolha dos dados necessários para a construção do *corpus*, e no sentido de estimular a produção textual através de um tema que apelasse ao interesse dos informantes, a tarefa consistiu na redação de um texto de opinião sobre a importância da profissão docente, com uma extensão aproximada de 20 linhas. Para tal, foi solicitado aos professores do PFICP<sup>52</sup> que lecionam o curso a entrega do enunciado<sup>53</sup> aos seus alunos. A escolha do tema da produção escrita teve em atenção o perfil do público-alvo, as suas

---

<sup>50</sup> Cf. nota 4.

<sup>51</sup> Cf. anexo 1.

<sup>52</sup> Cf. nota 44.

<sup>53</sup> Cf. anexo 1.

motivações e o seu futuro profissional, pois não nos podemos esquecer de que a formação que se encontravam a frequentar visa habilitá-los para o exercício de funções docentes.

O questionário e o exercício de produção escrita<sup>54</sup> foram aplicados no dia 17 de novembro de 2014<sup>55</sup> em cada uma das turmas dos 3 polos de formação do INFORDEPE (Díli, Baucau e Maliana). Nesse dia, dos 313 inscritos, estiveram presentes 279 estudantes que constituem, assim, a amostra de sujeitos participantes no presente estudo.

## 2.2. Perfil dos informantes

Tendo em conta que a amostra de informantes integra alunos do ESUP leste-timorense, importa, antes de mais, apresentar, mesmo que de forma muito sucinta, as instituições de ESUP acreditadas em Timor-Leste.

Os dados apresentados pela Agência Nacional para a Avaliação e Acreditação Académica (ANAAA)<sup>56</sup> registam a existência de 11 instituições de ESUP acreditadas sendo, para já, a Universidade Nacional de Timor Lorosa'e (UNTL), a única instituição pública acreditada por esta agência.<sup>57</sup>

---

<sup>54</sup> Por forma a manter o anonimato dos alunos, foi atribuído um código constituído por 3 letras (as 3 primeiras letras do município de origem) e por um número que varia entre o 1 e o 279 (exemplos: DIL001 e COV279).

<sup>55</sup> O exercício de produção escrita e o questionário sociolinguístico foram aplicados pelos docentes do PFICP, tendo sido solicitado aos alunos que assinassem uma declaração a autorizar que os materiais produzidos fossem utilizados para investigações realizadas no âmbito de uma Dissertação de Mestrado (cf. anexo 3). Para tal, foram dadas algumas orientações por escrito aos professores aplicadores, no sentido de uniformizar as condições de aplicação do exercício de produção escrita (cf. anexo 2 – Orientações de aplicação do exercício). Após a entrega dos envelopes por parte dos professores aplicadores, os enunciados da produção escrita e os questionários foram ordenados por município de origem e foi atribuído um código a cada um dos informantes.

<sup>56</sup> A ANAAA foi criada através do Decreto-Lei n.º 21/2010 de 1 de dezembro e é um organismo autónomo cujo objetivo é proceder à acreditação das instituições de ESUP, através da avaliação da respetiva qualidade. Esta instituição, tal como a UNTL e o INFORDEPE, está na dependência direta do Ministério da Educação.

<sup>57</sup> Apesar de ter sido criado pelo Decreto-Lei n.º 04/2011 de 26 de janeiro, o INFORDEPE tem enfrentado algumas dificuldades na sua acreditação institucional. As indefinições jurídicas e políticas relativas ao seu estatuto levaram, inclusivamente, a que os 279 estudantes do curso de Licenciatura (os informantes deste estudo) e os 51 estudantes do curso de Mestrado tivessem sido, ao abrigo de um acordo assinado a 14 de dezembro de 2014, transferidos para a UNTL para poderem obter os seus diplomas académicos. Neste

A tabela que se segue apresenta essas 11 instituições indicando o número dos estudantes matriculados em cada uma delas, no ano de 2015<sup>58</sup>.

	Nome da Instituição	Acrónimo	N.º de alunos inscritos em 2015
1	Universidade Nacional Timor Lorosa'e	UNTL	11 158 <sup>59</sup>
2	Universidade da Paz	UNPAZ	12 355
3	Universidade de Dili	UNDIL	2 797
4	Universidade Oriental Timor Lorosa'e	UNITAL	9 161
5	Dili Institute of Technology	DIT	4 424
6	Instituto Católico para Formação de Professores	ICFP	212
7	Instituto de Ciências Religiosas "São Tomás de Aquino"	ICR	372
8	East Timor Institute of Business	IOB	3 688
9	Instituto Superior Cristal	ISC	6 326
10	East Timor Coffee Institute	ECTI	909
11	Instituto Profissional de Canossa	IPDC	459
<b>Total de estudantes inscritos em 2015</b>			<b>51 861</b>

Tabela 6: Instituições acreditadas pela ANAAA, com o respetivo n.º de alunos matriculados em 2015

Como referido anteriormente, no presente estudo participou um grupo de 279 alunos do ESUP leste-timorense, constituído por 123 sujeitos do género masculino e 156 do género feminino, com idades entre os 21 e os 28 anos. Tendo em consideração que, em 2015, existiam mais de 51 mil alunos matriculados nas várias instituições de ESUP, não se pretende, como é óbvio, generalizar as observações sobre o comportamento gráfico dos estudantes do ESUP timorense. No entanto, é de realçar que se trata de uma amostra bastante relevante do ponto de vista qualitativo, tendo por referência o universo dos estudantes de cursos de formação de professores do EB<sup>60</sup>.

momento, e aguardando uma decisão política clara sobre as competências legais do INFORDEPE, este instituto orienta as suas formações apenas para a vertente da formação contínua de professores.

<sup>58</sup> Segundo dados apresentados pela ANAAA em <http://www.anaaa.gov.tl/publicacoes-1/relatorio-estatistico> (consultado a 15 de novembro de 2016).

<sup>59</sup> Os informantes deste estudo estão incluídos no número de estudantes matriculados na UNTL para o ano de 2015.

<sup>60</sup> Segundo dados apresentados pela ANAAA, apenas 2 instituições de ESUP oferecem esta formação académica (Bacharelato e Licenciatura), a UNTL e o Instituto Católico de Formação de Professores (ICFP). Para se perceber melhor a relevância da amostra deste estudo, se atentarmos apenas na formação de professores do EB, é importante referir que estas 2 instituições graduaram, em 2015, apenas 185 estudantes (a UNTL, 133 e o ICFP, 52).

Para efeitos de caracterização do perfil dos informantes deste estudo, foram considerados os dados resultantes da aplicação do questionário sociolinguístico.

Avançamos desde já que, no que concerne ao município de origem, se verifica, na amostra, algum equilíbrio entre os 13 municípios que constituem o território de Timor-Leste. Ainda assim, o distrito de Díli é o que está menos representado (15 informantes), por oposição aos distritos de Ainaro, Baucau e Ermera (25 informantes cada).

	DISTRITO DE ORIGEM													TOTAL
	Díli	Ainaro	Aileu	Manufahi	Oe-Cusse	Liquiçá	Baucau	Viqueque	Lautém	Manatuto	Ermera	Bobonaro	Covalima	
N.º de informantes	15	25	21	24	24	24	25	16	18	17	25	23	22	279

Tabela 7: Distribuição dos informantes por município de origem

### 2.2.1. Apresentação dos dados do questionário<sup>61</sup>

Sabendo-se que, em Timor-Leste, o português não é a LM da grande maioria da população, era importante identificar a LM de cada um dos aprendentes, oriundos de todos os municípios, no sentido de tentar perceber possíveis relações entre os tipos de desvios registados e a LM de cada subgrupo de informantes. Com esse objetivo se formulou a **Questão n.º 1 – Em que língua começou a comunicar no seio familiar?**

No que diz respeito aos resultados, verifica-se a presença de 13 LM diferentes entre os aprendentes, sendo que o tétum assume uma posição dominante, com 116 sujeitos (41,5%). Também com algum destaque encontra-se o mambae, LM de 53 dos aprendentes, o que corresponde a 19% dos inquiridos. O facto de o mambae figurar em segundo lugar não será surpreendente se tivermos em conta que, em 2001, esse idioma era falado em cerca de

<sup>61</sup> Para uma informação mais detalhada dos resultados obtidos na aplicação do questionário, cf. anexo 4.

23% dos *sucos*<sup>62</sup> e encontrava-se seguido pelo tétum, com 20%<sup>63</sup>.

Das restantes LM identificadas, destaca-se a presença de línguas como o baikeno, o makasae, o tukudede, o fataluku, o kemak e o bunak, ainda com alguma representatividade no universo em estudo. Já a presença do idaté, do rahesuk, do lolein, do lakalei e do galoli é pouco expressiva, variando entre 4 e apenas 1 aprendente.

**Questão n.º 2 – Que línguas já aprendeu durante a sua vida? Ordene-as, começando por aquelas que domina melhor.**

Na resposta a esta questão, os inquiridos tinham de ordenar, numa escala de 1 a 5 (sendo que 1 significava a língua em que consideravam ter mais proficiência e 5, a língua com uma proficiência mais elementar), as línguas que conheciam. Com esta questão, pretendia-se traçar o perfil linguístico dos jovens para, mais uma vez, determinar a importância dessa variável na análise dos desvios apresentados.

Todos os informantes afirmam dominar a língua tétum, sendo que, para 139 (49,8%), essa é a língua que melhor dominam; para 133 (47,7%), é a língua que dominam em segundo lugar e, para os restantes 7 (2,5%), é a língua que dominam em terceiro lugar. De referir que sempre que o tétum aparece em segundo lugar, a LM é aquela que figura em primeiro.

Apenas um informante (AIN019) afirma que o português é a língua que melhor domina. Outros 43 aprendentes (15,4%) consideram este idioma a segunda língua que melhor dominam, 108 (38,7%) como sendo a terceira, 119 (42,6%) como sendo a quarta e, finalmente, 6 (2,1%) como sendo a quinta. Saliente-se que apenas dois informantes (LAU176 e LAU186) não incluem o português nas 5 línguas que melhor dominam.

---

<sup>62</sup> A página do governo de Timor-Leste (<http://timor-leste.gov.tl/?p=91&lang=pt>) define *suco* como sendo a menor divisão administrativa de Timor-Leste, composta por uma ou mais aldeias.

<sup>63</sup> *The 2001 Survey of Sucos*, publicado em outubro de 2001, pela *East Timor Transitional Administration*, o *Asian Development Bank*, o *World Bank*, e pelo *United Nations Development Programme*, p.32.

No que diz respeito ao indonésio, 1 jovem - ERM216 - (0,4%) afirmou ser essa a língua que melhor domina, 75 (26,8%) referiram-na como sendo a segunda, 133 (47,7%) a terceira, 40 (14,3%) a quarta e, por fim, 7 (2,5%) como sendo a quinta.

Como era expectável<sup>64</sup>, nenhum dos informantes indicou o inglês como sendo a língua que melhor domina. Três aprendentes (1%) indicaram-na como sendo a segunda, 24 (3,3%) consideram-na como sendo a terceira, 61 (21,9%) a quarta e, finalmente, 66 (23,7%) a quinta.

É ainda importante referir que, quando questionados sobre a língua que melhor dominam, a maioria indicou a LM. É ainda curioso verificar que há 23 aprendentes que afirmam que dominam melhor o tétum do que a sua própria LM. Este facto será facilmente explicável se tivermos em consideração que estes jovens estão a estudar em Díli já há alguns anos, havendo, por isso, um maior contacto com a língua veicular, o tétum.

Da análise dos dados apresentados depreende-se, portanto, que o contacto com a língua portuguesa acontece, geralmente, depois da exposição ao tétum, quando este assume o estatuto de língua materna, ou ao mesmo tempo do que o tétum, quando a LM é outra.

É igualmente relevante referir o facto de o primeiro contacto significativo com a língua portuguesa ocorrer em contexto formal (na escola). É importante não esquecer também que, para muitos destes jovens, sobretudo para os mais velhos, o Ensino Básico foi todo ele realizado em língua indonésia, o que, nestes casos, relega o português para a 4.ª posição, considerando a cronologia de exposição ao respetivo *input* linguístico.

### **Questão n.º 3 – Qual a língua que privilegia para comunicar em contexto informal?**

Sabendo, à partida, que os falantes de Timor-Leste adequam a utilização das línguas que dominam ao contexto em que se encontram, caracterizando-se a situação como de

---

<sup>64</sup> Os dados apresentados em *Timor-Leste Population and Housing Census 2015*, referem a existência de apenas 7271 cidadãos cuja LM é o inglês (disponível em <http://www.statistics.gov.tl/category/publications/census-publications>, consultado a 18 de novembro de 2016).



bilinguismo social com diglossia<sup>65</sup>, pretendeu-se, com esta questão, perceber que língua(s) os inquiridos mais utilizam no seu dia a dia e se, eventualmente, em alguma destas situações recorrem à língua portuguesa.

Os resultados apresentados<sup>66</sup> evidenciam que, em ambientes informais, 257 dos inquiridos, ou seja, 92,1% comunicam, maioritariamente, em tétum, sendo que apenas 10 (3,6%) afirmam recorrer à língua portuguesa. No que diz respeito às outras línguas utilizadas, 5 jovens afirmam recorrer ao mambae (1,8%), 4 ao bunak (1,4%) e um ao kemak e ao rahesuk (0,4%).

Da análise dos dados apresentados, pode concluir-se que o tétum, apesar de ser a LM de apenas 117 dos informantes, é a língua utilizada pela maioria dos inquiridos para comunicar em contexto informal, relegando, neste domínio, os outros idiomas para segundo plano.

#### **Questão n.º 4 – Qual a língua que privilegia para comunicar em contexto formal?**

Esta questão pretendia esclarecer quais as línguas utilizadas pelos informantes em contexto formal, por forma a aferir o uso real da língua portuguesa, uma das LO do país.

Tendo em consideração o panorama linguístico de Timor-Leste, já se antecipava que o tétum e o português seriam as línguas com maior predominância de uso em contextos formais.

Além disso, se se comparem estes com os resultados obtidos no âmbito da questão anterior, verifica-se que o tétum continua a ser a língua privilegiada, com um total de 185 aprendentes (66,3%) que afirmam preferir o seu uso em contextos formais. No entanto, assiste-se a um aumento significativo do recurso à língua portuguesa, pois 93 jovens (33,3%) afirmam usá-la neste tipo de contextos.

---

<sup>65</sup> Cf. nota 14.

<sup>66</sup> Cf. anexo 4.

### **Questão n.º 5 – Que língua considera a mais importante para o seu futuro?**

As respostas a esta questão permitem ter a noção da necessidade sentida por estes jovens em relação à aprendizagem da LP comparativamente com outras.

Através da análise dos resultados apresentados, percebe-se, desde logo, a importância que o português assume junto dos inquiridos. Em 279 questionários, 268 aprendentes, ou seja, 96% dos inquiridos apontaram a língua portuguesa como sendo a mais importante para o futuro, o que confirma a necessidade que estes jovens sentem em aprendê-la. Os restantes 11 inquiridos (4%) atribuem ao tétum um lugar de destaque. A escolha dos inquiridos estará seguramente ligada ao facto de, sendo LO, o seu conhecimento poder trazer-lhes alguns benefícios no futuro. No entanto, esses benefícios não são, para os inquiridos, comparáveis aos da aprendizagem do português, que é visto como a língua que lhes oferece mais garantias e utilidade para o futuro.

Apesar de serem consideradas línguas de trabalho, a língua indonésia e a língua inglesa não são vistas, por nenhum dos informantes, como línguas que possam ser úteis no futuro. Este resultado é muito interessante, se considerarmos a projeção que a língua inglesa tem a nível internacional e o facto de a Indonésia e a Austrália serem países vizinhos.

### **Questão n.º 6 – Que diploma(s) de língua portuguesa possui?**

Desde o ano de 2000 que o ensino da língua portuguesa em Timor-Leste tem estado, essencialmente, a cargo da Cooperação Portuguesa, responsável pela implementação de 3 projetos<sup>67</sup>, a saber: (i) o Projeto de Reintrodução da Língua Portuguesa (PRLP), entre 2000 e 2007; (ii) o Projeto de Consolidação da Língua Portuguesa (PCLP), entre 2008 e 2011, e (iii) o PFICP, entre 2012 e 2014<sup>68</sup>. Ao longo destes anos, estes projetos ofereceram vários tipos de

---

<sup>67</sup> Estes Projetos foram criados ao abrigo de um Protocolo de Cooperação entre o Ministério da Educação de Timor-Leste e o Ministério dos Negócios Estrangeiros e Cooperação de Portugal.

<sup>68</sup> No dia 22 de março de 2016, foi assinado um protocolo de cooperação entre o Ministério da Educação da RDTL e o Ministério dos Negócios Estrangeiros e Cooperação de Portugal, com o objetivo de criar o Projeto Formar Mais – Formação Contínua de Professores. Este Projeto prevê a colocação de um professor de língua portuguesa em cada uma das 11 maiores escolas do 3.º ciclo do EB, apresentando duas componentes essenciais: (i) a formação para o aperfeiçoamento de competências em língua portuguesa de professores do

curso de língua portuguesa a professores e funcionários públicos. Estes cursos encontravam-se divididos em vários tipos, a saber:

- a) Nível 1, Nível 2, Nível 3 e Preparação para o Bacharelato, destinados a professores;
- b) Oficinas de Língua e Especialização, para os funcionários públicos.

Refira-se que, muitas vezes, e no sentido de completar o número de formandos por turma, a Cooperação Portuguesa permitia a inclusão de membros da população em geral que manifestassem vontade de frequentar estes cursos. A inclusão desta questão no questionário permite compreender se há informantes nestas condições, fator indicativo da vontade que têm de aprender português.

A análise dos dados obtidos permite verificar que mais de metade dos aprendentes (163, equivalente a 58,5%) nunca frequentou qualquer curso de língua portuguesa. Dos restantes 119 informantes (42,6%), 39 (14%) possuem o certificado de Nível 1; 22 (7,8%), o de Nível 2; 10 (3,6%), o de Nível 3; 7 (2,5%), o de Preparação para o Bacharelato e 31 (11,1%), o de Oficinas de Língua. A informação recolhida através desta questão torna-se deveras importante, na medida em que pode concluir-se que quase metade destes jovens possui um certificado de um curso de língua portuguesa, tendo-o frequentado de livre e espontânea vontade, pois esta formação era destinada, maioritariamente, aos docentes timorenses.

#### **Questão n.º 7 – Em termos globais, como considera o seu domínio do português?**

A inclusão desta pergunta no questionário sociolinguístico tinha como objetivo obter a autoavaliação dos aprendentes relativamente ao seu domínio do português.

---

3.º Ciclo do EB e (ii) a formação para o desenvolvimento de competências no âmbito dos conteúdos programáticos dos 7.º, 8.º e 9.º anos de escolaridade da disciplina de Língua Portuguesa. Além disso, o Projeto Formar Mais prevê também o apoio de 15 docentes do Ensino Secundário Geral (ESG), um para cada uma das 15 disciplinas que compõem o currículo, com o objetivo de apoiar o INFORDEPE na formação científico-pedagógica de professores, no âmbito dos conteúdos programáticos das diversas áreas disciplinares do Currículo do ESG.

No que diz respeito à análise dos dados, verifica-se que 163 aprendentes (58,5%) consideram ter um bom domínio do português; 63 (22,6%) afirmam dominá-lo suficientemente e, finalmente, 52 inquiridos (18,6%) apontam o seu domínio como sendo muito bom. É com alguma estranheza que se constata que nenhum dos inquiridos afirmou possuir um domínio insuficiente ou muito insuficiente da LP.

### **Questão n.º 8 – Onde/Como aprendeu português?**

Sabendo-se, à partida, que o português não é a LM destes jovens, considerou-se pertinente a inclusão desta questão para verificar a importância do ensino formal como meio de acesso à língua portuguesa.

No que concerne aos resultados, 276 inquiridos (98,9%) afirmam ter aprendido português na escola, 2 (0,7%) através da leitura de jornais, livros e revistas e 1 (0,4%) em casa, sendo de referir que as opções “*Através da rádio ou da televisão*” e “*Outro local*” não foram selecionadas por nenhum dos inquiridos.

Os resultados evidenciam que a escola se assume, claramente, como o meio de difusão mais importante do português para estes jovens. Conhecendo a realidade da difusão da língua portuguesa no território timorense, não é de todo surpreendente que o ensino formal assuma este protagonismo. De facto, à exceção de um espaço informativo diário de cerca de 20 minutos na RTTL (Rádio e Televisão de Timor-Leste) e de uma edição semanal do jornal *O Semanário* (com edição bilingue em português e tétum), poucos são os conteúdos existentes em português na imprensa. Além disso, e apesar de a RTTL transmitir, fora do seu horário reduzido de emissão, a programação da Rádio Televisão Portuguesa Internacional (RTPI), os conteúdos aí apresentados não são, de maneira alguma, orientados para a população timorense em geral e para os inquiridos em particular.

### **Questão n.º 9 – Em que situações usa, habitualmente, o português?**

Pretendia-se, com esta questão, avaliar o uso do português em vários contextos. Tal como aconteceu na anterior, as respostas à presente questão indicam que a escola é o contexto principal no qual estes jovens universitários contactam com a língua portuguesa.

Como era expectável, a hipótese “*Nas aulas*” foi escolhida por 258 dos jovens, ou seja, por 92,5% dos inquiridos.

Em segundo lugar, a resposta “*Com amigos e colegas*”<sup>69</sup> foi assinalada 10 vezes, correspondendo a 3,6% dos inquiridos. A resposta “*Com colegas portuguesas ou brasileiros*” foi escolhida 8 vezes, representando 2,9% dos questionários aplicados. A hipótese “*Nos serviços públicos*” foi escolhida por apenas 3 dos inquiridos, correspondendo a uma percentagem de 1,1%. Por fim, as opções “*Em casa*” e “*Outro local*” não foram assinaladas por nenhum dos inquiridos.

#### **Questão n.º 10 – Onde costuma ouvir pessoas a falar em português?**

A inclusão desta questão afigurava-se essencial, no sentido de perceber o uso do português no dia a dia, em Timor-Leste.

“*Nas aulas de língua portuguesa*” foi a hipótese escolhida por 232 dos jovens, representando 83,1% do total dos inquiridos. Em segundo lugar, no que à frequência diz respeito, 33 jovens escolheram a resposta “*Na rádio ou na televisão*”, o que equivale a 11,8% dos questionários aplicados. Com valores mais baixos, encontra-se a resposta “*Nos serviços públicos*”, que foi assinalada por 11 dos jovens, o que corresponde a 3,9%. As opções “*No mercado e nas lojas*”, “*Na rua*”, “*Na Igreja*” e “*Noutro local*” não foram escolhidas por nenhum dos inquiridos, o que reforça a ideia de que, para estes jovens, o português é muito pouco utilizado fora do contexto escolar. Recolheram-se, ainda, 3 questionários em que esta questão não foi respondida.

#### **Questão n.º 11 – Em que local costuma ler em português?**

Pretendia-se, com esta questão, compreender a opinião dos inquiridos relativamente ao acesso a textos escritos em português.

---

<sup>69</sup> Foi explicado aos inquiridos que, nesta opção, não deveriam ser incluídos colegas portuguesas ou brasileiros.

Relativamente aos resultados, verificou-se que a hipótese “*Na escola*” foi, mais uma vez, a resposta escolhida pela maioria dos inquiridos. 267 jovens, ou seja, 96% dos informantes, escolheram a escola como sendo o local onde costumam ler em português. Esta escolha vem reforçar a ideia de que a escola assume uma posição de grande destaque como meio de difusão da LP.

Em segundo lugar, em termos de frequência, 7 jovens, ou seja, 2,5% dos inquiridos, escolheram a hipótese “*Em casa*” como sendo o local onde costumam ler em português.

As respostas “*Nos serviços públicos*” e “*Na televisão*” aparecem em terceiro lugar, com 1 jovem a escolher cada uma destas hipóteses, representando, em cada caso, apenas 0,4% da totalidade dos inquiridos.

Saliente-se, por fim, que nenhum dos aprendentes apresentou a opção “*Noutro local*” para responder a esta questão e que 3 dos jovens, ou seja, 1,1%, não apresentaram respostas.

#### **Questão n.º 12 – Em que situações costuma escrever em português?**

Pretendia-se, também aqui, perceber se estes jovens sentem a necessidade de recorrer à LP escrita fora do contexto escolar.

A análise dos resultados a esta questão mostra, mais uma vez, que a escola é o local quase exclusivo de produção escrita em língua portuguesa. Como se pode verificar, 276 jovens, representando 98,9% dos inquiridos, escolheram a resposta *Na escola*, e apenas 3 dos aprendentes escolheram a opção “*Nos serviços públicos*”, representando os restantes 1,1%.

#### **Questão n.º 13 – Aprender português é importante porque...**

Com esta questão pretendia-se verificar por que motivo é que os aprendentes consideravam importante a aprendizagem do português.

A resposta que surgiu com maior frequência foi a opção “*É uma das línguas oficiais de Timor-Leste*”, assinalada 254 vezes, correspondendo a 91% do total dos questionários.

Bastante distante aparece a hipótese “*Preciso dela para continuar a estudar*”, assinalada em 13 questionários, o que representa 4,6% dos aprendentes. A hipótese “*Pode servir para comunicar com pessoas de outros países*” é selecionada 5 vezes, o que representa 1,8% dos inquiridos, e 4 inquiridos, ou seja 1,4%, escolheram a hipótese “*Preciso dela para conseguir um bom trabalho*”. De registar que as hipóteses “*Preciso dela para comunicar no dia a dia*” e “*Não é importante*” não foram selecionadas por nenhum dos inquiridos. É de referir também que esta questão não foi respondida em 3 questionários, ou seja, em 1,1% dos casos.

Em suma, à exceção de 1 caso (BOB240) que afirmou ter aprendido português em casa, podemos afirmar que a língua portuguesa chegou a estes jovens através da escola. Salvo raras exceções, a escola é o espaço privilegiado de contacto com esta língua nova sendo, inclusivamente, o único espaço onde estes aprendentes escrevem, leem, ouvem e usam a língua portuguesa. O tétum assume um lugar de destaque na comunicação diária (contextos formais e informais) mas é à língua portuguesa que é atribuída maior importância para a vida futura.

### 2.3. Procedimentos de tratamento dos dados

Como já foi referido anteriormente, o presente estudo tem como principal objetivo identificar as dificuldades na representação gráfica das consoantes sibilantes do PE por parte dos sujeitos da amostra, i.e., 279 estudantes da Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do EB do INFORDEPE, oriundos de cada um dos 13 municípios de Timor-Leste, para os quais a LP não é LM. É importante referir que a seleção dos informantes obedeceu, apenas, a um critério: a frequência desta Licenciatura do INFORDEPE.

Quanto à tarefa propriamente dita, esta foi, à semelhança do que aconteceu com o questionário sociolinguístico, aplicada pelos professores do PFICP. Recolhidas as 279 produções escritas<sup>70</sup>, procedeu-se, nos 24.728 vocábulos que constituem os textos

---

<sup>70</sup> Dado o elevado número de produções escritas, não se procedeu à transcrição dos originais. No entanto, é importante salientar que muitos deles apresentam segmentos riscados e ilegíveis. Os segmentos riscados não foram contabilizados por considerar-se que se tratou de um lapso e que o informante não pretendeu incluí-lo na sua produção.

produzidos, ao levantamento de todas as palavras que continham grafemas que representam sibilantes. Neste conjunto de 1.043 palavras registaram-se 11.076 ocorrências de grafemas que representam as sibilantes (ocorrências quer divergentes quer convergentes com a norma gráfica europeia), pois existem várias palavras repetidas e muitas delas contêm mais do que uma sibilante.

Por forma a encontrar padrões de uso, calcular médias e construir as tabelas de registo de resultados, recorreu-se, para o tratamento da informação recolhida, ao programa *Excel* do *Microsoft Office*. Recorrendo a essa ferramenta, foi criada uma tabela para a organização dos dados com o objetivo de registar: (i) a posição silábica em que se registava a sibilante (coda ou ataque); (ii) a unidade fonológica alvo (/s, z, ʃ e ʒ/) e a respetiva realização fonética; (iii) a representação grafemática alvo e, por último, (iv) as representações gráficas registadas, conforme se pode verificar na tabela exemplificativa (imagem 2):

Forma gráfica alvo dos vocábulos com sibilantes ocorridas no corpus	Forma gráfica dos vocábulos com sibilantes registadas no corpus	Fonema/Fone alvo	Grafema correto	Representação desviante
Total de palavras				
aborrecido	aborrecido	/s/ - [s]	<c>	
académicas	académicas	/s/ - [ʃ]	<s>	
acessibilidade	acesibilidade	/s/ - [s]	<c>	
	acesibilidade	/s/ - [s]	<ss>	<s>
acesso	aceso	/s/ - [s]	<c>	
	aceso	/s/ - [s]	<ss>	<s>
acho	acho	/ʃ/ - [ʃ]	<ch>	

Imagem 2: Tratamento, em Excel, dos dados recolhidos (convergentes e divergentes)

Em seguida, apresenta-se uma outra tabela apenas com as ocorrências desviantes com o objetivo de identificar, também, as possibilidades descritivas/interpretativas dos desvios, conforme se pode verificar na imagem que se segue:



Forma gráfica alvo dos vocábulos com sibilantes ocorridas no corpus	Forma gráfica dos vocábulos com sibilantes registadas no corpus	Fonema/Fone alvo	Gratema correto	Representação desviante	Possibilidade descritiva/interpretativa			Outros
Total de palavras					Opacidade da ortografia portuguesa	Desconhecimento das regras contextuais de uso dos grafemas	Desconhecimento da identidade do segmento fonológico	
	aces <b>ibilidade</b>	/s/ - [s]	<ss>	<s>		x	x	
	aces <b>o</b>	/s/ - [s]	<ss>	<s>		x	x	
	a <b>ç</b> ho	/ʃ/ - [ʃ]	<ch>	<çh>				x
acrescenta	acrec <b>enta</b>	/s.s/ - [ʃs]	<sc>	<c>				x
acrescentar	acrec <b>entar</b>	/s.s/ - [ʃs]	<sc>	<c>				x
agressiva	agres <b>iva</b>	/s/ - [s]	<ss>	<s>		x	x	
agressivo	agres <b>ivo</b>	/s/ - [s]	<ss>	<s>		x	x	
ajudar	azud <b>ar</b>	/ʒ/ - [z]	<j>	<z>			x	
	aprend <b>is</b> agem	/z/ - [z]	<z>	<s>	x			

Imagem 3: Tratamento, em Excel, das ocorrências desviantes<sup>71</sup>

Do conjunto das ocorrências desviantes a considerar no estudo, excluíram-se todos os termos estrangeiros que não têm correspondência em português, como é o caso de *suco*<sup>72</sup>, assim como formas cuja segmentação vocábular não foi respeitada (*temser*; *porexemplo*; etc.) bem como neologismos não atestados no léxico do PE padrão (*planejador*; *controleação*; etc.), desde que evidenciassem uma representação gráfica das sibilantes plausível em português. Dito isto, foram incluídos na análise os casos de representação de sibilantes em neologismos como *profissionalidade* que, no restrito domínio da representação da sibilante, denotam uma representação gráfica desviante.<sup>73</sup>

<sup>71</sup> Para uma visão mais detalhada, cf. anexo 7.

<sup>72</sup> Cf. nota 62.

<sup>73</sup> Para uma visão mais pormenorizada dos casos de exclusão, cf. anexo 8.

---

## **3. Apresentação e discussão dos resultados**

---

Após uma análise global dos dados, procurar-se-á, neste subcapítulo, proceder à apresentação dos desvios encontrados na representação gráfica das consoantes sibilantes do português, por parte de aprendentes do ESUP timorense. Para tal, distribuir-se-ão os resultados, em função das ocorrências convergentes e divergentes na representação das sibilantes, atendendo ao município de origem e à LM do informante. Depois disso, organizar-se-ão os desvios ocorridos para, assim, poder verificar-se (i) qual a sibilante cuja representação gráfica se afigura mais problemática; (ii) qual a posição silábica em que ocorrem mais desvios e, (iii) se existe uma relação entre a LM do aprendente e o número de desvios.

### 3.1. Representação gráfica das consoantes sibilantes no *corpus* recolhido: ocorrências convergentes e divergentes

Uma análise global dos dados permite observar que, dos 279 textos recolhidos, 121 apresentam desvios na representação gráfica das consoantes sibilantes, o que representa 43% das produções.

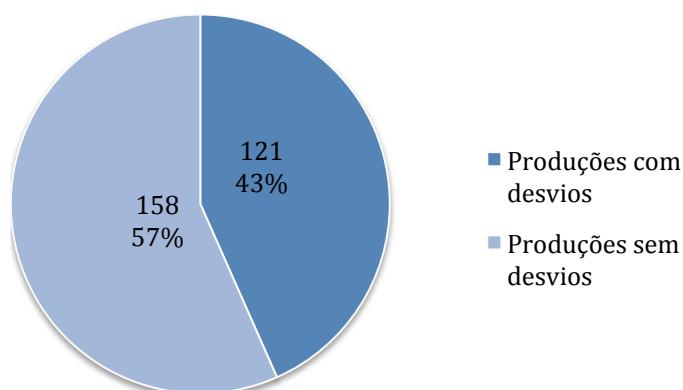


Gráfico 1: Distribuição das produções com e sem desvios na representação das sibilantes

Se for considerado o município de origem dos aprendentes, como é visível no gráfico 2, verifica-se que os informantes oriundos do enclave de Oe-Cusse são os que apresentam menor percentagem de textos com desvios (8%), enquanto que os naturais do distrito de Ermera são aqueles que geram uma maior percentagem de textos com desvios (68%). Saliente-se que, no caso do enclave de Oe-cusse, assiste-se à presença de “apenas” duas LM

distintas (tétum com 2 informantes e baikenu com 4) enquanto que, no caso do distrito de Ermera, verifica-se a presença de três (tétum com 14; mambae com 6 e kemak com 5).

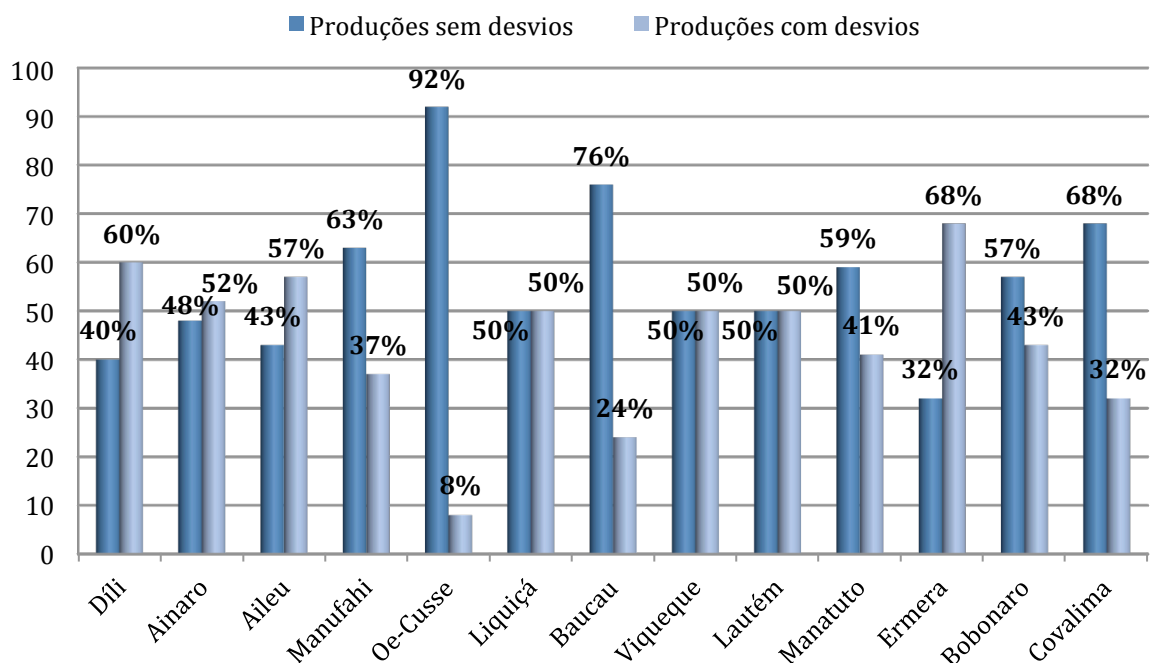


Gráfico 2: Percentagem de produções com e sem desvios por distrito de origem

Relativamente à LM dos aprendentes, e pelo facto de muitas delas terem uma presença muito residual no universo do estudo<sup>74</sup> (cf. anexo 4, questão 1), foram considerados apenas 2 grupos: os aprendentes cuja LM é o tétum e os aprendentes com outra LM. Assim, em termos percentuais, verifica-se algum equilíbrio entre os dois grupos, com o tétum a ser a LM de 116 dos jovens (41,5%) e as restantes 12 a serem a LM dos outros 163 (58,5%).

No que diz respeito aos resultados apresentados no gráfico 3, verifica-se que o desempenho é semelhante. Em 53% dos textos produzidos por aprendentes cuja LM é o tétum não se verificaram desvios na representação gráfica das sibilantes enquanto que, no outro grupo, não se verificaram desvios em 59% das produções.

<sup>74</sup> Cf. anexo 4, questão 1.

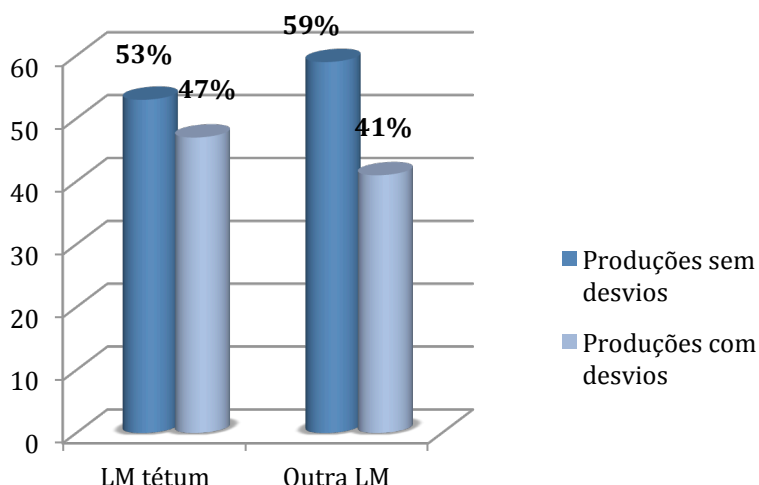


Gráfico 3: Percentagem de produções com e sem desvios na representação gráfica das sibilantes, por LM

Por forma a verificar a percentagem de representações desviantes em relação à norma-padrão do PE, procedeu-se, num primeira fase, à contabilização da totalidade das palavras com representação gráfica das sibilantes (excluindo as que são referidas no subcapítulo 2.3.).

Assim, e face à totalidade das sibilantes representadas, verificou-se que a percentagem de ocorrências divergentes é bastante reduzida, 2,01%, o que corresponde a 223 ocorrências desviantes num universo de 11.076 representações<sup>75</sup>.

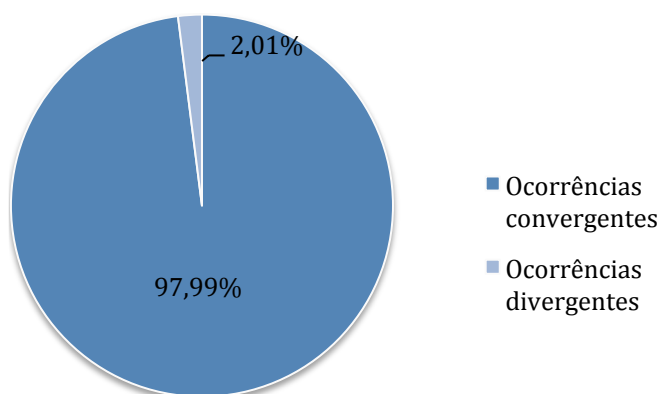


Gráfico 4: Distribuição da totalidade das ocorrências - convergentes e divergentes

<sup>75</sup> Canas (2014: 36), num estudo sobre a representação gráfica das sibilantes por aprendentes de português LS com diferentes LM (espanhol, italiano, alemão e chinês), concluiu que a percentagem de ocorrências desviantes face à totalidade das sibilantes representadas no seu *corpus* foi de apenas 0,81%. Este dado permite concluir que, no presente estudo, os aprendentes apresentam cerca de 2,5 vezes mais ocorrências desviantes.

### 3.1.1. Análise dos dados por município

As 279 produções que constituem o *corpus* de análise deste trabalho – constituído por um total de 11.076 ocorrências gráficas de sibilantes - permitiram a recolha de uma panóplia diversificada de desvios à norma ortográfica do português na representação gráfica das sibilantes. A tabela que se segue apresenta a síntese dos dados recolhidos.

Município	N.º de informantes	Total de palavras gráficas <sup>76</sup>	Média de palavras por informante/município	Total de ocorrências de sibilantes	Total de ocorrências com desvios na representação de sibilantes por município	Média da percentagem de desvios na representação de sibilantes por informante/município
Díli	15	1.465	97,67	653	20	3,06%
Ainaro	25	1.947	77,88	881	27	3,06%
Aileu	21	1.941	92,43	837	29	3,46%
Manufahi	24	2.027	84,46	921	14	1,52%
Oe-cusse	24	1.970	82,03	859	3	0,35%
Liquiçá	24	2.246	93,53	905	18	1,99%
Baucau	25	2.742	109,68	1304	6	0,46%
Viqueque	16	1.459	91,19	749	13	1,74%
Lautém	18	1.576	87,56	755	25	3,31%
Manatuto	17	1.837	108,06	890	12	1,35%
Ermera	25	2.106	84,24	940	28	2,98%
Bobonaro	23	1.730	75,22	734	13	1,77%
Covalima	22	1.682	76,45	719	15	2,09%
<b>TOTAL</b>	<b>279</b>	<b>24.728</b>	<b>88,63</b>	<b>11.076</b>	<b>223</b>	<b>2,01%</b>

Tabela 8: Síntese dos dados do *corpus* por município

A análise dos dados da tabela 8 permite concluir que Baucau é o município que regista produções escritas mais extensas (média de 110 palavras por texto) e que Bobonaro é o que apresenta textos de dimensões mais reduzidas (média de 75 palavras por produção).

Se considerarmos os 223 desvios, conclui-se que Aileu é o município que regista a percentagem de desvios mais elevada (3,46%) e que Oe-Cusse apresenta apenas a ocorrência de 3 desvios em 859 representações (0,35%)<sup>77</sup>.

<sup>76</sup> Para efeitos de contabilização, consideraram-se palavras gráficas todas as unidades linguísticas dotadas de sentido separadas por espaços brancos. Assim, em casos como *ensinar-vos*; *fazê-lo* ou *deve-se*, foi contabilizada apenas uma palavra gráfica.

<sup>77</sup> Nos resultados obtidos por Caetano (2014: 110) verifica-se, também, que os informantes do município de Oe-Cusse (professores do Ensino Secundário Técnico-vocacional) apresentam, em média, uma percentagem menos elevada de desvios nos exercícios de produção escrita, com 5,67% de desvios. Apesar de se tratar de um

### 3.1.2. Análise dos dados por LM

Como já foi referido anteriormente, pelo facto de muitas LM terem uma presença muito residual no universo da amostra (cf. quadro 2), foram considerados, para os propósitos deste estudo, apenas 2 grupos: os aprendentes cuja LM é o tétum e os aprendentes com outra LM. Neste âmbito, tentar-se-á, aqui, verificar o desempenho dos aprendentes tendo em consideração a respetiva LM. Para esse efeito, proceder-se-á à análise da distribuição da totalidade das ocorrências, convergentes e divergentes.

Em termos globais, verifica-se que os 2 grupos produzem uma percentagem muito idêntica de desvios: 1,99% para os aprendentes de LM tétum e 2,03% para os jovens com outra LM. A análise dos dados da tabela 9 permite concluir também que os informantes cuja LM é o tétum registam produções escritas ligeiramente mais extensas (média de 92 palavras por texto).

LM	N.º de informantes	Total de palavras gráficas	Média de palavras por informante / LM	Total de ocorrências de sibilantes	Total de ocorrências com desvios na representação de sibilantes por LM	Média da percentagem de desvios na representação de sibilantes por informante/ LM
TÉTUM	116	10.631	91,65	4.765	94	1,97%
OUTRA	163	14.097	86,48	6.311	129	2,04%
<b>TOTAL</b>	<b>279</b>	<b>24.728</b>	<b>88,63</b>	<b>11.076</b>	<b>223</b>	<b>2,01%</b>

Tabela 9: Síntese dos dados do *corpus* por LM

### 3.2. Distribuição dos desvios por possibilidade descritiva/interpretativa

Como já foi referido anteriormente, vários são os fatores que contribuem para a complexidade da ortografia do português e que colocam entraves à aprendizagem da escrita. Assim, podemos afirmar que há desvios que decorrem do facto de:

- a) uma unidade fónica poder ser representada, nos mesmos contextos, por grafemas diferentes. Por exemplo, em ataque (e especificamente em posição intervocálica), /s/ pode ser representado por <ss>, <ç> / <c> e, muito raramente, por <x>;

estudo que incide sobre todo o tipo de desvios, este dado pode ser interessante e motivar futuros estudos sobre a aquisição do português por parte de aprendentes cuja LM é o baiqueno.

- b) o aprendente não dominar regras de natureza contextual que distribuem as possibilidades ortográficas. Por exemplo, o grafema <ç>, em posição de ataque, só pode representar /s/ se a vogal seguinte (núcleo da sílaba) não for <e> ou <i>;
- c) o aprendente desconhecer a identidade fonológica do segmento sibilante (por exemplo, não saber se a unidade é [+] ou [-] [voz]).

Com o objetivo de delimitar claramente esses fatores e, assim, identificar o mais produtivo, foram delimitadas as seguintes possibilidades interpretativas:

1. Desvios motivados pela “opacidade” da ortografia portuguesa;
2. Desvios motivados pelo desconhecimento das regras contextuais de uso dos grafemas;
3. Desvios motivados pelo desconhecimento da identidade do segmento fonológico;
4. Desvios motivados pelo desconhecimento das regras contextuais de uso dos grafemas e/ou pelo desconhecimento da identidade do segmento fonológico (casos ambíguos)<sup>78</sup>;
5. Outros desvios<sup>79</sup>.

Como se pode verificar no quadro seguinte, a possibilidade de interpretação que originou o maior número de desvios foi a possibilidade 4 (desconhecimento das condições contextuais de uso dos grafemas e/ou da identidade do segmento fonológico), responsável por 66 desvios, o que corresponde a 29,5% do total. Por outro lado, o possível

---

<sup>78</sup> Esta categoria engloba os desvios cuja motivação decorre da impossibilidade de confirmar se se trata da situação 2 ou 3. Por exemplo, no caso da palavra *acesso*, se grafada *aceso*, colocam-se duas hipóteses: (i) o aprendente não evidencia domínio do segmento fonológico, que é /s/ e não /z/; e (ii) o aprendente até sabe qual a estrutura fonológica da palavra alvo, mas desconhece que, em posição intervocálica, o grafema <s> vale /z/ (= [z]).

<sup>79</sup> Nesta categoria englobam-se todos os desvios cuja motivação não se deve a nenhuma das 4 possibilidades anteriores. Assim, neste grupo, encontram-se, por exemplo, desvios relacionados com o desconhecimento da forma do dígrafo (*açho*); com a assimilação de segmentos fonológicos (*acrecenta*) ou com a epêntese de segmentos fonológicos (*pasciência*).



desconhecimento da identidade do segmento fonológico foi aquilo que causou menores dificuldades com “apenas” 14 desvios, o que equivale a 6,3% do total de ocorrências.

Possibilidade descritiva/interpretativa	N.º de ocorrências	Porcentagem
1. “Opacidade” da ortografia portuguesa	64	28,7%
2. Desconhecimento das regras contextuais de uso dos grafemas	57	25,6%
3. Desconhecimento da identidade do segmento fonológico	14	6,3%
4. Desconhecimento das regras contextuais de uso dos grafemas e/ou da identidade do segmento fonológico	66	29,5%
5. Outros fatores	22	9,9%
<b>Total</b>	<b>223</b>	<b>100%</b>

Tabela 10: Distribuição dos desvios por possibilidade descritiva/interpretativa

### 3.3. Distribuição dos desvios por posição silábica e por segmentos afetados

Depois de proceder ao levantamento da totalidade das ocorrências, convergentes e divergentes, procurou-se organizá-las por forma a verificar (i) a posição silábica em que se registava a sibilante; (ii) a unidade fonológica alvo e a respetiva realização fonética; (iii) a representação grafemática alvo e, por último, (iv) as representações gráficas convergentes e divergentes.

Desta forma, os desvios encontram-se organizados em 3 grandes classes correspondendo às posições silábicas de ataque e de coda e às sequências fonológicas que, pela sua especificidade, serão tratadas à parte (cf. anexo 6).

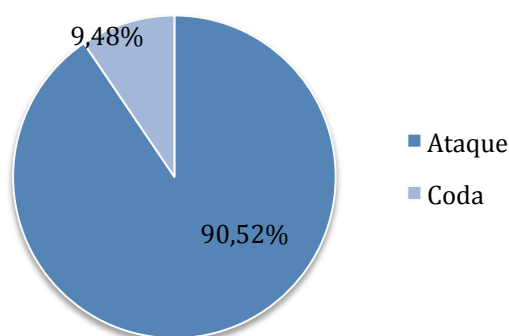


Gráfico 5: Distribuição dos desvios por posição silábica

Conforme se pode observar no gráfico 5, verifica-se que 90,52% do total dos desvios apresentados (ou seja, 191 de um total de 223 ocorrências) ocorrem em posição de ataque

silábico. Como já referido anteriormente, as unidades fonológicas, nesta posição silábica, assumem apenas uma realização fonética ([s, z, ʃ e ʒ]) mas várias representações gráficas.

Relativamente ao fone cuja representação gráfica gera maiores dificuldades, verifica-se que [s] é o responsável por um número mais elevado de desvios. Conforme se pode observar no gráfico 6, 74,41% dos desvios apresentados (o que corresponde a 157 desvios) ocorrem na representação desta sibilante apicodental surda<sup>80</sup>.

O gráfico mostra-nos também que é na representação gráfica da palatal sonora [ʒ] que os aprendentes apresentam menos dificuldades (1,90%, o equivalente a 4 desvios)<sup>81</sup>.

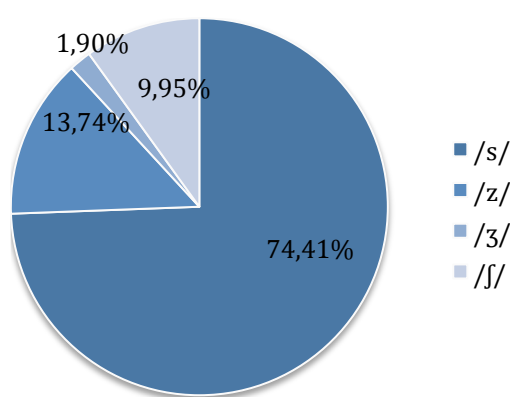


Gráfico 6: Distribuição dos desvios em função do segmento alvo

Observem-se, nas secções seguintes, e com maior detalhe, os resultados obtidos por posição silábica.

### 3.3.1. Ataque

Como se pode verificar na tabela 11, a representação gráfica dos segmentos fonológicos em posição de ataque silábico originou 191 desvios (3,06%), num total de 6.237 representações.

<sup>80</sup> Também no estudo de Canas (2014) se verifica- que a representação gráfica do fonema /s/ é a que causa maiores dificuldades. Verifica-se, no entanto, que, no caso dos estudantes timorenses, este desvio assume uma proporção ainda mais elevada (74,41% contra 52%).

<sup>81</sup> Nos dados do estudo de Canas (2014), verifica-se que é a sibilante palatal surda [ʃ] a que causa menos dificuldades aos aprendentes, com 3% dos desvios.

Posição silábica	Unidade fonológica alvo	Realização fónica alvo	Representação correta (grafema)	Ocorrências		Porcentagem	
				Convergentes	Divergentes	Ocorrências Convergentes	Ocorrências Divergentes
Ataque	/s/	[s]	<ç>	376	40	90,38	<b>9,62%</b>
			<ss>	1.737	24	98,64	1,36%
			<s>	2.063	7	99,66	0,34%
			<c>	813	86	90,43	9,57%
			<x>	23	0	<b>100%</b>	0%
	/z/	[z]	<z>	233	5	97,9%	2,1%
			<s>	412	24	94,50	5,50%
			<x>	69	0	<b>100%</b>	0%
	/ʒ/	[ʒ]	<g>	178	3	98,34	1,66%
			<j>	107	1	99,07	0,93%
			<ch>	17	1	94,44	5,56%
	/ʃ/	[ʃ]	<x>	18	0	<b>100%</b>	0%
<b>TOTAL</b>				<b>6.046</b>	<b>191</b>	<b>96,94</b>	<b>3,06%</b>

Tabela 11: Distribuição das ocorrências convergentes e divergentes em posição de ataque silábico

Se atentarmos às unidades afetadas em posição de ataque silábico, verifica-se que a representação gráfica da unidade fónica [s] é aquela que, incontestavelmente, causa maiores dificuldades, com 82,2% da totalidade dos desvios registados nesta posição silábica.

O gráfico 7 ilustra os resultados por unidade afetada em posição de ataque.

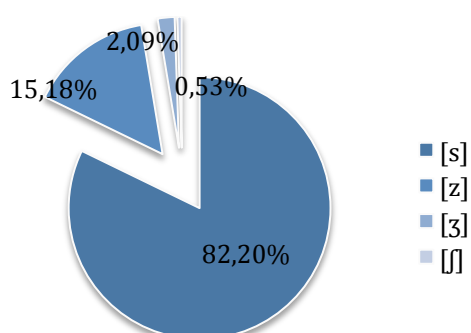


Gráfico 7: Distribuição dos desvios em posição de ataque silábico em função da unidade alvo

### 3.3.1.1. Desvios na representação gráfica do segmento fonológico /s/

Conforme referido anteriormente, o segmento fonológico /s/, na norma padrão do PE, admite, em posição de ataque silábico, apenas uma realização fonética ([s]), mas várias representações gráficas (<s>, <ss>, <ç>, <c>, <x>). Esta relação múltipla entre fonema e

grafema, gera, obviamente, dificuldades entre os aprendentes e origina desvios como aqueles que encontramos na tabela 12.

Neste âmbito, verificou-se que, dos 5 grafemas alvo possíveis na representação gráfica de /s/, 4 geraram ocorrências desviantes. Os desvios ocorreram em casos nos quais deveriam surgir os grafemas <c> (54,77%), <ç> (25,48%) e <s> (4,46%), e o dígrafo <ss> (15,49%). Curiosamente, o grafema <x>, tendo em conta os valores que pode assumir em ataque silábico, foi sempre grafado corretamente (23 ocorrências)<sup>82</sup>. Verifica-se, simultaneamente, que o grafema que gerou uma maior diversidade de desvios foi o grafema alvo <c>, com 4 registos distintos: <s>; <ç>; <sc> e <ss>.

Unidade fónica alvo – /s/ = [s]					
Grafema alvo	Número e percentagem de ocorrências desviantes	Grafema/ dígrafo desviante	Forma alvo	Forma desviante	Número de ocorrências
<ç>	40 (25,48%)	<c>	atenção	atenção	1
			avaliações	avaliacões	1
			começa	comeca	1
			começando	comecando	1
			comunicação	comunicação	2
			cooperação	cooperacão	1
			crianças	criancas	8
			dedicação	dedicacão	1
			educação	educacão	1
			esforçar	esforcar	2
			explicação	explicacão	3
			formação	formacão	1
			função	funcão	3
			informações	informaco	1
			nação	nacão	1
			organização	organização	1
			planificação	planificacão	3
			prequiosas	prequicosas	1
			preocupação	preocupacão	3
			relação	relacão	2
situações	situacões	1			
<ss>	24 (15,29%)	<s>	prequioso	prequioso	1
			acessibilidade	acesibilidade	1
			acesso	aceso	1
			agressiva	agresiva	1
			agressivo	agresivo	1
			assiduidade	asissuidade	2
			assiduidade	asesuidade	1
			interesse	interrese	2
			processo	proceso	1
			professor	profesor	3
			profissão	profisaun	1
			profissão	profisão	3
profissional	profisional	2			
sucesso	suceso	1			

<sup>82</sup> Este resultado poderá dever-se ao seu uso menos comum, em léxico restrito: cf. *auxílio, máximo, próximo, trouxe* (Martins e Festas, 2012).

		<c>	<i>interesse</i>	<i>interece</i>	1
			<i>profissional</i>	<i>profecional</i>	1
			<i>profissional</i>	<i>proficional</i>	1
		<ç>	<i>posso</i>	<i>poco</i>	1
<s>	7 (4,46%)	<c>	<i>ensinar</i>	<i>encinar</i>	2
			<i>sincero</i>	<i>cincerio</i>	2
		<ç>	<i>ensinar</i>	<i>ençinar</i>	1
			<i>sério</i>	<i>cerio</i>	1
		<ss>	<i>ensinar</i>	<i>enssinar</i>	1
<c>	86 (54,77%)	<s>	<i>capacidade</i>	<i>capasidade</i>	2
			<i>capacitar</i>	<i>capasitar</i>	1
			<i>cedo</i>	<i>sedo</i>	1
			<i>cidadão</i>	<i>sidadaun</i>	1
			<i>cidadãos</i>	<i>sidadãos</i>	1
			<i>cientistas</i>	<i>sientistas</i>	1
			<i>dicionário</i>	<i>disionário</i>	1
			<i>fácil</i>	<i>fasil</i>	1
			<i>licenciado</i>	<i>lisenciado</i>	1
			<i>nacional</i>	<i>nasional</i>	1
			<i>paciência</i>	<i>paciensia</i>	2
				<i>pasciênsia</i>	1
				<i>pasiencia</i>	7
				<i>pasiência</i>	1
				<i>pasiencia</i>	3
		<i>pasiênsia</i>		2	
		<i>pasiênsia</i>		2	
		<i>pasiensia</i>		6	
		<i>pasiensia</i>		6	
		<i>pasencia</i>		2	
		<i>pasência</i>	1		
		<i>precisa</i>	<i>presica</i>	2	
			<i>presisa</i>	2	
		<i>precisamos</i>	<i>presiza</i>	1	
			<i>presijamos</i>	1	
		<i>preciso</i>	<i>presisamos</i>	1	
			<i>persiso</i>	1	
			<i>persizo</i>	3	
			<i>presico</i>	1	
			<i>presico</i>	1	
		<i>principalmente</i>	<i>presiso</i>	4	
			<i>prinsipalmente</i>	1	
		<ç>	<i>pacência</i>	1	
			<i>pacência</i>	6	
			<i>pasiencia</i>	3	
			<i>passencia</i>	1	
		<i>precisa</i>	<i>preçisa</i>	3	
			<i>precisamos</i>	1	
		<sc>	<i>paciência</i>	3	
			<i>pasciência</i>	2	
		<ss>	<i>pasciênsia</i>	1	
			<i>capacidade</i>	<i>capassidade</i>	2
			<i>paciência</i>	<i>passencia</i>	1
<b>TOTAL</b>					<b>157 (100%)</b>

Tabela 12: Distribuição das representações gráficas desviantes de /s/

Como já foi referido anteriormente, na variedade acroletal do tétum<sup>83</sup>, o fonema /s/ apenas admite a realização [s] e a representação gráfica <s><sup>84</sup> nesta posição silábica,

<sup>83</sup> Cf. nota 35.

<sup>84</sup> Cf. tabela 4.

enquanto que, em português, admite as representações <ç>, <ss>, <s>, <c> e <x>, não sendo, por isso, de estranhar que esta apicodental surda tenha sido o segmento que mais desvios gerou.

Tendo em conta, por um lado, que <c>, <ç> e <ss> não constam, no tétum, dos grafemas representativos do segmento /s/, e que, por outro, a utilização destes grafemas em português requer o conhecimento de algumas regras contextuais, os motivos que levaram a estas representações desviantes poderão estar relacionados com o efeito cumulativo de estes dois fatores. Aliás, as coincidências lexicais entre as duas línguas e a maior simplicidade e transparência das relações fonema-grafema do tétum neste caso específico, potenciam a transferência da solução gráfica do tétum para a escrita da língua alvo (LA), como se pode verificar através de vocábulos como *preguisoso*, *fásil* e *nasional*.

Também nesta posição silábica se verificam casos de desvios que aparentam indiciar o desconhecimento da estrutura fonológica da palavra da LA ou das regras contextuais específicas em que determinado grafema adquire um dado valor fónico. Casos como os desvios em vocábulos como *atenção*, *avaliações* ou *começa* mostram isso mesmo, visto que, em português, o grafema <c> só pode assumir o valor de /s/ quando seguido dos grafemas vocálicos <e> ou <i>. Registam-se também casos em que o grafema <ç> é utilizado em vez de <s>, <ss> e <c>, como é o caso dos vocábulos *ençinar*, *poço* e *preçisa*, que parecem dever-se à opacidade da ortografia do português.

### 3.3.1.2. Desvios na representação gráfica do segmento fonológico /z/

Como se pode verificar no gráfico 7, o segmento fonológico /z/ é aquele que, depois de /s/, causa maiores dificuldades em posição de ataque silábico, com 15,18% dos desvios a ocorrerem na representação desta apicodental sonora. Tal como com o segmento /s/, também o fonema /z/ se realiza, em posição de ataque silábico, através de uma única realização fonética, neste caso [z], admitindo, contudo, 3 representações gráficas (<z>, <s> e <x>).

Neste âmbito, verificou-se que, dos 3 grafemas alvo possíveis na representação gráfica de [z], e cuja correta utilização está documentada no *corpus*, 2 geraram ocorrências

desviantes. Os desvios ocorreram maioritariamente onde deveria constar o grafema <s> (82,76%) e, com menor frequência, onde o segmento deveria ser grafado com <z> (17,24%). À semelhança do que acontece com o segmento fonológico /s/, também para grafar /z/ o grafema <x> foi sempre grafado corretamente (69 ocorrências)<sup>85</sup>.

A tabela que se segue ilustra os desvios encontrados na representação deste segmento.

Unidade fónica alvo – /z/ = [z]					
Grafema alvo	Número e percentagem de ocorrências desviantes	Grafema/dígrafo desviante	Forma alvo	Forma desviante	Número de ocorrências
<z>	5 (17,24%)	<s>	<i>aprendizaem</i>	<i>aprendisaem</i>	3
			<i>clareza</i>	<i>claresa</i>	1
			<i>responsabilizar</i>	<i>responsabilisar</i>	1
<s>	24 (82,76%)	<j>	<i>atrasado</i>	<i>atroiado</i>	1
			<i>desejo</i>	<i>dejejo</i>	1
			<i>precisa</i>	<i>precija</i>	3
			<i>precisamos</i>	<i>presijamos</i>	1
		<c>	<i>precisa</i>	<i>presica</i>	2
			<i>preciso</i>	<i>presico</i>	1
		<ç>	<i>preciso</i>	<i>presico</i>	1
		<ss>	<i>posição</i>	<i>possicão</i>	1
			<i>precisa</i>	<i>precissa</i>	1
			<i>preciso</i>	<i>precisso</i>	3
			<i>rigorosa</i>	<i>riqorossa</i>	1
		<z>	<i>Brasil</i>	<i>Brazil</i>	1
			<i>desenvolvimento</i>	<i>dezenvolvimento</i>	1
			<i>fase</i>	<i>faze</i>	1
			<i>precisa</i>	<i>presiza</i>	1
<i>precisar</i>	<i>precizar</i>		1		
		<i>preciso</i>	<i>persizo</i>	3	
<b>TOTAL</b>				<b>29 (100%)</b>	

Tabela 13: Distribuição das representações gráficas desviantes de [z]

Como referido anteriormente, no tétum, o fonema /z/ admite apenas a realização [z] e a representação gráfica <z><sup>86</sup>, não sendo, por isso, de estranhar que esta apicodental sonora tenha sido, depois de /s/, o segmento que mais desvios gerou (cf. gráfico 7).

Vejamos, por exemplo, os casos de *presica* e *presico*. Este desvios parecem evidenciar o desconhecimento do segmento fonológico (que é /z/ e não /k/, no primeiro caso, ou /s/ no

<sup>85</sup> Tal situação poderá dever-se ao facto de <x> assumir a realização fonética de [z] apenas em início de palavras iniciadas por <e> e quando seguido de vogal (à exceção de palavras que admitem mais do que uma pronúncia como é o caso de exógeno) como acontece, por exemplo, com *exame*, *exercer*, *existir*, *exorcismo*, *exuberante*. (Martins e Festas, 2012: 9)

<sup>86</sup> Cf. tabela 4.

segundo). O mesmo se verifica nos casos de *atrajado*, *dejejo*, *precija* e *precijamos*, que indiciam o desconhecimento da forma fonológica da palavra na LA e/ou a transferência da LM, pois o tétum, na sua variedade mesoletal, pode motivar o grafema <j> na representação de /z/, como é o caso de /a'zuda/ para <ajuda>.<sup>87</sup>

Já os casos de *aprendisagem*, *claresa* e *responsabilisar* parecem estar relacionados com a opacidade da ortografia do português. Esta opacidade na representação gráfica destes fonemas poderá ter levado estes jovens a recorrerem ao conhecimento gráfico que têm do tétum, nomeadamente em desvios como *dezenvolvimento*, *preciza* e *Brazil*.

### 3.3.1.3. Desvios na representação gráfica do segmento fonológico /ʃ/

O gráfico 7 evidencia que a representação do segmento fonológico /ʃ/ é a que, em posição de ataque silábico, causa menores dificuldades, com apenas 1 desvio a ocorrer nas 36 representações desta consoante palatal surda, ou seja, em 0,54% dos casos.

Como se viu anteriormente, e tal como com os restantes segmentos sibilantes, o fonema /ʃ/ apenas admite, em posição de ataque silábico, uma realização fonética: [ʃ]. No entanto, são possíveis 2 representações gráficas, <ch> e <x>.

Relativamente às ocorrências desviantes, a tabela que se segue mostra o único desvio encontrado na representação desta unidade.

Unidade fónica alvo – /ʃ/ = [ʃ]					
Grafema alvo	Número e percentagem de ocorrências desviantes	Grafema/dígrafo desviante	Forma alvo	Forma desviante	Número de ocorrências
<ch>	1 (100%)	<çh>	<i>acho</i>	<i>açho</i>	1
TOTAL				1 (100%)	

Tabela 14: Distribuição das representações gráficas desviantes de [ʃ]

Tendo em conta que <ch> não consta, no tétum, dos dígrafos representativos do segmento fonológico /ʃ/, e que, como referido anteriormente, este segmento apenas

<sup>87</sup> Cf. nota 35.



admite, nesta posição silábica, a representação gráfica <x><sup>88</sup>, talvez fosse de esperar um número de desvios mais elevado. No entanto, verificou-se apenas um desvio na representação desta sibilante (*açho*). Ao que tudo indica, esta representação desviante poderá estar relacionada com a opacidade da ortografia do português ou, apenas, com alguma falta de concentração ou insegurança do informante MNT196, uma vez que o dígrafo <çh> também não existe na ortografia do português. É interessante referir, também, que não ocorreu qualquer desvio sempre que a palatal foi grafada com <x> (16 ocorrências).

### 3.3.1.4. Desvios na representação gráfica do segmento fonológico /ʒ/

Como se pode verificar no gráfico 7, 2,06% dos desvios em posição de ataque silábico ocorreram com a representação do segmento fonológico /ʒ/. Como já referido anteriormente, este fonema, nesta posição silábica, admite apenas uma realização fónica, [ʒ], e duas representações gráficas, <g> e <j>. Apesar de se tratar de uma das consoantes em análise cujo registo gráfico originou menos desvios, verifica-se que, nas 4 ocorrências desviantes, 75% resultam de ocorrências desviantes para grafar o <g> e os restantes 25% para grafar o <j>.

A tabela que se segue ilustra os desvios encontrados na representação deste segmento.

Unidade fónica alvo – /ʒ/ = [ʒ]					
Grafema alvo	Número e percentagem de ocorrências desviantes	Grafema desviante	Forma alvo	Forma desviante	Número de ocorrências
<g>	3 (75,00%)	<j>	<i>gerir</i>	<i>jerir</i>	1
			<i>sugestões</i>	<i>sujestões</i>	1
		<z>	<i>corrigir</i>	<i>corrizir</i>	1
<j>	1 (25,00%)	<z>	<i>ajudar</i>	<i>azudar</i>	1
<b>TOTAL</b>				<b>4 (100%)</b>	

Tabela 15: Distribuição das representações gráficas desviantes de [ʒ]

Estas representações desviantes parecem estar relacionadas com a opacidade da ortografia do português, nos casos de *jerir*, *sugestões*, e com o desconhecimento da

<sup>88</sup> Cf. tabela 4.

estrutura fonológica da LA nos casos de *corrizar* e *azudar*. Tendo, contudo, em conta que <g> não consta, no tétum, dos grafemas representativos do segmento /ʒ/, e que, como referido anteriormente, na variedade acroletal do tétum, o fonema /ʒ/ admite apenas a realização [ʒ] e a representação gráfica <j><sup>89</sup>, pode também, existir, nos casos de *jerir*, *sujestões*, um efeito transferência da grafia do tétum para a da LA, efeito esse eventualmente também operante nos casos de *corrizar* e *azudar*, pois, como já se referiu anteriormente, o tétum, na sua variedade mesoletal, pode motivar o grafema <z> na representação de /ʒ/ (como em <azuda> para /a'ʒuda/).

### 3.3.2. Coda

Como já se observou no gráfico 7, só 9,48% dos desvios apresentados ocorreram em coda silábica, sendo que a representação gráfica dos segmentos fonológicos nesta posição silábica originou 20 desvios (0,42%) num total de 4.801 representações.

Posição silábica	Unidade fonológica alvo	Realização fónica alvo	Representação gráfica correta	Ocorrências		Percentagem	
				Convergentes	Divergentes	Convergentes	Divergentes
Coda	/s/	[ʃ], [ʒ], [z]	<x>	57	5	91,94%	8,06%
			<s>	4693	13	99,72%	0,28%
			<z>	31	2	93,94%	6,06%
TOTAL				4781	20	99,58%	0,42%

Tabela 16: Distribuição das ocorrências convergentes e divergentes em posição de coda silábica

Ao inverso do que acontece em posição de ataque silábico, onde se assiste a uma correspondência unívoca fonema-fone, em coda verifica-se que o único segmento fonológico sibilante que pode surgir nesta posição, o fonema /s/, se atualiza através de 3 realizações fonéticas ([ʃ], [ʒ] e [z]), podendo ser representado graficamente por <s>, <z> e <x>, em função do contexto em que se encontra.

<sup>89</sup> Cf. tabela 4.

Depois de recolhidos todos os desvios em coda silábica, verifica-se que 13 surgem em final de palavra e que 7 ocorrem em posição de coda silábica interna (cf. tabela 17).

Coda - Unidades fónicas alvo – [ʃ], [ʒ] e [z]					
Grafema alvo	Número e percentagem de ocorrências desviantes	Grafema/dígrafo desviante	Forma alvo	Forma desviante	Número de ocorrências
<s>	13 (65,00%)	<z>	<i>através</i>	<i>atravez</i>	7
			<i>através</i>	<i>atravéz</i>	4
		<ss>	<i>postura</i>	<i>posstura</i>	1
		<x>	<i>teste</i>	<i>texte</i>	1
<x>	5 (25,00%)	<s>	<i>experiência</i>	<i>esperiência</i>	1
			<i>explica</i>	<i>esplica</i>	1
			<i>explicar</i>	<i>esplicar</i>	3
<z>	2 (10,00%)	<s>	<i>produz</i>	<i>produs</i>	1
			<i>vez</i>	<i>ves</i>	1
TOTAL				20 (100%)	

Tabela 17: Distribuição das representações gráficas desviantes de [ʃ], [ʒ] e [z], em coda silábica

Verifica-se, ainda, que o grafema alvo <s>, além de ter sido o que gerou um número mais elevado de desvios (13, o que corresponde a 65%), foi também aquele que originou uma maior diversidade de desvios, com 3 registos distintos: <z>, <ss> e <x>. Além disso, verifica-se que, em vez de <x> e <z>, nesta posição silábica, os aprendentes optam sempre, nestas ocorrências desviantes, pelo grafema <s>.

É interessante verificar que não foram registados quaisquer desvios sempre que a unidade fónica sibilante continha valor morfológico, pois os plurais nominais e adjetivais foram sempre grafados corretamente nas 2430 ocorrências. Este resultado sugere que a função morfológica associada ao grafema é um fator promotor de representações gráficas convergentes com as da LA.

Por fim, e por se tratar de desvios que não alteram a estrutura fónica das palavras, estas representações desviantes parecem estar relacionadas com a opacidade da ortografia do português. Em casos como *experiência*, *explica* e *produs*, os comportamentos desviantes também poderão estar relacionados com a transferência gráfica da língua tétum (*esperiência*, *esplika* e *prodús*), admitindo-se, assim, e mais uma vez, um efeito cumulativo de fatores promotores do desvio.

### 3.3.3. Sequências

Pela sua especificidade, e pelo facto de, muitas vezes, não ser possível apurar, com rigor, qual o segmento fonológico afetado, nem a respetiva posição silábica, os desvios na representação da sequência fonológica /s.s/ foram tratados à parte.

Como se pode verificar na tabela que se segue, os dados recolhidos mostram que os aprendentes manifestaram muitas dificuldades em representar esta sequência fonológica, pois verificam-se 12 desvios de representação em 47 ocorrências, o que corresponde a uma percentagem de 25,53%.

Sequência fonológica alvo (coda/ataque)	Sequência fónica alvo	Representação gráfica correta	Ocorrências		Percentagem	
			Convergentes	Divergentes	Convergentes	Divergentes
/s.s/	[[s]	<sc>	35	12	74,47%	25,53%
TOTAL			35	12	74,47%	25,53%

Tabela 18: Distribuição das ocorrências convergentes e divergentes de /s.s/

Depois de recolhidos os desvios, verifica-se, de acordo com a tabela 19, que o grafema <c> surge em nove ocorrências para representar esta sequência, <s> ocorre em duas e o dígrafo <ss> regista-se apenas em uma.

Sequência fónica alvo – [[s]					
Sequência gráfica alvo	Número e percentagem de ocorrências desviantes	Grafema/dígrafo desviante	Forma alvo	Forma desviante	Número de ocorrências
<sc>	12 (100%)	<c>	<i>acrescenta</i>	<i>acrecenta</i>	1
			<i>acrescentar</i>	<i>acrecentar</i>	1
			<i>consciência</i>	<i>conciencia</i>	1
			<i>consciência</i>	<i>conciência</i>	2
			<i>conscientes</i>	<i>concientes</i>	1
			<i>disciplina</i>	<i>diciplina</i>	3
		<s>	<i>consciência</i>	<i>consência</i>	1
			<i>disciplinado</i>	<i>disiplinado</i>	1
		<ss>	<i>disciplina</i>	<i>dissiplina</i>	1
TOTAL				12 (100%)	

Tabela 19: Distribuição das representações gráficas desviantes de [[s]

Estes desvios revelam um claro desconhecimento de que se trata, em português, de uma sequência fonológica. Se, para além disso, tivermos em conta que, no tétum, os vocábulos *disciplina*<sup>90</sup> e *consciente* são grafados *dixiplina* e *konxiente*<sup>91</sup>, podemos admitir que o processo de simplificação fonológica transposto para a escrita tenha uma motivação acrescida, sendo igualmente expressão de um efeito de transferência fonológica do tétum.

---

<sup>90</sup> É curioso verificar, certamente devido a uma maior exposição a este vocábulo, que a palavra *disciplina(s)*, sempre que foi utilizada com o significado de matéria escolar, foi grafada corretamente. Isto confirma o interesse que pode ter a observação dos hábitos de uso do português e da familiarização com o texto escrito.

<sup>91</sup> *Disionáriu Nasional ba Tetun Ofisiál*, PP. 135 e 447.

---

## Conclusão

---

O presente estudo permite concluir que a amostra de 279 aprendentes do Ensino Superior leste-timorense possui, de uma forma geral, muitas fragilidades na representação gráfica das sibilantes.

Embora os desvios não sejam, do ponto de vista quantitativo, consideráveis (representam um total de 2,01% da totalidade das ocorrências), este estudo permite-nos verificar que esta área da ortografia do português é, também em Timor-Leste, uma área crítica.

No primeiro capítulo deste trabalho, procurámos descrever o panorama linguístico de Timor-Leste e identificar o papel que a língua portuguesa desempenha neste contexto multilingue. Concluiu-se que, apesar de ser uma das LO do país e a língua do estado, o português assume o estatuto de LM de um número muito reduzido de elementos da população. Na verdade, o português é língua segunda (LS) para a maioria dos timorenses que o falam. Este idioma é utilizado por uma pequena fração da população que, dependendo das fontes, varia entre os 5% e os 37% de falantes. Concluiu-se, também, que o tétum é a língua veicular, a língua indonésia é, essencialmente, utilizada nas transações comerciais e o inglês é uma língua de trabalho, nomeadamente nos contactos internacionais.

No primeiro capítulo, descreveram-se ainda as relações complexas entre unidades fónicas e gráficas no domínio das sibilantes da língua portuguesa. Referiu-se que, em posição inicial de sílaba (absoluta ou interior), é possível a ocorrência das quatro unidades fonológicas sibilantes (/s, z, ʃ, ʒ/), e que se regista, aí, o maior conjunto de grafemas para as representar (<ç, ss, s, c, x, z, g, j, ch>). Referiu-se, também, que, apesar de apenas o fonema /s/ poder ocupar a posição de coda silábica, a sua atualização é variável e contextualmente condicionada, podendo ser representado graficamente por <s, z, x>.

Paralelamente, descreveram-se as relações entre unidades fónicas e gráficas no domínio das sibilantes do tétum e constatou-se que essas relações são bem menos complexas. Em posição de ataque silábico, os grafemas representam biunivocamente um

fonema (<s, z, j, x> representam, respetivamente, /s, z, ʒ e ʃ/). Já em posição de coda, verifica-se que o único segmento fonológico aí possível pode ser representado por vários grafemas (/s/ é representado por <s> quando realiza [ʃ] e é representado por <z> quando realiza [ʒ]). Concluiu-se, por isso, que, de uma forma geral, a ortografia do tétum é mais fonética do que fonológica.

No segundo capítulo, foram apresentados os dados recolhidos através do questionário sociolinguístico, que permitiram traçar o perfil dos informantes. A análise destes dados permitiu observar que o tétum é a língua materna da maioria dos inquiridos (41,5%) e, conseqüentemente, a língua que melhor dominam (49,8%). Permitiu, ainda, confirmar que o tétum é a língua privilegiada para comunicar tanto em contextos formais como informais.

Ainda neste capítulo, procedeu-se à descrição e justificação da metodologia adotada, apresentando-se os procedimentos de recolha do *corpus* onde foram contabilizadas, não só as ocorrências desviantes, mas também as convergentes.

No terceiro capítulo, analisaram-se os dados de acordo com os objetivos propostos na Introdução. A análise dos dados conduziu-nos às seguintes respostas:

- (i) Se atentarmos ao segmento cuja representação gráfica se afigurou mais problemática, verificamos que /s/ é o responsável por um número mais elevado de desvios, com 74,41% das ocorrências (o que corresponde a 157 desvios). Verificou-se, também, que é na representação gráfica do segmento fonológico /ʒ/ que os aprendentes apresentaram menos dificuldades (1,90%, o equivalente a 4 desvios).
- (ii) A representação gráfica das sibilantes em posição de ataque silábico foi a que originou maiores dificuldades, verificando-se que 90,52% do total dos desvios apresentados (ou seja, 191 de um total de 223 ocorrências) ocorreram nesta posição.
- (iii) Em termos globais, verificou-se que o grupo de aprendentes de LM tétum e o grupo de estudantes com outra LM produziram uma percentagem muito idêntica de desvios com 1,99% e 2,03%

respetivamente.

- (iv) Conclui-se que os aprendentes naturais do município de Aileu foram os que registaram a percentagem de desvios mais elevada (3,46%) e que os de Oe-Cusse apresentaram, apenas, a ocorrência de 3 desvios em 859 representações (0,35%);
- (v) Concluiu-se que o desconhecimento das condições contextuais de uso dos grafemas e/ou da identidade do segmento fonológico poderá explicar o maior número de desvios (66 ocorrências, o que corresponde a 29,5% do total).

Resta-nos salientar que todo os desvio ortográfico do aprendente deve incitar o professor a refletir sobre o sistema de escrita pois, é a partir dessa reflexão que se identificam as necessidades de encaminhamento dos educandos. No caso de Timor-Leste, os desvios ortográficos deverão ser, no futuro, objeto de atenção por parte dos pedagogos. Estudos deste tipo serão, por isso, necessários e úteis para o ensino da Língua Portuguesa neste país .

Perspetivando, neste momento, as implicações decorrentes deste estudo, espera-se que as conclusões retiradas deste trabalho possam originar outras perguntas de investigação.



---

## Referências

---

- Albuquerque, D. (2011). *O Português de Timor-Leste: contribuições para o estudo de uma variedade emergente*, in PAPIA 21 (1).
- Albuquerque, D. (2012). *Bilinguismo e Multilinguismo em Timor-Leste: Aquisição, Interação e Estudo de Caso*, in Revista PerCursos Linguísticos, vol. 2, n.º 6.
- Assembleia Constituinte de Timor-Leste (2002). *Constituição da República Democrática de Timor-Leste*, Díli.
- Baltazar, A. (2013). *Contribuições para a construção de materiais didáticos para o ensino do português, por professores timorenses, no ensino secundário técnico-vocacional de Timor-Leste*. Dissertação de Mestrado em Língua e Literatura Materna. Universidade de Coimbra. Disponível em: <https://estudogeral.sib.uc.pt/bitstream/10316/25198/1/tese%20de%20mestrado.pdf> (consultado a 12 de novembro de 2016).
- Barbeiro, L. F. (2008). *Ortografia – quadro-geral*. In Maria Helena Mira Mateus, Dulce Pereira e Glória Fischer (orgs.) *Diversidade linguística na escola portuguesa*. Lisboa: Fundação Calouste Gulbenkian; CD2. Disponível em <http://www.iltec.pt/divling/index.html> (consultado a 18 de novembro de 2016).
- Barroso, H. (2011). *Forma e Substância da Expressão da Língua Portuguesa*, Coimbra: Livraria Almedina.
- Batoréo, H. J. (2009). *A Língua Portuguesa em Timor: de que forma deve o ensino de Português adaptar-se às diferentes realidades nacionais*, in Estudos Linguísticos/*Linguistic Studies*, 4, Lisboa: Centro de Linguística da Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa.
- Batoréo, H. J. (2010). *Ensinar Português no Enquadramento Poliglóstico de Timor-Leste*, in Palavras, 37, Lisboa: Associação de Professores de Português, Primavera de 2010.
- Brito, R. H. P. (2010). *Temas para a compreensão do atual quadro linguístico de Timor-Leste*, in Ciências e Letras, n.º 48, Porto Alegre.

- Brito, R. H. P. & Martins, M. L. (2005). *Moçambique e Timor-Leste: onde também se fala o português*, in Actas do III SOPCOM, VI LUSOCOM e II IBÉRICO, vol. III.
- Caetano, A. M. (2014). *Desvios na representação grafemática em aprendentes de Português Língua Segunda. Análise de uma amostra do universo de Professores do Ensino Secundário Técnico-Vocacional de Timor-Leste*. Dissertação de Mestrado em Ciências da Linguagem. Universidade do Minho.
- Canas, A. (2014). *Representação gráfica das sibilantes por aprendentes de Português L2*. Dissertação de Mestrado em Português como Língua Estrangeira e Língua Segunda. Universidade de Coimbra. <https://estudogeral.sib.uc.pt/handle/10316/27563>
- Carvalho, M. J. A. (2001). *Panorama linguístico de Timor. Identidade regional, nacional e pessoal*, in *Timor Lorosa'e*. Camões: Revista de Letras e Culturas Lusófonas, n.º 14.
- Corte-Real, B. & Brito R. (2006). *Aspectos da política-lingüística de Timor-Leste: desvendando contracorrentes*. Texto apresentado no X Congresso Brasileiro de Língua Portuguesa, realizado em São Paulo pela Pontifícia Universidade Católica, entre 28 de Abril e 1 de Maio de 2004, pp. 123-131.
- Costa, L. (2005). *Línguas de Timor*, in Fernando Cristóvão (Dir. e Coord. de), Dicionário Temático da Lusofonia, Lisboa, Texto Editores.
- Correia, A. et al. (2005). *Disionáriu Nasionál ba Tetun Ofisiál*. Díli: Instituto Nacional de Línguística.
- Esperança, J. P. T. (2001). *Estudos de Linguística Timorese*, Aveiro, SUL – Associação de Cooperação para o Desenvolvimento.
- Fernandes, V. H. P. M. (2008). *Os erros ortográficos em língua materna dos alunos do Ensino Básico na perspectiva do trabalho curricular dos professores*, Dissertação apresentada à Universidade Portucalense Infante D. Henrique para a obtenção do Grau de Mestre em Administração e Planificação da Educação, Porto.
- Franco, M. et al. (2003). *Domínio da comunicação, linguagem e fala. Perturbações específicas de linguagem em contexto escolar – Fundamentos*, Lisboa: Ministério da Educação. <http://www.scribd.com/doc/2685039/comunic-fala11> (consultado a 12 de novembro de 2016).

- Hull, G. (1999). *Manual de Língua Tétum para Timor-Leste*. Díli: Sebastião Aparício da Silva Project.
- Hull, G. (2001). *Timor-Leste: Identidade, Língua e Política Educacional*. Díli: Instituto Camões.
- Hull, G. (2002). *The Languages of East Timor: Some Basic Facts*. Díli: Instituto Nacional de Linguística, Universidade Nacional de Timor Lorosa'e. Disponível em: [http://www.portphillip.vic.gov.au/default/CommunityGovernanceDocuments/The\\_Languages\\_of\\_East\\_Timor\\_Some\\_Basic\\_Facts.pdf](http://www.portphillip.vic.gov.au/default/CommunityGovernanceDocuments/The_Languages_of_East_Timor_Some_Basic_Facts.pdf) (consultado a 20 de janeiro de 2017).
- Hull, G. & Eccles, L. (2005). *Gramática da Língua Tétum* (traduzido por Idalina Morgado a partir da edição em língua inglesa, 2001), Lisboa: Lidel – Edições Técnicas.
- Jornal da República, 1ª SÉRIE, Nº 1, Decreto do Governo N.º 1/2004 de 1 de abril de 2004, *O padrão ortográfico da língua tétum*, República Democrática de Timor-Leste.
- Jornal da República, 1ª SÉRIE, Nº 43, Decreto-Lei N.º 21/2010 de 1 de dezembro de 2010, Ministério da Educação de Timor-Leste.
- Jornal da República, 1ª SÉRIE, Nº 46, Decreto-Lei N.º 23/2010 de 9 de dezembro de 2010, Ministério da Educação de Timor-Leste.
- Jornal da República, 1ª SÉRIE, Nº 3, Decreto-Lei N.º 4/2011 de 26 de janeiro de 2011, Ministério da Educação de Timor-Leste.
- Martins, C. (1994). *O desaparecimento do mirandês na cidade de Miranda do Douro: uma leitura dos estudos de filologia mirandesa de José Leite de Vasconcelos*. In: *Variação linguística no espaço, no tempo e na sociedade*. Lisboa: Associação Portuguesa de Linguística/ Ed. Colibri, 95-105. Disponível em <http://hdl.handle.net/10316/18270> (consultado a 21 de janeiro de 2017).
- Martins, C. & Festas, I. (2012). *Palavras irregulares em testes de leitura. Para uma revisão dos critérios de irregularidade grafema-fone em português*. In: A. Costa & I. Duarte (Ed.), *Nada na linguagem lhe é estranho. Estudos em homenagem a Isabel Hub Faria* (pp. 247-262). Porto: Edições Afrontamento.

- Mateus, M. H. M. *et al.* (2005). *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta.
- SSTL (2001). *The Survey of Sucos: initial analysis and implications for poverty reduction, Dili, East Timor Transitional Administration (ETTA), Asian Development Bank (ADB), World Bank (WB) and United Nations Development Program (UNDP)*.
- Stern, H. (1983). *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.
- Thomaz, L. F. (2002). *Babel Lorosa'e. O problema linguístico de Timor-Leste*. Lisboa: Instituto Camões.
- Veloso, J. (2005). *A língua na escrita e a escrita da língua. Algumas considerações gerais sobre transparência e opacidade fonémicas na escrita do português e outras questões. In Da investigação às práticas. Estudos de natureza Educacional; Escola Superior de Educação de Lisboa, Centro Interdisciplinar de Estudos Educacionais, Vol. VI, Nº 1, 2005, pp. 49-69. Disponível em <http://repositorio-aberto.up.pt/bitstream/10216/25333/2/joaovelosolingua000100296.pdf> (consultado a 23 de janeiro de 2017).*
- Veloso, J. (2007). *Da influência do conhecimento ortográfico sobre o conhecimento fonológico*. München: Lincom Europa. Disponível em <https://sites.google.com/site/joaomanuelveloso/publica%C3%A7%C3%B5es> (consultado a 23 de janeiro de 2017).

#### **Sites consultados**

- <http://www.anaaa.gov.tl>
- <http://www.cplp.org>
- <http://timor-leste.gov.tl>
- <http://www.ethnologue.com>
- <http://www.jornal.gov.tl>
- <http://www.statistics.gov.tl>
- <http://www.wikiwand.com>

<http://www.portaldalinguaportuguesa.org>

<http://www.infordepe.gov.tl>

<http://www.untl.edu.tl>

---

## Anexos

---

- Anexo 1: Questionário e exercício de produção escrita
- Anexo 2: Orientações para os professores aplicadores
- Anexo 3: Autorização dos informantes
- Anexo 4: Apresentação dos dados do questionário
- Anexo 5: Principais línguas de Timor-Leste
- Anexo 6: Distribuição das ocorrências convergentes e divergentes
- Anexo 7: Tratamento dos dados em *Excel*
- Anexo 8: Ocorrências desviantes excluídas da análise

**QUESTIONÁRIO [KESTIONARIU]**

Código: [kodigo:] \_\_\_\_\_

Este questionário é anónimo. Não deve escrever o seu nome em nenhum sítio.

[Kestionariu ida ne'e la tau naran, lalika hakerek ita nia naran iha fatin ruma.]

**Naturalidade (Distrito) :** [Fatin Moris (Distritu)]: \_\_\_\_\_

**Morada (Distrito):** [Hela Moris (Distritu)]: \_\_\_\_\_

**Sexo:** Masculino [Mane]  Feminino [Feto]  **Idade:** [Tinan]: \_\_\_\_\_

**ATENÇÃO:** Responda, com sinceridade, às questões que lhe são colocadas, assinalando, com um X, à exceção da questão 2, a resposta que considera mais adequada.

[ATENSAUN: Hatán ho loloos, lia-husuk sira tuirmai ne'e, marka ho X. Ba lia-husuk muak hatán tuir ida ne'ebe ita considera loos liu.]

1- Em que língua começou a comunicar no seio familiar? [Dalen ida ne'be ita hatu koalía iha uma?]

- Tétum; [Tetun;]
- Português; [Portugés;]
- Indonésio; [Indonéziu;]
- Inglês; [Inglés;]
- Outra. Qual? [Seluk. Ida ne'be?] \_\_\_\_\_

2- Que línguas já aprendeu durante a sua vida? Ordene-as, começando por aquelas que domina melhor. [Lia saída deit mak ita aprende ona iha ita nia moris? Hakerek tuir malu, hahú housi sira ne'be ita hatene liu.]

- 1.º \_\_\_\_\_
- 2.º \_\_\_\_\_
- 3.º \_\_\_\_\_
- 4.º \_\_\_\_\_
- 5.º \_\_\_\_\_

3- Qual a língua que privilegia para comunicar em ambientes informais? [Lian ida ne'ebe mak diak liu, hodi komunika iha fatin informal sira ka fatin baibain?]

- Tétum; [Tetun;]
- Português; [Portugés;]
- Indonésio; [Indonéziu;]
- Inglês; [Inglés;]

- Outra. Qual? [Seluk. Ida ne'be?] \_\_\_\_\_
- 4- Qual a língua que privilegia para comunicar em ambientes formais? [Dalen ida ne'ebe mak diak liu, atu hodi komunika iha fatin formál sira ka iha servisu fatin?]
- Tétum; [Tetun;]
- Português; [Portugés;]
- Indonésio; [Indonéziu;]
- Inglês; [Inglés;]
- Outra. Qual? [Seluk. Ida ne'be?] \_\_\_\_\_
- 5- Que língua considera mais importante para o seu futuro? [Lia saida mak importante liu ba ita boot nia moris iha futuro?]
- Tétum; [Tetun;]
- Português; [Portugés;]
- Indonésio; [Indonéziu;]
- Inglês; [Inglés;]
- Outra. Qual? [Seluk. Ida ne'be?] \_\_\_\_\_
- 6- Que diploma(s) de língua portuguesa possui? [Diploma dalen Portugés nian, ida ne'be mak ita iha ona?]
- Nenhum; [La iha;]
- Nível 1; [Nivel 1;]
- Nível 2; [Nivel 2;]
- Nível 3; [Nivel 3;]
- Preparação para o Bacharelato; [Preparasaun ba Baxarelatu;]
- Oficina de Língua; [Ofisina Dalen nia;]
- Oficina de especialização; [Ofisina Espesializasaun;]
- Outro. Qual? [Seluk. Ida ne'be?] \_\_\_\_\_
- 7- Em termos globais, como considera o seu domínio do português? [Iha termu global, oinsá ita domina Dalen Portugés?]
- Muito bom; [Diák liu;]
- Bom; [Diák;]
- Suficiente; [Ladún diák ;]
- Insuficiente; [La diák ;]
- Muito insuficiente; [La diák liu;]



8- Onde/Como aprendeu português? [*Ne 'be/Oinsá ita aprende dalen português?*]

- Na escola; [*Iha escola, ho mestre;*]
- Em casa; [*Iha uma, ho família;*]
- Através da leitura de jornais, livros, revistas...; [*Liu husi lê jornal, livru sira, revista sira...*]
- Através da rádio ou da televisão; [*Liu husi radiu ka no televisaun.*]
- De outra forma. Qual? [*Ho forma Seluk. Ida ne 'be?*] \_\_\_\_\_

9- Em que situações usa habitualmente o português? [*Iha situausaun ne 'ebé mak uza beibeik dalen português?*]

- Nas aulas; [*Iha aula sira;*]
- Com amigos e colegas; [*Ho amigo sira no colega sira;*]
- Em casa; [*Iha uma;*]
- Nos serviços públicos; (Banco, Hospital, Correios, por exemplo); [*Iha servisu públiku (Banku, Ospital, Koreiu, nudar ezemplu);*]
- Com colegas portugueses ou brasileiros; [*Ho colega sira Portugal no Brazil;*]
- Outro local. [*Fatin seluk*]: \_\_\_\_\_

10- Onde costuma ouvir pessoas falarem em português? [*Ne 'ebe rona ema makdalen na 'in Dalen Portugés?*]

- No mercado e nas lojas; [*Iha bazar no loja sira;*]
- Nos serviços públicos; [*Iha fatin servisu públiku;*]
- Na rua; [*Iha luron;*]
- Na rádio ou na televisão; [*Iha radiu eh televizaun;*]
- Nas aulas de Língua Portuguesa; [*Iha aula sira dalen português nian;*]
- Na igreja; [*Iha uma kreda;*]
- Outro local. [*Fatin seluk*]: \_\_\_\_\_

11- Em que situações costuma ler em português? [*Iha situausaun sira ne 'ebe mak ita toman lee Portugés?*]

- Na escola; [*Iha escola;*]
- Nos serviços públicos; [*Iha fatin servisu públiku;*]
- Na televisão; [*Iha televizaun;*]
- Em casa; [*Iha uma;*]
- Outro local. [*Fatin seluk*]: \_\_\_\_\_

12- Em que situações costuma escrever em português? [*Iha situasaun saída mak ita hakerek portugues?*]

- Na escola; [*Iha escola;*]
- Nos serviços públicos; [*Iha fatin servisu publiku;*]
- Em casa; [*Iha uma;*]
- Outro local. [*Fatin seluk*]: \_\_\_\_\_

13- Aprender português é importante porque: [*Aprende dalen português importante tamba:*]

- É uma das línguas oficiais de Timor-Leste; [*Tamba dalen ida ne'e halo parte mós dalen ofisiál Timor-Leste;*]
- Preciso dela para comunicar no dia a dia; [*Ha'u presiza nia atu hato'o lia loronloron;*]
- Preciso dela para continuar a estudar; [*Há'u presiza nia atu estuda ba oin;*]
- Pode servir para comunicar com pessoas de outros países; [*Bele sai mós hodi hato'o lia ba ema iha estranjeiru;*]
- Preciso dela para conseguir um bom trabalho; [*Ha'u presiza nia hato'o hetan servisu di'ak;*]
- Não é importante. [*La importante.*]

---

## **PRODUÇÃO ESCRITA**

Num máximo de 20 linhas, dá a tua opinião sobre como deve ser o professor ideal.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

Muito obrigado pela sua colaboração! [*Obrigadu wain ba ita boot nia tulun*]

**APLICAÇÃO DO QUESTIONÁRIO E DO EXERCÍCIO DE PRODUÇÃO ESCRITA****INSTRUÇÕES GERAIS PARA OS PROFESSORES APLICADORES**

Tendo em conta que a normalidade e a qualidade do processo de recolha dos dados é fundamental para a sua validação e para garantir o tratamento equitativo entre os informantes de todas as turmas, solicita-se que cada docente adote os seguintes procedimentos:

**Antes da entrega do exercício de produção escrita**

- 1- Cada envelope contém 25 questionários sociolinguísticos, 25 exemplares do exercício de produção escrita e 25 declarações de autorização para utilizar os dados recolhidos neste estudo. Deverão entregar, em primeiro lugar, o exercício de produção escrita sem fazer referência, nessa altura, ao facto de poder vir a ser utilizado num estudo;
- 2- Informar os alunos que dispõem de 30 minutos para a realização da tarefa, escrevendo, no quadro, a hora de início e término do exercício;
- 3- Informar os alunos que o exercício é anónimo e certificar-se que nenhum deles se identifica na folha do exercício;
- 4- Informar que não é permitido o uso do dicionário de Língua Portuguesa durante o exercício;
- 5- Entregar o enunciado, proceder à sua leitura e alertar para o cumprimento das indicações que constam dele;

**Depois do exercício de produção escrita**

- 1- Os enunciados deverão permanecer na mesa de cada aluno;
- 2- Entregar, a cada aluno, o questionário sociolinguístico, referindo o propósito do exercício e do questionário (recolha de dados para uma dissertação de Mestrado);
- 3- Tal como no exercício, informar os alunos que o questionário é anónimo e certificar-se que nenhum aluno se identifica;
- 4- Elucidar os alunos no caso de existir alguma dúvida no preenchimento do questionário;
- 5- Assegurar-se que os dados do cabeçalho foram preenchidos corretamente;

**Depois do questionário sociolinguístico**

- 1- Os enunciados do exercício e do questionário deverão permanecer na mesa de cada aluno;
- 2- Distribuir, a cada aluno, a declaração de autorização para a utilização dos dados recolhidos, explicando os propósitos do estudo;
- 3- No caso de algum aluno se recusar a assinar a declaração poderá inutilizar os enunciados que se encontram na sua mesa (o professor deverá registar o nome do aluno);
- 4- Para aqueles que concordarem assinar, deverá recolher o questionário e o exercício de cada aluno, agrafando-os um ao outro, certificando-se, novamente, que não há qualquer indicação que identifique o aluno;
- 5- Colocar os enunciados do exercício e do questionário num envelope e indicar o número de exemplares recolhidos.
- 6- Colocar as declarações num envelope separado e indicar o número de exemplares recolhidos, assegurando-se que o número seja idêntico ao do outro envelope.

Faculdade de Letras da Universidade de Coimbra  
Mestrado em Português Língua Estrangeira/Língua Segunda

---

## DECLARAÇÃO

Eu, \_\_\_\_\_, aluno(a) do polo de \_\_\_\_\_, da Licenciatura em Formação de Professores dos 1.º e 2.º Ciclos do Ensino Básico, no INFORDEPE, autorizo que os materiais por mim produzidos sejam utilizados para investigações realizadas da Dissertação de Mestrado de Rui Manuel Ferreira Pinto, no âmbito do Mestrado em Português Língua Estrangeira/Língua Segunda.

Data: \_\_\_/\_\_\_/2014

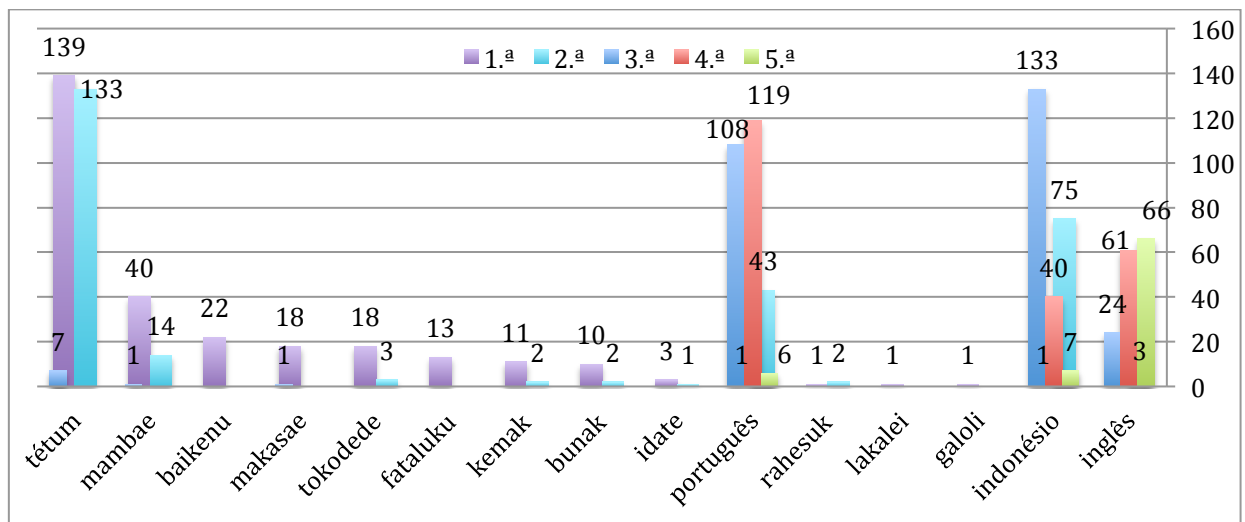
Assinatura: \_\_\_\_\_

**Questão n.º 1 - Em que língua começou a comunicar no seio familiar?**

N.º de informantes	LÍNGUA MATERNA													TOTAL
	tétum	mambae	baikeno	makasae	tukodede	fataluku	kemak	bunak	idaté	rahesuk	lolein	lakalei	galoli	
	116	53	22	19	18	16	13	12	4	3	1	1	1	279

Distribuição dos aprendentes da amostra por LM

**Questão n.º 2 - Que línguas já aprendeu durante a sua vida? Ordene-as, começando por aquelas que domina melhor.**



Distribuição dos aprendentes por níveis relativos de proficiência nas línguas conhecidas

**Questão n.º 3 – Qual a língua que privilegia para comunicar em contexto informal?**

Língua	tétum	português	mambae	bunak	kemak	rahesuk	NS/NR <sup>92</sup>	TOTAL
<b>N.º de informantes</b>	<b>257</b>	10	5	4	1	1	1	279
<b>Percentagem</b>	<b>92,1%</b>	3,6%	1,7%	1,4%	0,4%	0,4%	0,4%	100%

Língua dominante em contexto informal

**Questão n.º 4 – Qual a língua que privilegia para comunicar em contexto formal?**

Língua	tétum	português	indonésio	inglês	outra	NS/NR	TOTAL
<b>N.º de informantes</b>	<b>185</b>	93	0	0	0	1	279
<b>Percentagem</b>	<b>66,3%</b>	33,3%	0%	0%	0%	0,4%	100%

Uso das línguas em contexto formal

**Questão n.º 5 – Que língua considera a mais importante para o seu futuro?**

Língua	tétum	português	indonésio	inglês	outra	TOTAL
<b>N.º de informantes</b>	11	<b>268</b>	0	0	0	279
<b>Percentagem</b>	4%	<b>96%</b>	0%	0%	0%	100%

Relevância da língua para o futuro

<sup>92</sup> NS/NR: Não sabe ou não responde.

**Questão n.º 6 – Que diploma(s) de língua portuguesa possui?**

Diploma	Nenhum	Nível 1	Nível 2	Nível 3	Preparação para o Bacharelato	Oficina de Língua	Oficina de especialização	NS/NR	TOTAL
<b>N.º de informantes</b>	<b>163</b>	39	22	10	7	31	0	7	279
<b>Percentagem</b>	<b>58,5%</b>	14%	7,8%	3,6%	2,5%	11,1%	0%	2,5%	100%

Certificação do nível de língua portuguesa

**Questão n.º 7 – Em termos globais, como considera o seu domínio do português?**

Domínio	Muito Bom	Bom	Suficiente	Insuficiente	Muito Insuficiente	NS/NR	TOTAL
<b>N.º de informantes</b>	52	<b>163</b>	63	0	0	1	279
<b>Percentagem</b>	18,6%	<b>58,5%</b>	22,5%	0%	0%	0,4%	100%

Autoavaliação dos aprendentes em relação ao seu domínio do português

**Questão n.º 8 – Onde/Como aprendeu português?**

Local/forma	Escola	Casa	Jornais, livros e revistas	Rádio ou TV	Outro	TOTAL
<b>N.º de informantes</b>	<b>276</b>	1	2	0	0	279
<b>Percentagem</b>	<b>98,9%</b>	0,4%	0,7%	0%	0%	100%

Meios de acesso e de aprendizagem da/à língua portuguesa



**Questão n.º 9 – Em que situações usa, habitualmente, o português?**

Contexto	Nas aulas	Com amigos e colegas	Em casa	Nos serviços públicos	Com colegas portugueses ou brasileiros	Outra situação	TOTAL
<b>N.º de informantes</b>	<b>258</b>	10	0	3	8	0	279
<b>Percentagem</b>	<b>92,5%</b>	3,6%	0%	1,1%	2,9%	0%	100%

Contextos de uso do português

**Questão n.º 10 – Onde costuma ouvir pessoas a falar em português?**

Local	No mercado e nas lojas	Nos serviços públicos	Na rua	Na rádio ou na televisão	Nas aulas de língua portuguesa	Na igreja	Noutro local	NS/NR	TOTAL
<b>N.º de informantes</b>	0	11	0	33	<b>232</b>	0	0	3	279
<b>Percentage m</b>	0%	3,9%	0%	11,8%	<b>83,1%</b>	0%	0%	1,2%	100%

Contextos de exposição (auditiva) ao português

**Questão n.º 11 – Em que local costuma ler em português?**

Local	No escola	Nos serviços públicos	Na televisão	Em casa	Noutro local	NS/NR	TOTAL
<b>N.º de informantes</b>	267	1	1	7	0	3	279
<b>Percentagem</b>	96%	0,4%	0,4%	2,5%	0%	1,1%	100%

Distribuição dos aprendentes da amostra pelas situações em que leem em língua portuguesa

**Questão n.º 12 – Em que situações costuma escrever em português?**

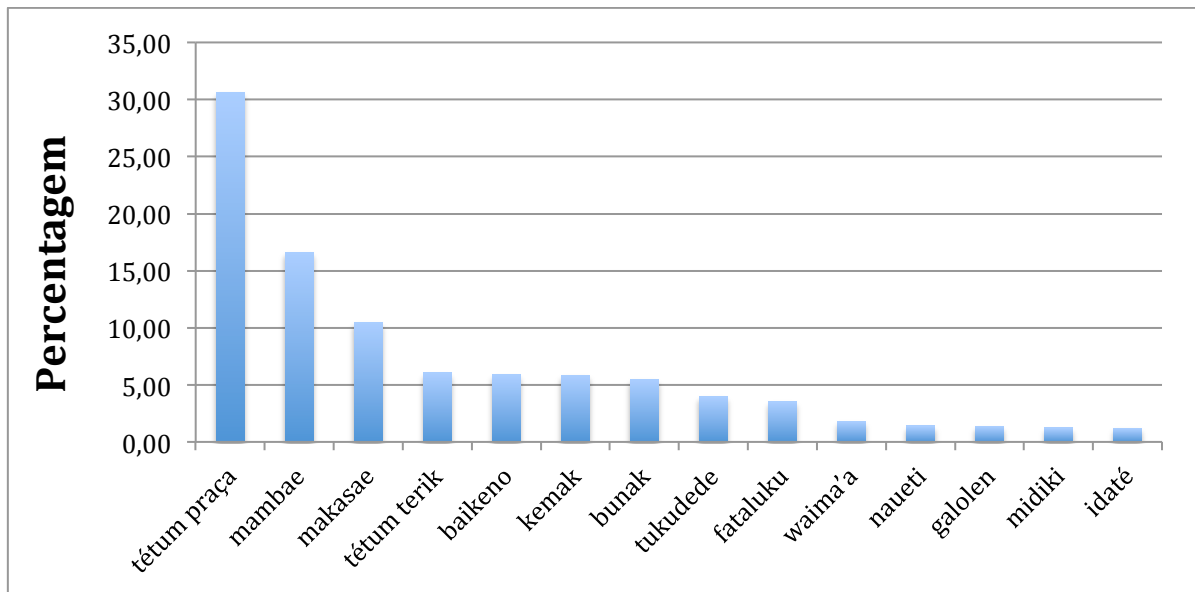
Local	Na escola	Nos serviços públicos	Em casa	Noutro local	TOTAL
<b>N.º de informantes</b>	276	3	0	0	279
<b>Percentagem</b>	98,9%	1,1%	0%	0%	100%

Escrever em português

**Questão n.º 13 – Aprender português é importante porque...**

Importância	Língua oficial	Comunicar no dia a dia	Estudar	Comunicar com estrangeiros	Conseguir trabalho	Não é importante	NS/NR	TOTAL
<b>N.º de informantes</b>	254	0	13	5	4	0	3	279
<b>Percentagem</b>	91%	0%	4,6%	1,8%	1,4%	0%	1,1%	100%

Distribuição dos aprendentes da amostra pela importância atribuída à aprendizagem do português



Principais línguas de Timor-Leste (% da população/LM) – Fonte: [http://www.wikiwand.com/pt/L%C3%ADnguas\\_de\\_Timor-Leste](http://www.wikiwand.com/pt/L%C3%ADnguas_de_Timor-Leste) (consultado a 18 de novembro de 2016).

Posição silábica	Unidade fonológica alvo	Realização fônica alvo	Representação correta (grafema)	Ocorrências		Porcentagem		
				Convergentes	Divergentes	Convergentes	Divergentes	
Ataque	/s/	[s]	<ç>	376	40	90,38%	9,62%	
			<ss>	1737	24	98,64%	1,36%	
			<s>	2063	7	99,66%	0,34%	
			<c>	813	86	90,43%	9,57%	
			<x>	23	0	100%	0%	
	/z/	[z]	<z>	233	5	97,9%	2,1%	
			<s>	412	24	94,50%	5,50%	
			<x>	69	0	100%	0%	
	/ʒ/	[ʒ]	<g>	178	3	98,34%	1,66%	
			<j>	107	1	99,07%	0,93%	
	/ʃ/	[ʃ]	<ch>	17	1	94,44%	5,56%	
			<x>	18	0	100%	0%	
	<b>TOTAL</b>				6046	191	<b>96,94%</b>	<b>3,06%</b>
	Coda	/s/	[ʃ], [ʒ], [z]	<x>	57	5	91,94%	8,06%
<s>				4693	13	99,72%	0,28%	
<z>				31	2	93,94%	6,06%	
<b>TOTAL</b>				4781	20	99,58%	0,42%	

Distribuição das ocorrências convergentes e divergentes em função da posição silábica, da unidade fonológica alvo, da unidade fônica alvo e da representação gráfica correta.

Sequência fonológica alvo (coda/ataque)	Sequência fónica alvo	Representação correta (grafemas)	Ocorrências		Porcentagem	
			Convergentes	Divergentes	Convergentes	Divergentes
/s.s/	[[s]	<sc>	35	12	74,47%	25,53%
<b>TOTAL</b>			35	12	74,47%	25,53%

Distribuição das ocorrências convergentes e divergentes em função da sequência fonológica alvo, da sequência fónica alvo e da representação gráfica correta.

Forma gráfica alvo dos vocábulos com sibilantes ocorridas no corpus	Forma gráfica dos vocábulos com sibilantes registadas no corpus	Fonema/Fone alvo	Grafema correto	Representação desviante	Possibilidade descritiva/interpretativa			Posição silábica		TOTAL DE OCORRÊNCIAS	
					Opacidade da ortografia portuguesa	Desconhecimento das regras contextuais de uso dos grafemas	Desconhecimento da identidade do segmento fonológico	Outros	ataque		coda
	Total de palavras									24728	
	acesibilidade	/s/ - [s]	<ss>	<ç>		x	x		x		1
	aceso	/s/ - [s]	<ss>	<ç>		x	x		x		1
	açho	/ʃ/ - [ʃ]	<ch>	<ç>				x desconhecimento da forma do dígrafo	x		1
acrescenta	acrecenta	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			1
acrescentar	acrecentar	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			1
agressiva	agresiva	/s/ - [s]	<ss>	<ç>		x	x		x		1
agressivo	agresivo	/s/ - [s]	<ss>	<ç>		x	x		x		1
ajudar	azudar	/ʒ/ - [ʒ]	<ç>	<ç>			x		x		1
	aprendisagem	/z/ - [z]	<z>	<s>	x				x		3
assiduidade	asissuidade							Desconhecimento da própria forma lexical			2
	asesuidade										1
	atencão	/s/ - [s]	<ç>	<c>		x			x		1
atrasado	atrajado	/z/ - [z]	<ç>	<ç>			x		x		1
	atravez	/ʒ/ - [ʒ]	<ç>	<z>	x					x	7
	atravéz	/ʒ/ - [ʒ]	<ç>	<z>	x				x		4
avaliações	avaliacões	/s/ - [s]	<ç>	<c>		x			x		1
Brasil	brazil	/z/ - [z]	<ç>	<z>	x				x		1
	capacidade	/s/ - [s]	<ç>	<s>		x	x		x		2
	capassidade	/s/ - [s]	<ç>	<ss>	x				x		2
	capasitar	/s/ - [s]	<ç>	<s>		x	x		x		1
cedo	sedo	/s/ - [s]	<ç>	<s>	x				x		1
	sidadaun	/s/ - [s]	<ç>	<s>	x				x		1
	sidadãos	/s/ - [s]	<ç>	<s>	x				x		1
cientistas	sientistas	/s/ - [s]	<ç>	<s>	x				x		1
clareza	claresa	/z/ - [z]	<z>	<s>	x				x		1
	comeca	/s/ - [s]	<ç>	<c>		x			x		1
começando	comecando	/s/ - [s]	<ç>	<c>		x			x		1
	comunicaçã	/s/ - [s]	<ç>	<c>		x	x		x		2
consciência	conciencia	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			1
	conciência	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			2
	consência	/s.s/ - [ʃs]	<sc>	<s>				x assimilação de segmentos fonológicos			1
conscientes	concientes	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			1
	cooperaçã	/s/ - [s]	<ç>	<c>		x			x		1
	corrizar	/ʒ/ - [ʒ]	<ç>	<z>			x		x		1
crianças	criancas	/s/ - [s]	<ç>	<c>		x			x		8
dedicação	dedicacão	/s/ - [s]	<ç>	<c>		x			x		1
desejo	dejejo	/z/ - [z]	<ç>	<ç>			x		x		1
	de desenvolviment	/z/ - [z]	<ç>	<z>	x				x		1
	disionário	/s/ - [s]	<ç>	<s>		x	x		x		1
disciplina	diciplina	/s.s/ - [ʃs]	<sc>	<c>				x assimilação de segmentos fonológicos			3
	dissiplina	/s.s/ - [ʃs]	<sc>	<ss>				x assimilação de segmentos fonológicos			1
	disiplinado	/s.s/ - [ʃs]	<sc>	<s>				x assimilação de segmentos fonológicos			1
educação	educaçã	/s/ - [s]	<ç>	<c>		x			x		1
ensinar	encinar	/s/ - [s]	<ç>	<c>	x				x		2
	ençinar	/s/ - [s]	<ç>	<ç>		x			x		2
	ensinar	/s/ - [s]	<ç>	<ss>		x			x		1
	esforçar	/s/ - [s]	<ç>	<c>		x			x		3
experiência	esperiência	/s/ - [ʃ]	<ç>	<s>	x					x	1
explica	esplica	/s/ - [ʃ]	<ç>	<s>	x					x	2
	explicacão	/s/ - [s]	<ç>	<c>		x			x		1
explicar	esplicar	/s/ - [ʃ]	<ç>	<s>	x					x	2
	fasil	/s/ - [s]	<ç>	<s>		x	x		x		1
fase	faze	/z/ - [z]	<ç>	<z>	x				x		1
formação	formacão	/s/ - [s]	<ç>	<c>		x			x		3
função	função	/s/ - [s]	<ç>	<c>		x			x		1
	jerir	/ʒ/ - [ʒ]	<ç>	<ç>	x				x		1

informações	informações	/s/ - [s]	<ç>	<ç>		x			x		1
interesse	interesse	/s/ - [s]	<ss>	<ç>	x				x		1
	interesse	/s/ - [s]	<ss>	<ç>		x	x		x		2
licenciado	licenciado	/s/ - [s]	<ç>	<ç>		x	x		x		1
nação	nação	/s/ - [s]	<ç>	<ç>		x			x		3
	nacional	/s/ - [s]	<ç>	<ç>		x	x				1
	organização	/s/ - [s]	<ç>	<ç>		x			x		1
paciência	paciência	/s/ - [s]	<ç>	<ç>		x			x		1
	paciência	/s/ - [s]	<ç>	<ç>		x			x		6
	paciência	/s/ - [s]	<ç>	<ç>	x				x		2
	paciência	/s/ - [s]	<ç>	<sc>					x epêntese de segmento fonológico		3
	paciência	/s/ - [s]	<ç>	<sc>					x epêntese de segmento fonológico		2
	paciência	/s/ - [s]	<ç>	<sc>					x epêntese de segmento fonológico		1
	paciência	/s/ - [s]	<ç>	<ç>	x				x		1
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		7
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		1
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		3
	paciência	/s/ - [s]	<ç>	<ç>		x			x		3
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		2
	paciência	/s/ - [s]	<ç>	<ç>	x				x		2
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		6
	paciência	/s/ - [s]	<ç>	<ç>	x				x		6
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		2
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		1
	paciência	/s/ - [s]	<ç>	<ss>	x				x		1
	paciência	/s/ - [s]	<ç>	<ç>		x			x		1
	paciência	/s/ - [s]	<ç>	<ç>		x	x		x		0
planificação	planificação	/s/ - [s]	<ç>	<ç>		x			x		3
posição	posição	/z/ - [z]	<ç>	<ss>			x		x		1
posso	poço	/s/ - [s]	<ss>	<ç>	x				x		1
	postura	/s/ - [j]	<ç>	<ss>		x			x		1
	precisa	/z/ - [z]	<ç>	<ç>			x		x		3
	precisa	/s/ - [s]	<ç>	<ç>		x			x		3
	precisa	/z/ - [z]	<ç>	<ss>			x		x		1
	precisa	/s/ - [s]	<ç>	<ç>		x	x		x		2
	precisa	/z/ - [z]	<ç>	<ç>		x	x		x		2
	precisa	/s/ - [s]	<ç>	<ç>		x	x		x		2
	precisa	/s/ - [s]	<ç>	<ç>		x	x		x		1
	precisa	/z/ - [z]	<ç>	<ç>	x				x		1
	precisamos	/s/ - [s]	<ç>	<ç>		x			x		1
	precisamos	/s/ - [s]	<ç>	<ç>		x	x		x		1
	precisamos	/z/ - [z]	<ç>	<ç>			x		x		1
	precisamos	/s/ - [s]	<ç>	<ç>		x	x		x		1
	precizar	/z/ - [z]	<ç>	<ç>	x				x		1
	persiso	/s/ - [s]	<ç>	<ç>	x				x		1
	persizo	/s/ - [s]	<ç>	<ç>	x				x		3
	persizo	/z/ - [z]	<ç>	<ç>	x				x		3
	precisso	/z/ - [z]	<ç>	<ss>			x		x		3
	presico	/s/ - [s]	<ç>	<ç>		x	x		x		1
	presico	/z/ - [z]	<ç>	<ç>		x	x		x		1
	presico	/s/ - [s]	<ç>	<ç>		x	x		x		1
	presico	/z/ - [z]	<ç>	<ç>		x	x		x		1
	presiso	/s/ - [s]	<ç>	<ç>		x	x		x		4
preguiçoso	preguiçoso	/s/ - [s]	<ç>	<ç>		x	x		x		1
preguiçosas	preguiçosas	/s/ - [s]	<ç>	<ç>		x			x		1
preocupação	preocupação	/s/ - [s]	<ç>	<ç>		x			x		1
	principalmente	/s/ - [s]	<ç>	<ç>	x				x		1
	processo	/s/ - [s]	<ss>	<ç>		x	x		x		1
produz	produz	/s/ - [j]	<ç>	<ç>	x				x		1
professor	professor	/s/ - [s]	<ss>	<ç>		x	x		x		3
	profissao	/s/ - [s]	<ss>	<ç>		x	x		x		1
	profissão	/s/ - [s]	<ss>	<ç>		x	x		x		3
profissional	profissional	/s/ - [s]	<ss>	<ç>	x				x		1
	profissional	/s/ - [s]	<ss>	<ç>	x				x		1
	profissional	/s/ - [s]	<ss>	<ç>		x	x		x		2
	relação	/s/ - [s]	<ç>	<ç>		x			x		3
	responsabilizar	/z/ - [z]	<ç>	<ç>	x				x		1
rigorosa	rigorosa	/z/ - [z]	<ç>	<ss>			x		x		1
sério	çerio	/s/ - [s]	<ç>	<ç>		x			x		1
sincero	çincerio	/s/ - [s]	<ç>	<ç>	x				x		1
	situações	/s/ - [s]	<ç>	<ç>		x			x		1
	sucedo	/s/ - [s]	<ss>	<ç>		x	x		x		1
	sustentões	/s/ - [s]	<ç>	<ç>		x			x		1
teste	teste	/s/ - [j]	<ç>	<ç>	x				x		1
vez	ves	/s/ - [j]	<ç>	<ç>	x				x		1

Forma gráfica alvo dos vocábulos com sibilantes ocorridas no <i>corpus</i>	Forma gráfica dos vocábulos com sibilantes registadas no <i>corpus</i>	Fonema/Fone alvo	Grafema correto	Representação desviante	Posição silábica	
					ataque	coda
Total de palavras						
etecetera	etc					
controlo	controlação					
Leonardo da Vinci	Leonardo da Vinci					
planejador	planejador					
por exemplo	porexemplo					
suco (aldeia)	suco					
suporte	suportador					
tem de ser	tenser					
dez	10					
cinquenta	50					
por cento	%					
três	3					