

TÂNIA SANTOS FERREIRA

**PADRÕES NA AQUISIÇÃO/APRENDIZAGEM DA
MARCAÇÃO DO GÊNERO NOMINAL EM
PORTUGUÊS COMO L2**

Volume I



**Faculdade de Letras
Universidade de Coimbra
2011**

Faculdade de Letras da Universidade de Coimbra

**PADRÕES NA AQUISIÇÃO/APRENDIZAGEM DA
MARCAÇÃO DO GÉNERO NOMINAL EM
PORTUGUÊS COMO L2**

Volume I

TÂNIA SANTOS FERREIRA

Dissertação de Mestrado em Português Língua Estrangeira/ Língua Segunda, área de especialização em Linguística Aplicada, apresentada à Faculdade de Letras da Universidade de Coimbra, sob a orientação das Professoras Doutoras Cristina dos Santos Pereira Martins e Maria Joana Vieira dos Santos.

Dedico esta dissertação aos meus pais, Joaquim e Isaura, e à minha irmã Carina que estiveram sempre comigo ao longo deste percurso

Índice Geral

Índice de Quadros	iii
Índice de Gráficos.....	iv
Lista de Abreviaturas.....	vi
Resumo	vii
Abstract.....	viii
Agradecimentos	ix
INTRODUÇÃO.....	1
CAPÍTULO 1	3
A AQUISIÇÃO/APRENDIZAGEM DE L2 – CONCEITOS FUNDAMENTAIS	3
1.1. Introdução e Objetivos.....	4
1.2. Língua Materna e L2.....	5
1.2.1. Aquisição da LM vs. Aquisição/Aprendizagem de L2	6
1.3. Interlíngua.....	12
CAPÍTULO 2	18
GÉNERO GRAMATICAL	18
2.1. Introdução e Objetivos.....	19
2.2. Para uma definição do conceito de Género Gramatical.....	20
2.3. O Género Nominal em português	26
2.3.1. Uma categoria da <i>flexão</i> ?	26
2.3.2. Critérios de marcação de Género dos nomes em português	32
2.4. Implicações para o Contexto de Aquisição/Aprendizagem de PL2	40
CAPÍTULO 3	41
TRATAMENTO E ANÁLISE DOS DADOS	41
3.1. Introdução e Objetivos.....	42
3.2. Metodologia.....	43

3.2.1. Descrição do <i>corpus</i>	43
3.2.2. Perfil dos informantes	46
3.3. Tratamento dos dados	51
3.3.1. Tipologia de desvios.....	51
3.4. Quantificação e análise dos dados apurados	61
3.4.1. Frequência de desvios do <i>corpus</i>	61
3.4.2 Frequência de desvios em função do nível de proficiência dos aprendentes .	66
3.4.2.1 Categoria 1.....	73
3.4.2.2 Categoria 2.....	77
3.4.2.3 Categoria 3.....	82
3.4.2.4 Categoria 4.....	86
3.4.2.5 Categoria 5.....	90
3.4.3 Síntese dos Resultados Apurados.....	94
CONSIDERAÇÕES FINAIS.....	96
BIBLIOGRAFIA.....	100

Índice de Quadros

QUADRO 2.1 – DISTRIBUIÇÃO DOS NOMES EM PORTUGUÊS QUANTO AO IT	31
QUADRO 2.2 - DISTRIBUIÇÃO DOS NOMES NÃO-SEXUADOS QUANTO AO IT E RESPECTIVO VALOR DE GEN	33
QUADRO 2.3 – FREQUÊNCIA DOS LEMAS RECENSEADOS NA CORLEX EM FUNÇÃO DO GEN E DA CLASSE TEMÁTICA DOS NOMES SIMPLES E DERIVADOS	34
QUADRO 2.4 - RELAÇÃO ENTRE OS SUFIXOS DERIVACIONAIS E RESPECTIVO GEN DOS PRODUTOS NOMINAIS	38
QUADRO 3.1 – DISTRIBUIÇÃO DE TEXTOS, PALAVRAS E NOMES POR GRUPOS DE INFORMANTES EM FUNÇÃO DO NÍVEL DE PROFICIÊNCIA EM PL2	45
QUADRO 3.2 – DISTRIBUIÇÃO DOS INFORMANTES POR LM	48
QUADRO 3.3 – CASOS DE HESITAÇÃO.....	52
QUADRO 3.4 – TIPOLOGIA DE DESVIOS	56
QUADRO 3.5 – CASOS DE HIPERCORREÇÃO	58
QUADRO 3.6 – DISTRIBUIÇÃO DO NÚMERO DE DESVIOS POR CATEGORIA DE DESVIO DE MARCAÇÃO DE GEN NOMINAL	62
QUADRO 3.7 – DISTRIBUIÇÃO DO NÚMERO DE DESVIOS POR SUBCATEGORIA DE DESVIO DE MARCAÇÃO DE GEN NOMINAL	64
QUADRO 3.8 – FREQUÊNCIA DE DESVIOS PRODUZIDOS POR NÍVEL	66
QUADRO 3.9 – DISTRIBUIÇÃO DE NOMES COM DESVIO DA CATEGORIA 1 E RESPETIVAS FREQUÊNCIAS DE OCORRÊNCIA DE DESVIOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	75
QUADRO 3.10 – DISTRIBUIÇÃO DE NOMES COM DESVIO DA CATEGORIA 2 E RESPETIVAS FREQUÊNCIAS DE OCORRÊNCIA DE DESVIOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	80
QUADRO 3.11 – DISTRIBUIÇÃO DE NOMES COM DESVIO DA CATEGORIA 3 E RESPETIVAS FREQUÊNCIAS DE OCORRÊNCIA DE DESVIOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	84
QUADRO 3.12 – DISTRIBUIÇÃO DE NOMES COM DESVIO DA CATEGORIA 4 E RESPETIVAS FREQUÊNCIAS DE OCORRÊNCIA DE DESVIOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	88
QUADRO 3.13 – DISTRIBUIÇÃO DE NOMES COM DESVIO DA CATEGORIA 5 E RESPETIVAS FREQUÊNCIAS DE OCORRÊNCIA DE DESVIOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	92

Índice de Gráficos

GRÁFICO 2.1 - DISTRIBUIÇÃO EM PERCENTAGEM DAS OCORRÊNCIAS DOS ITENS NOMINAIS EM FUNÇÃO DA CLASSE TEMÁTICA E DO VALOR DE GEN.....	35
GRÁFICO 2.2 – PERCENTAGEM DAS OCORRÊNCIAS DE NOMES MASCULINOS DE ACORDO COM A CLASSE TEMÁTICA A QUE PERTENCEM	35
GRÁFICO 2.3 – PERCENTAGEM DAS OCORRÊNCIAS DE NOMES FEMININO DE ACORDO COM A CLASSE TEMÁTICA A QUE PERTENCEM	36
GRÁFICO 2.4 – PERCENTAGEM DAS OCORRÊNCIAS DE NOMES COMUNS DE DOIS GEN DE ACORDO COM A CLASSE TEMÁTICA A QUE PERTENCEM	36
GRÁFICO 3.1 – PERCENTAGEM DE TEXTOS DO <i>CORPUS</i> COM DESVIOS DE MARCAÇÃO DE GEN.....	44
GRÁFICO 3.2 – PERCENTAGEM DE INFORMANTES POR CURSO DE PL2	46
GRÁFICO 3.3 – DISTRIBUIÇÃO DOS INFORMANTES POR IDADE	47
GRÁFICO 3.4 – DISTRIBUIÇÃO DOS INFORMANTES PELO ANO EM QUE INICIARAM O ESTUDO DE PL2	47
GRÁFICO 3.5 – DISTRIBUIÇÃO DOS INFORMANTES POR NÍVEL COMUM DE REFERÊNCIA DO QECRL.....	49
GRÁFICO 3.6 – PERCENTAGEM DE DESVIOS DE ATRIBUIÇÃO DE GEN NOMINAL	51
GRÁFICO 3.7 – DISTRIBUIÇÃO EM PERCENTAGEM DOS DESVIOS PRODUZIDOS NO <i>CORPUS</i> POR CATEGORIA.....	62
GRÁFICO 3.8 – DISTRIBUIÇÃO EM PERCENTAGEM DAS SUBCATEGORIAS DE DESVIOS PRODUZIDOS NO <i>CORPUS</i>	64
GRÁFICO 3.9 – PERCENTAGEM DE DESVIOS EM FUNÇÃO DOS NOMES PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA EM PL2	67
GRÁFICO 3.10 – PERCENTAGEM DE DESVIOS POR CATEGORIA PRODUZIDOS NO NÍVEL A1/A1+	68
GRÁFICO 3.11 – PERCENTAGEM DE DESVIOS POR CATEGORIA PRODUZIDOS NO NÍVEL A2/A2+	68
GRÁFICO 3.12 – PERCENTAGEM DE DESVIOS POR CATEGORIA PRODUZIDOS NO NÍVEL B1	68
GRÁFICO 3.13 – PERCENTAGEM DE DESVIOS POR CATEGORIA PRODUZIDOS NO NÍVEL B2	68

GRÁFICO 3.14 – PERCENTAGEM DE DESVIOS POR CATEGORIA PRODUZIDOS NO NÍVEL C1	68
GRÁFICO 3.15 – PERCENTAGEM TOTAL DE CATEGORIAS DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA DOS APRENDENTES.....	69
GRÁFICO 3.16 – PERCENTAGEM DE DESVIOS POR SUBCATEGORIA PRODUZIDOS NO NÍVEL A1/A1+	70
GRÁFICO 3.17 – PERCENTAGEM DE DESVIOS POR SUBCATEGORIA PRODUZIDOS NO NÍVEL A2/A2+	70
GRÁFICO 3.18 – PERCENTAGEM DE DESVIOS POR SUBCATEGORIA PRODUZIDOS NO NÍVEL B1	70
GRÁFICO 3.19 – PERCENTAGEM DE DESVIOS POR SUBCATEGORIA PRODUZIDOS NO NÍVEL B2	70
GRÁFICO 3.20 – PERCENTAGEM DE DESVIOS POR SUBCATEGORIA PRODUZIDOS NO NÍVEL C1	70
GRÁFICO 3.21 – PERCENTAGEM TOTAL DAS SUBCATEGORIAS DE DESVIO PRODUZIDAS POR NÍVEL DE PROFICIÊNCIA	71
GRÁFICO 3.22 – PERCENTAGEM DE DESVIOS DA CATEGORIA 1 EM FUNÇÃO DO CONJUNTO TOTAL DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA	73
GRÁFICO 3.23 – PERCENTAGEM DE DESVIOS DA CATEGORIA 2 EM FUNÇÃO DO CONJUNTO TOTAL DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA	77
GRÁFICO 3.24 – PERCENTAGEM DE DESVIOS DA CATEGORIA 3 EM FUNÇÃO DO CONJUNTO TOTAL DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA	82
GRÁFICO 3.25 – PERCENTAGEM DE DESVIOS DA CATEGORIA 4 EM FUNÇÃO DO CONJUNTO TOTAL DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA	86
GRÁFICO 3.26 – PERCENTAGEM DE DESVIOS DA CATEGORIA 5 EM FUNÇÃO DO CONJUNTO TOTAL DE DESVIOS PRODUZIDOS POR NÍVEL DE PROFICIÊNCIA	90

Lista de Abreviaturas

ASL – Aquisição de Segundas Línguas

CORLEX – Léxico Multifuncional Computorizado do Português Contemporâneo

GEN – Género gramatical

GN – Grupo nominal

GU – Gramática Universal

IL – Interlíngua

IT – Índice Temático

L2 – Língua não-materna

LA – Língua-alvo

LM – Língua materna

LO – Língua de origem

MCP – Memória a Curto Prazo

MLP – Memória a Longo Prazo

PEAPL2 - *Corpus* de Produções Escritas de Aprendentes de PL2

PL2 – Português língua não-materna

QECRL – Quadro Europeu Comum de Referência para as Línguas

Resumo

A presente dissertação tem como principal objetivo a identificação de padrões variáveis na aquisição/aprendizagem da marcação do género gramatical dentro do grupo nominal, a partir da análise de desvios produzidos por alunos estrangeiros de diferentes níveis de proficiência em português língua não-materna (PL2).

O objetivo específico deste trabalho é apurar, tendo em conta as particularidades da categoria de GEN em português, que tipo de características formais e semânticas dos nomes suscitam maiores dificuldades aos aprendentes durante o processo de aquisição/aprendizagem da marcação de GEN.

Assim, partindo da definição do conceito de *Interlíngua* (IL) proposto por Larry Selinker (1972) e da descrição das propriedades do sistema de marcação de GEN em português, seguindo atentamente os trabalhos de G. Corbett (1991) e Villalva (1994; 2003; 2008), procedeu-se à análise detalhada dos desvios de atribuição de GEN retirados de um conjunto diversificado de textos produzidos por alunos de PL2 que, entre Maio de 2009 e Maio de 2010, frequentaram cursos de português para estrangeiros na Faculdade de Letras da Universidade de Coimbra.

Deste modo, as correlações estabelecidas entre tipos de desvios, marcas de GEN e nível de proficiência dos aprendentes de PL2 permitem concluir que: i) em todos os níveis de proficiência em PL2 analisados se registam desvios de marcação de GEN; ii) os nomes cuja marcação de GEN não depende de critérios de natureza quer semântica quer formal suscitam maiores dificuldades aos aprendentes; iii) o avultado número de casos de atribuição de GEN nominal desviantes registado parece sugerir que, no processo de aquisição/aprendizagem lexical, a assimilação do valor de GEN de um nome é progressiva e relativamente lenta.

Palavras-chave: aquisição/aprendizagem de português língua não-materna, interlíngua, género nominal

Abstract

This research aims to identify the variable patterns in grammatical gender marking acquisition / learning within the noun group based on the analysis of errors produced by foreign students of various proficiency levels when learning Portuguese as a Foreign Language.

The specific aim of this thesis is to ascertain, considering the particularities of the gender category in Portuguese, which of the nouns' formal and semantic characteristics are more difficulties for learners during gender marking acquisition / learning.

Starting from the definition of Interlanguage proposed by Larry Selinker (1972) and the description of the Portuguese gender marking system based on the works of G. Corbett (1991) and Villalva (1994; 2003; 2008), the gender errors of a diverse group of texts produced by students of Portuguese as a Foreign Language were analysed. The students in this study attended the Portuguese as a Foreign Language course at the University of Coimbra's Faculdade de Letras between May 2009 and May 2010.

The correlation between different types of errors, gender marking and the proficiency level of the Portuguese as a Foreign Language learners allowed us to conclude that: i) gender marking errors were found in all levels of proficiency analysed; ii) the nouns in which gender marking does not depend on formal or semantic criteria cause more difficulties for learners; iii) the great number of noun gender marking errors seems to suggest that, during the lexical acquisition / learning process, the assimilation of a noun's gender is progressive and relatively slow.

Keywords: Portuguese as a Foreign Language acquisition / learning, interlanguage, nominal gender marking

Agradecimentos

Gostaria de agradecer, de forma muito especial, às Professoras Cristina dos Santos Pereira Martins e Maria Joana Vieira dos Santos, da Universidade de Coimbra, por todo o apoio, pela dedicação e pela orientação que me concederam ao longo deste último ano. Agradeço a confiança, em mim, o incentivo, a paciência que sempre demonstraram desde o primeiro momento. Estou eternamente grata pelos conselhos dados que me permitiram progredir não só como investigadora, mas, principalmente, como pessoa.

Agradeço à Professora Maria Carmen de Frias e Gouveia, por me ter disponibilizado, tão prontamente, os seus artigos científicos. Muito obrigada pelo carinho.

Agradeço a toda a minha família, em especial aos meus pais, à minha irmã, por todo o apoio que me concederam durante este ano.

Agradeço ao meu namorado Pedro Santos, por estar sempre a meu lado, e a toda a sua família, pelo carinho e pelas palavras de incentivo.

Agradeço a todos os meus amigos, em especial às minhas amigas Sandra Simões, Geraldine Garcia e Filipa Marques, pelos conselhos e sugestões concedidas durante a concretização deste trabalho. Obrigada pelo apoio, pelas palavras motivadoras, por toda a vossa preciosa ajuda e pela paciência que tiveram comigo durante este ano.

Agradeço ainda a todos aqueles que, de forma direta ou indireta, contribuíram para a conclusão desta dissertação. A todos vós, **MUITO OBRIGADA!**

Introdução

A presente dissertação alicerça-se num trabalho de investigação sobre a aquisição/aprendizagem da marcação do género (GEN) gramatical dentro do grupo nominal (GN), a partir da análise dos desvios relativos a esta categoria gramatical selecionados em textos escritos produzidos por aprendentes tardios de português como língua não-materna (PL2).

A escolha desta temática decorre, sobretudo, do facto de a categoria de GEN possuir um estatuto especial entre as demais categorias gramaticais, principalmente quando integrada num contexto de aquisição/aprendizagem de uma L2. Em boa verdade, e como se assinalará neste estudo, nem todas as línguas possuem a categoria de GEN. Acrescente-se ainda o facto de, mesmo nos sistemas linguísticos que possuem esta categoria, os critérios de atribuição de GEN serem muito variáveis. No caso específico do português, o valor de GEN de um nome não é inequivocamente dedutível a partir de marcadores morfológicos. Consequentemente, o valor de GEN dos nomes corresponde a uma área da gramática que não é, na sua plenitude, governada nem por um raciocínio semântico fiável, nem por regras passíveis de serem aplicadas com razoável grau de sucesso à totalidade dos itens lexicais assimilados e desde o início do processo de aquisição/aprendizagem no desempenho linguístico dos falantes aprendentes de PL2,

Deste modo, sendo esta uma categoria possivelmente “resistente” durante o processo de aquisição/aprendizagem de PL2, o objetivo central deste trabalho é o de tentar delimitar, com base na análise de desvios de marcação de GEN nominal, produzidos em textos escritos redigidos por aprendentes com diferentes níveis de proficiência em PL2, a existência de fases características do processo de construção dos sistemas interlinguísticos destes aprendentes. Mais concretamente, pretende-se responder às seguintes questões: i) é possível verificar a existência de padrões variáveis na aquisição/aprendizagem da marcação do GEN dentro do grupo nominal em alunos de PL2, em função dos seus níveis de aprendizagem?; e ii) que tipo de características formais e semânticas dos nomes suscitam mais dificuldades na aquisição/aprendizagem da marcação do GEN por parte dos aprendentes de PL2?

Assim, para o desenvolvimento deste trabalho procedeu-se à revisão dos principais conceitos ligados à área da Aquisição de Segundas Línguas, com especial enforque na noção de *Interlíngua*, seguindo as linhas teóricas de autores como L. Selinker (1992; 1993; 2008), S. Gass (1996; 2008) e R. Ellis (1986; 2000; 2003).

No que concerne à análise e descrição da categoria de GEN gramatical, teve-se em conta os estudos produzidos por G. Corbett (1991) e, para uma abordagem das principais características do sistema de marcação de GEN em português, seguiu-se atentamente os trabalhos desenvolvidos por J. Mattoso Câmara Jr. (1966; 1994) e Alina Villalva (1994; 2003; 2008).

Após a revisão teórica, procedeu-se à análise de um conjunto de produções escritas por alunos de PL2 com desvios de marcação de GEN nominal. Estes textos foram recolhidos a partir de um conjunto mais alargado de produções que integram o “*Corpus* de Produções Escritas de Aprendentes de PL2” (coord. Cristina Martins).

Refira-se, neste ponto, a organização da dissertação. Na presente *Introdução* apresenta-se os objetivos deste trabalho. Seguem-se três capítulos, sendo que os dois primeiros correspondem ao enquadramento teórico da investigação. O Capítulo 1 diz respeito às noções fundamentais ligadas ao domínio da Aquisição de Línguas Segundas, sublinhando as principais particularidades que revestem o processo de aquisição/aprendizagem da L2, bem como a descrição do conceito de *Interlíngua*, conceito central cunhado por L. Selinker em 1972, e dos fatores atuantes na construção dos sistemas interlinguísticos dos aprendentes de L2.

No capítulo seguinte, dedicado ao género gramatical, procede-se, numa primeira parte, à revisão das principais características que governam os sistemas de marcação de GEN nas línguas do mundo, com base no trabalho desenvolvido por Corbett (1991) e, numa segunda parte, apresenta-se o sistema de marcação de GEN nominal em português, bem como as suas particularidades.

No Capítulo 3 procede-se ao tratamento e análise dos desvios apurados neste trabalho, com a descrição detalhada dos métodos de recolha do *corpus*, tendo em conta as principais dificuldades na marcação do GEN nominal que os informantes em estudo apresentam.

Por fim, em *Considerações Finais*, apresenta-se um conjunto de reflexões desencadeadas a partir dos resultados obtidos na dissertação.

CAPÍTULO 1
A AQUISIÇÃO/APRENDIZAGEM DE L2 –
CONCEITOS FUNDAMENTAIS

1.1. INTRODUÇÃO E OBJETIVOS

Nas últimas décadas, muitos investigadores, tanto na área da Linguística como da Psicolinguística, têm procurado identificar fatores de diversa natureza (linguística, cognitiva, cultural e social) para explicitar as diferenças subjacentes aos processos de aquisição de línguas maternas e de línguas não-maternas. Tal interesse levou, a partir, essencialmente, dos anos 60 do século XX, à elaboração de numerosos estudos que constituem a área de investigação em Aquisição de Línguas Segundas (ALS, tradução de SLA: *Second Language Acquisition*).

Uma vez que este trabalho incide na aquisição/aprendizagem de português como língua não-materna por aprendentes tardios, no presente capítulo serão tratadas as noções que, no âmbito da investigação em ALS, dizem respeito às principais particularidades deste processo.

Deste modo, num primeiro momento, apresenta-se uma breve descrição da terminologia usada no campo da ALS, mais precisamente os conceitos de *língua materna* e *língua não-materna* (L2) (secção 1.2).

Após a exposição teórica tendente à descrição conceptual dos termos a utilizar, proceder-se-á à sucinta apresentação das questões mais relevantes no que se refere ao processo de aquisição/aprendizagem de línguas segundas (L2) e, em particular, das diferenças que apresentam os aprendentes tardios em relação aos aprendentes precoces (secção 1.2.1). Tais divergências são notórias se considerarmos: i) o *estádio inicial* dos aprendentes; ii) o *estádio final* do processo, i.e., as competências que o falante adquire e com que grau de proficiência as adquire; e iii) o próprio processo de aquisição/aprendizagem de uma L2 em si.

Por último, descritas as várias características do processo de aquisição/aprendizagem de uma L2, procede-se à descrição do conceito de *Interlíngua* (secção 1.3) no que respeita às suas particularidades e aos fatores que atuam na construção dos sistemas interlinguísticos dos falantes aprendentes de uma L2. Este ponto é fundamental, visto que se pretende identificar, no estudo empírico, com base na análise dos desvios de marcação de GEN nominal produzidos pelos informantes, as diferentes fases do processo de construção da interlíngua dos aprendentes de português língua não-materna (PL2).

1.2. LÍNGUA MATERNA E L2

O termo *língua materna* (LM) assinala a língua que o falante adquire nos primeiros anos de vida. Alguns investigadores preferem designar este primeiro idioma como *língua nativa* ou L1 (Gass & Selinker, 2008:7). A LM constitui o primeiro sistema linguístico de socialização da criança adquirido em contexto familiar. A assimilação das estruturas linguísticas nativas ocorre por *aquisição* natural e espontânea, numa fase precoce do desenvolvimento do falante.

Quando a criança é precocemente exposta a dois sistemas linguísticos, é possível admitir que possui duas LM e, nestes casos, o termo L2 poderá corresponder a um desses idiomas maternos. No entanto, Hamers & Blanc (1990) preferem utilizar as designações LA e LB para designarem as línguas que, ao serem adquiridas em simultâneo, constituem os dois sistemas linguísticos nativos da criança (1990:10). Para estes autores, o termo L2 designa, exclusivamente, a segunda língua adquirida pelo falante depois de já ter adquirido a LM, i.e., a L2 corresponde a um sistema linguístico não materno (cf. Hamers & Blanc, 1990:10). Tendo em conta a proposta defendida por estes autores, neste trabalho optou-se por usar o termo L2¹ como sinónimo de língua não-materna.

Quanto ao processamento das estruturas linguísticas uma L2, principalmente em fases tardias do desenvolvimento do falante, não existe consenso entre a comunidade científica sobre a conceptualização dos termos “aquisição” e “aprendizagem” de uma L2. Tal indefinição resulta, sobretudo, da dificuldade em obter critérios que, de forma sólida, permitam distinguir os dois conceitos. Não sendo o nosso objetivo discutir as diferentes posições tomadas pelos investigadores a este respeito², convém, no entanto,

¹ Preferiu-se o uso do termo genérico L2 de forma a evitar as dificuldades inerentes ao recurso de termos como *língua segunda* (LS) e *língua estrangeira* (LE) que, apesar de designarem uma língua não-nativa, não se referem à mesma situação, ou seja, apresentam divergências quanto à realidade sociolinguística que designam (Leiria, 2004). O termo LS é utilizado para se referir à língua não-materna que, numa determinada comunidade, possui um estatuto sócio-político definido (Leiria, 2004:1). A LS pode corresponder a uma das línguas oficiais de um país, é ensinada nas escolas e desempenha um papel crucial dentro das fronteiras territoriais onde é utilizada (Leiria, 2004). Pelo contrário, a noção de LE diz respeito a um sistema linguístico aprendido, principalmente em contexto formal e é utilizado dentro de uma comunidade onde essa língua não possui um estatuto definido (Leiria, 2004).

² De um modo geral, o termo “aquisição” é utilizado para designar o processamento das estruturas de uma língua em fases precoces do desenvolvimento humano, que ocorre em contextos naturais e de forma espontânea, quase “automática”. Já a noção “aprendizagem” designa o processo de assimilação de uma língua em fases tardias e está, tipicamente, circunscrito a um contexto formal. O aprendente tem, para além do mais, uma clara consciência do processo de aprendizagem do novo idioma. Todavia, para muitos

referir que alguns dos fatores associados à distinção destes termos contribuem, igualmente, para compreender as diferenças inerentes entre o processo de aquisição de uma LM e a aquisição ou aprendizagem de uma L2. Tais diferenças serão apresentadas na próxima secção (secção 1.2.1).

Acrescente-se ainda o facto de existirem autores que não reconhecem que os processos “aquisição” e “aprendizagem”, numa e noutra fase de desenvolvimento, sejam qualitativamente distintos, na medida em que envolverão os mesmos mecanismos básicos (cf. White, 1995:87). Em suma, e não sendo este um assunto consensual, compreende-se a dificuldade em preferir a utilização de um termo em detrimento do outro. Assim sendo, optou-se na presente dissertação pela designação “aquisição/aprendizagem de L2”. Tal opção terminológica revela-se mais eficaz, na medida em que abrange as distintas situações dos informantes que produziram o *corpus* de produções escritas utilizado nesta investigação (cf. Capítulo 3, secção 3.2.2).

1.2.1. AQUISIÇÃO DA LM VS. AQUISIÇÃO/APRENDIZAGEM DE L2

No domínio da investigação em ALS é comumente aceite que o processo de aquisição da LM difere, em alguns aspetos, do processo de aquisição/aprendizagem da L2. Esta distinção compreende-se devido às seguintes evidências: i) quando o falante adquire/aprende uma L2, já possui um sistema linguístico prévio (a LM); ii) somente uma minoria dos aprendentes tardios de uma L2 adquire um grau de competência e proficiência linguísticas que se possam considerar equivalentes às dos falantes nativos dessa L2 e, mesmo esses, conseguem-no com muito esforço (Gass & Selinker, 2008).

Tais evidências são cruciais para compreender em que medida o processo de aquisição/aprendizagem da L2 se distingue do processo de aquisição da LM. Por conseguinte, procede-se a uma breve exposição das principais particularidades que possuem tanto o *estádio inicial* como o *estádio final* dos aprendentes tardios de L2, tendo em conta o contraste que se pode estabelecer com os respetivos estádios no processo de aquisição da LM. Por fim, e referidas as especificidades que envolvem o processo de aquisição/aprendizagem de uma L2, na próxima secção (1.3) descrevem-se as características do sistema linguístico atualizado pelo falante aprendente de uma L2,

investigadores, os critérios evocados não constituem uma forma clara de distinguir a “aquisição” de “aprendizagem” de uma L2. Sobre este assunto, veja-se Krashen, 1981; Ellis, 2000.

i.e., a *interlíngua*, que, como se depreenderá, apresenta um conjunto de características específicas que serão consideradas e ilustradas pelos dados do nosso estudo empírico (Capítulo 3).

Considerando, em primeiro lugar, o *estádio inicial* da criança aquando da aquisição da LM existem duas perspetivas teóricas básicas a considerar: i) a posição empirista³; e ii) a posição inatista⁴.

Segundo a tradição behaviorista, corrente representativa da posição empirista, muito em voga nas décadas de 40 e 50 do século XX, a criança inicia o processo de aquisição da LM sem possuir qualquer tipo de conhecimento prévio. Assim, todo o conhecimento linguístico que vai adquirindo provém da experiência, ou seja, da interação com o mundo real. De acordo com esta posição⁵, entende-se que a aquisição da linguagem se processa por uma sucessão de hábitos⁶, i.e., de respostas aos estímulos que advém do mundo extralinguístico (Bhatia & Ritchie, 1996:5).

Em oposição à teoria behaviorista surge um modelo diferente para a compreensão da aquisição da linguagem: a versão inatista. Noam Chomsky (1975) argumenta que a teoria behaviorista não explica, por exemplo, como um bebé, ao estar exposto a estruturas linguísticas simples, consegue reconhecer e usar outras estruturas mais complexas⁷. Este argumento da *pobreza do estímulo* conduz, de forma inevitável, à postulação de estruturas mentais inatas que atuam sobre o *input* linguístico (Bhatia & Ritchie, 1996:6).

Deste modo, N. Chomsky (1975:141) reconhece a existência de um dispositivo cognitivo para adquirir línguas (*Language Acquisition Device*), independente de outros mecanismos cognitivos e especificamente programado para a aquisição dos sistemas

³ Defendida, entre outros, pelos behavioristas.

⁴ Proposta enquadrada no âmbito da Gramática Generativa.

⁵ A posição empirista foi partilhada pelos linguistas estruturalistas americanos (cf. Bhatia & Ritchie, 1996).

⁶ Ao tratar as diferentes teorias que influenciaram os estudos em ALS, e em particular a teoria behaviorista da aquisição da linguagem, Rod Ellis descreve, de acordo com este modelo, a noção de hábito: “Habits are formed when learners respond to stimuli in the environment and subsequently have their response reinforced so that they are remembered. Thus, a habit is a stimulus-response connection” (2003:31).

⁷ Chomsky (1975:141) afirma que: “parece evidente que a aquisição da linguagem se baseia na descoberta pela criança daquilo que, de um ponto de vista formal, constitui uma teoria profunda e abstrata – uma gramática generativa da sua língua – da qual muitos dos conceitos e princípios se encontram apenas remotamente relacionados com a experiência através de cadeias longas e complexas de etapas inconscientes e de natureza quase dedutiva. Considerações acerca da natureza da gramática que é adquirida, da qualidade inferior e da extensão dos dados à disposição do sujeito (...) não permitem pensar que uma grande parte das estruturas da linguagem possa ser aprendida por um organismo não informado inicialmente sobre a sua natureza geral”.

linguísticos. Segundo o autor, neste dispositivo estariam integradas as informações sobre a forma possível que a gramática de qualquer língua poderia assumir e, partindo deste pressuposto, define o importante conceito da *Gramática Universal* (GU). A GU corresponde a um conjunto de princípios e parâmetros comuns a todas as línguas que configuram a capacidade inata para a linguagem (Bhatia & Ritchie, 1996:705). Para L. White (2003) este conceito é relevante para o domínio da aquisição das línguas, uma vez que a “UG is postulated as an explanation of how it is that learners come to know properties of grammar that go far beyond the input, how they know that certain things are not possible, why grammars are of one sort rather than another” (2003:20).

A teoria chomskyana da aquisição da linguagem tem especial interesse para o campo da aquisição/aprendizagem da L2. Com efeito, em meados dos anos 80, investigadores que integram o quadro teórico da Gramática Gerativa começaram a avaliar se os princípios e parâmetros da GU estariam disponíveis durante a aquisição/aprendizagem da L2. No que se refere à definição do *estádio inicial* do aprendente tardio de uma L2 os investigadores dividem-se em função das suas posições em relação a duas questões básicas: por um lado, o grau de acessibilidade à GU e, por outro, o papel que a LM tem no início do processo de aquisição/aprendizagem da L2.

Assinale-se, num primeiro momento, a proposta dos investigadores que negam o acesso à GU no âmbito da aquisição/aprendizagem da L2. Segundo esta perspectiva, os aprendentes baseiam-se, simplesmente, num conjunto de estratégias de aprendizagem geral e, tal facto justifica, segundo os defensores desta posição, não só as diferenças entre a aquisição da LM e a aquisição/aprendizagem da L2, mas também o facto de os aprendentes tardios da L2 dificilmente adquirirem um grau de proficiência linguística nativa (Franceschina, 2005).

Outra perspectiva entende que o *estádio inicial* do aprendente de uma L2 se caracteriza pelo acesso pleno à GU. Os defensores desta posição assumem que a aquisição/aprendizagem da L2 é semelhante ao da aquisição da LM, e é possível que os aprendentes consigam atingir o mesmo grau de proficiência dos falantes nativos (cf. White, 1995).

Por sua vez, alguns autores entendem ainda que a LM constitui o *estádio inicial* da aquisição/aprendizagem da L2. Neste sentido Schwartz & Sprouse propõem que o aprendente, ao iniciar o processo de aquisição/aprendizagem da L2, tem acesso à GU, através da gramática da LM: “All the principles and parameter values as instantiated in

the L1 grammar immediately carry over as the initial state of a new gramatical system on first exposure to input the target language” (1996:41, *apud* Han, 2004:66).

Apesar das várias posições teóricas sobre as características do *estádio inicial* do aprendente de L2, a verdade é que, quando este inicia o processo de aquisição/aprendizagem, possui conhecimentos prévios que vão condicionar não só o modo como o falante assimila as estruturas linguísticas dessa L2, mas também os resultados obtidos, i.e., o *estádio final* da aquisição/aprendizagem. De facto, as crianças muito novas conseguem, facilmente, adquirir um grau de proficiência nativa das línguas às quais estão expostas (Gass & Selinker, 2008:164), mas, em contrapartida, aprendentes tardios de uma L2 dificilmente conseguem atingir um grau de proficiência linguística nativo. Selinker admite que somente uma escassa minoria de aprendentes adultos adquire competência nativa na L2 (1992:83).

Perante estas diferenças, os investigadores têm procurado compreender quais os fatores que concorrem para a sua presença. Muitos têm advogado que elas radicam em questões maturacionais, i.e., nas especificidades das distintas fases do desenvolvimento neurocognitivo do aprendente e na sua capacidade de resposta a *input* linguístico.

Deste modo, e no que se refere à cronologia em que a aquisição linguística da LM ocorre, muitos investigadores consideram que existe um período crítico para adquirir uma língua, apesar de haver divergências quanto ao seu limite cronológico (Ellis, 1986:107). A título de exemplo, refira-se Lenneberg (1967) que defende o fim do período crítico na puberdade, sendo este coincidente com o final do processo de lateralização⁸ das funções linguísticas no hemisfério esquerdo do cérebro (Lenneberg, 1967). Por sua vez, Gleitman & Newport (1995:3) defendem o fim do período crítico aos 5/6 anos de idade, uma vez que nesta altura a criança já adquiriu estruturas linguísticas complexas, e por volta dos cinco anos “youngsters sound essentially adult” (1995:3).

Quanto à aquisição/aprendizagem de L2, não existe consenso no que diz respeito à existência de um período crítico. Aliás, muitos investigadores consideram que os resultados obtidos na aquisição/aprendizagem de uma L2 são condicionados por um conjunto variado de fatores tais como a motivação e os próprios objetivos do aprendente

⁸A lateralização consiste, nas palavras de Bhatia & Ritchie (1996:697) em: “localization of a particular brain function in one side of the brain or the other”. Trata-se de um processo que perpassa distintas fases, uma vez que a lateralização de um conjunto de funções – tanto linguísticas como também motoras, tal como as que estão envolvidas na perceção visual e auditiva – resulta do amadurecimento do cérebro. Por outras palavras, o cérebro de um recém-nascido não evidencia a lateralização de funções que se observa no cérebro de um adulto.

durante o processo (cf. Ioup, 2005:419). Tal evidência justifica, em parte, o facto de não ser possível equacionar um limite cronológico no âmbito da aquisição/aprendizagem de uma L2.

Também os subsistemas de memória mobilizados na aquisição da LM e na aquisição/aprendizagem da L2 contribuem, igualmente, para compreender as diferenças existentes. A memória humana está dividida em: Memória a Curto Prazo (MCP) ou Memória de Trabalho; e Memória a Longo Prazo (MLP). A MLP subdivide-se, ainda, em MLP procedimental e MLP declarativa (que, por sua vez se segmenta em MLP declarativa semântica e MLP declarativa episódica) (Baddeley, 1999). Cada subsistema é responsável por distintas tarefas e, no caso específico da assimilação e codificação das estruturas linguísticas, importa considerar o papel da MLP procedimental e da MLP declarativa.

A MLP procedimental é o subsistema responsável por “the learning of new, and the control of established, sensori-motor and cognitive “habits”, “skills”, and other procedures, such as riding a bicycle and skilled game playing” (Ullman, 2004:237), i.e., por saberes práticos, acessíveis e de utilização imediata, automática. No que diz respeito às representações linguísticas, e dado que a aquisição das representações armazenadas na MLP procedimental é feita sem um esforço consciente e por mera exposição de *input* linguístico, Ullman (2004) e Paradis (2004) defendem, nos seus estudos relacionados com o papel dos subsistemas de memória nestes processos, que a aquisição da gramática da LM é suportada por este subtipo de MLP.

Quando a exposição a estruturas linguísticas da L2 ocorre em fases tardias do desenvolvimento do falante, mais dificilmente ele parece conseguir assimilar essas mesmas estruturas pela MLP procedimental. Com efeito, e segundo Ullman (2004) e Paradis (2004), a aquisição/aprendizagem da gramática de uma L2 é alcançada com a participação da MLP declarativa, o que implica um esforço consciente do aprendente durante o processamento linguístico da L2. Esta posição explica por que razão são escassos os casos em que os aprendentes tardios de uma L2 atingem um grau de proficiência linguística, passível de ser considerado ‘nativo’. Sintetizando, segundo estes autores o processo de assimilação das estruturas linguísticas da LM e da L2 é diferente, pelo que os resultados óbitos em cada processo também o são.

Acrescente-se ainda que, geralmente, os aprendentes tardios de L2 acusam, a determinado momento do seu desenvolvimento, dificuldades em progredir, apesar de uma contínua exposição a *input* linguístico dessa língua, de uma forte motivação para

aprender o novo sistema linguístico, e de suficientes oportunidades para praticar e usar essa língua. Tal facto revela a existência de um fenómeno típico, referido na bibliografia sobre a ALS como *fossilização* (cf. secção 1.3).

Em suma, e através do que ficou dito, percebe-se que o processo de aquisição/aprendizagem de L2 é, grosso modo, distinto do processo de aquisição da LM. Além disso, o próprio sistema linguístico atualizado pelo aprendente de L2 apresenta distintas particularidades, pois não corresponde, na maioria dos casos, aos enunciados produzidos por um falante nativo dessa L2. Esta mesma evidência deu origem ao conceito de *Interlíngua* (IL), noção central na área da investigação em ALS, que se define na próxima secção.

1.3. INTERLÍNGUA

As interlínguas (IL) são sistemas linguísticos intermédios que o aprendente tardio, durante o seu processo de aquisição/aprendizagem da L2, vai construindo em direção à língua-alvo (LA)⁹. Do ponto de vista estrutural, as IL não vão corresponder nem às estruturas da língua de origem (LO), nem às estruturas da LA.

Diversos investigadores (Corder, 1992; Selinker, 1992) compreenderam que o sistema linguístico atualizado pelos aprendentes tardios da L2 apresenta características próprias. Segundo estes autores, os “erros”¹⁰ cometidos pelos falantes aprendentes de uma L2 não devem ser considerados fruto do acaso e, por isso, necessitam de ser analisados, no sentido de comprovarem a existência de um sistema linguístico intermédio, condicionado por regras próprias. Corder admite que “los errors de los alumnos prueban la existencia de este sistema y son sistemáticos en sí mismos” (Corder, 1992:37). A partir deste ponto de vista, Selinker propõe, no seu artigo seminal *Interlanguage*, publicado em 1972, o conceito de *Interlíngua* para designar o sistema linguístico independente e único que os aprendentes de L2 evidenciam e que apresenta uma organização baseada em regras e princípios (Selinker, 1992).

O reconhecimento da IL pressupõe que os aprendentes de L2 terão interiorizado uma gramática mental, um sistema linguístico abstrato que se rege por um conjunto de regras e princípios que vão condicionar tanto o domínio da compreensão como o de produção da L2. Por este motivo, Selinker (1992) assume a existência de uma estrutura psicológica latente, mobilizada para a aquisição/aprendizagem da L2, que corresponde a uma alternativa ao conceito chomskyano de GU. Segundo o autor, esta estrutura psicológica, apesar de estar ao serviço da aquisição/aprendizagem da L2, nem sempre é ativada, ou seja, nem sempre a aquisição/aprendizagem da L2 é bem-sucedida (Selinker, 1992). Para a formulação deste conceito, Selinker (1992) inspirou-se na proposta de Lenneberg (1967) que defende a existência da estrutura linguística latente. No entanto,

⁹ A designação de língua-alvo (LA) corresponde à L2 que o falante pretende adquirir e “se restringe unicamente a uma norma de un dialeto que constituye el foco interlingual de atención del alumno” (Selinker, 1992:83).

¹⁰ Corder (1992) nota que nem todos os erros cometidos pelo falante devem ser objeto de análise. Para o autor é necessário distinguir erros sistemáticos dos não sistemáticos, sendo que os primeiros – designados como “faltas” – correspondem a lapsos casuais produzidos pelo aprendente e “no tienen ninguna importancia para el proceso del aprendizaje de la lengua” (1992:37) e os segundos – que se designam “erros” – se referem a desvios recorrentes na produção do falante, devendo ser, por isso, analisados. No entanto, Corder sublinha que distinguir as “faltas” dos “erros” pode ser uma tarefa muito complicada para o investigador e implica uma análise exaustiva dos dados, concluindo que só o contexto da produção do aprendente permite indicar a presença de um “erro” (1992:39).

face à proposta de Lenneberg, a estrutura psicológica latente apresenta algumas divergências. Com efeito, a estrutura linguística latente conduz, segundo Lenneberg (1967), a uma competência e proficiência nativas na língua que a atualiza e, além disso, não pode ser ativada durante todas as fases do desenvolvimento humano (Lenneberg, 1967). O mesmo autor revela que esta estrutura apenas pode ser mobilizada até à puberdade, ou seja, está correlacionada com a hipótese do período crítico para a aquisição da linguagem (cf. secção 1.2.1).

Outros investigadores optam por uma posição distinta e entendem que, na construção da IL, é possível inferir a influência total ou parcial da GU (White, 1995:19) (cf. secção 1.2.1). Corder (1992) assume que, quer na aquisição da LM, quer na aquisição/aprendizagem da L2, a GU atua, independentemente de outros fatores, como a idade do aprendente aquando da exposição ao *input* linguístico (Corder, 1992). Por conseguinte, os resultados obtidos, no que respeita aos domínios de produção e compreensão da L2, ficar-se-ão a dever, por exemplo, à própria motivação do falante (Corder, 1992).

Considera-se que as IL correspondem a sistemas linguísticos, uma vez que durante a construção da IL, os fenómenos desviantes produzidos pelos aprendentes em relação à LA são, de certo modo, previsíveis. Tendo em conta este facto, na presente dissertação pretende-se verificar se os desvios produzidos pelos alunos do *corpus* permitem identificar o processo de construção da IL, i.e., se é possível inferir diferentes fases do desenvolvimento linguístico do falante com base na correlação entre os fenómenos desviantes observados na marcação de género nominal e o nível de proficiência / competência linguística dos aprendentes de PL2 (cf. Capítulo 3).

Com efeito, à semelhança do que ocorre na aquisição da LM, os desvios sistemáticos dos falantes serão representativos das diferentes etapas por que os aprendentes passam ao longo da aquisição/aprendizagem da L2, i.e., evidenciam um processo gradual, transicional, reflexo do que Corder (1992) designa por “built-in-syllabus”. O processo de construção da IL é, deste modo, encarado como um *continuum* que perpassa por distintos estágios de desenvolvimento (Ellis, 2003).

Os sistemas interlinguísticos caracterizam-se por uma certa instabilidade, ou seja, são sistemas flutuantes, de transição e em permanente reestruturação, o que pode resultar tanto de fatores externos (através do *input* que o aprendente recebe), como de fatores internos. Segundo Ellis (1986), o dinamismo observável na passagem dos diferentes estádios da IL, ao longo do *continuum*, é resultante das hipóteses que o

falante coloca em relação ao próprio funcionamento da L2. O autor ilustra este dinamismo com o seguinte caso: “For example, early WH questions are typically non-inverted (e.g. ‘What you want?’), but when the learner acquires the subject-inversion rule, he does not apply it immediately to all WH questions. To begin with he restricts the rule to a limited number of verbs and to particular WH pronouns (e.g. ‘who’ and ‘what’). Later he extends the rule, by making it apply both to an increasing range of verbs and to other WH pronouns” (Ellis, 1986:50). Este caso específico retrata o modo como os aprendentes processam e revêm as regras da L2, o que resulta, por vezes, num certo “descontrole” da dinâmica do processo, verificando-se a aplicação das novas regras à totalidade dos contextos linguísticos, até que o aprendente alcance uma nova fase de desenvolvimento (Ellis, 1986:50).

Selinker (1992) identifica vários processos atuantes na construção dos sistemas interlinguísticos. Um desses processos designa-se por transferência de instrução, que se caracteriza por manifestações desviantes em relação à LA atribuíveis a hábitos ou métodos de ensino (cf. Selinker, 1992:83). Outro fator diz respeito às estratégias de aprendizagem da LA, observáveis no facto de o aprendente tender à simplificação do material linguístico da L2 (Selinker, 1992:88), evidenciando, por exemplo, o não emprego dos artigos, a ausência de marcação do plural ou a omissão de formas verbais.

A sobregeneralização das regras da LA atua igualmente na construção das IL e também pode ser vista como resultando da estratégia de simplificação da LA pelos aprendentes. O produto de tais sobregeneralizações observa-se, essencialmente, nos planos morfológico e morfo-sintático através da regularização dos paradigmas flexionais (Selinker, 1992). Quando o aprendente assume que atingiu, de forma satisfatória, os objetivos de comunicação da LA, este pode considerar que a sua IL está apta a desempenhar as funções comunicativas dessa língua e, nesse caso, tende a “estabilizar” determinadas estruturas linguísticas, cessando o desenvolvimento da IL (Selinker, 1992)

Outro processo indicado por Selinker (1992) corresponde à transferência linguística. Genericamente, a transferência designa a presença de informação linguística da LM nos enunciados da L2. O papel que o primeiro sistema linguístico tem na aquisição/aprendizagem da L2 tem sido amplamente discutido por diferentes

perspetivas teóricas.¹¹ Os behavioristas procederam à análise contrastiva da LM e da LA¹², pois acreditavam que grande parte dos desvios produzidos pelos aprendentes de L2 correspondia à influência do sistema linguístico nativo do informante (Gass, 1996:319).

Todavia, para alguns investigadores, a presença de desvios na produção dos falantes não é, necessariamente, reveladora de um processo de transferência da LM e por isso, rejeitam o modelo de transferência proposto pelos behavioristas (Gass, 1996:318). Deste modo, preferem utilizar o termo “influência” da LM para designar a presença de estruturas linguísticas do sistema nativo do falante nos enunciados que produz na L2 (Gass, 1996:318). Esta influência pode surtir um efeito positivo ou negativo, dependendo do grau de afinidade estrutural entre a LM e a L2. Por um lado, quando o aprendente, ao comunicar na L2, transfere diversas estruturas da sua LM que não correspondem à LA, evidencia-se a presença de transferência negativa, também designada como *interferência* e, nestes casos, o desvio é facilmente detetável. Por outro, a transferência positiva corresponde ao facto de a LM, pela semelhante partilha de estruturas com a L2, poder constituir um fator positivo na aquisição da L2 (Gass & Selinker, 2008:94).

Fundamentalmente, e integrada na perspetiva da IL, a influência da LM corresponde a um dos processos atuantes na construção do sistema interlinguístico e, ainda, contribui para a formulação de hipóteses sobre o *input* linguístico durante o processo de aquisição/aprendizagem da L2 (cf. Ellis, 1986:48).

Por fim, convém salientar que as ILs são sistemas que não correspondem, integralmente, a um conjunto de estruturas desviantes em relação à LA. Resultam, sobretudo, de o facto de os aprendentes irem colocando hipóteses em relação ao *input* que vão recebendo e, desta forma, constituírem um processo de construção criativa (Corder, 1992). Em última instância, uma vez que somente uma pequena percentagem dos falantes aprendentes da L2 atinge um grau de competência e proficiência linguística nativas (cf. Selinker, 1992:83), a IL caracteriza-se, necessariamente, pela presença de fenómenos linguísticos fossilizados. Convém, por isso, definir o conceito de *fossilização* e de que forma tem sido abordado na bibliografia em ALS.

¹¹ Para uma revisão aprofundada do fenómeno da transferência, veja-se Gass, 1996:317-345.

¹² A análise contrastiva corresponde, nas palavras de Bhatia & Ritchie (1996:692): “a procedure whereby the structures of two languages are compared with the purpose of determining what difficulties a native speaker of one of the languages will have in acquiring the other one”.

A presença permanente de desvios na produção dos falantes constituiu o ponto de partida para a formulação do conceito de *fossilização*. Procede-se, em seguida, a uma abordagem geral deste fenómeno, principalmente no que respeita às dificuldades inerentes à sua descrição. Como se verificará, tais dificuldades têm contribuído para alguns investigadores preferirem o conceito de “estabilização” ao de “fossilização”.

Larry Selinker é o primeiro a referir-se ao termo *fossilização*. Para este autor, os fenómenos linguísticos fossilizados são “ítems, reglas y subsistemas lingüísticos que los hablantes de una LM particular tienden a conservar en su IL en relación com una LO dada sin importar cuál sea la edad del alumno o cuánto entrenamiento haya recibido en la LO” (1992:85). De acordo com esta formulação, as estruturas fossilizadas são persistentes, resistentes a fatores externos e além disso, afetam a generalidade dos aprendentes de uma L2.

Além disso, Selinker refere que, por vezes, o aprendente denunciará um movimento de *backsliding*, i.e., “las estructuras fosilizables tienden a permanecer como actuación potencial, reemergiendo en la producción de una IL incluso cuando parecían erradicadas” (1992:85). Esta regressão manifesta-se no plano formal (fonético-fonológico e morfo-sintático) e ocorre, geralmente, em situações comunicativas, em que o aprendente está concentrado no efeito comunicativo do seu enunciado (Selinker, 1992).

A partir desta conceptualização inicial, Selinker foi redefinindo o conceito. Em 1978, Selinker & Lamendella explicam o fenómeno como “(...) the permanent cessation of IL learning before the learner has attained target language norms at all levels of linguistic structure and in all discourse domains in spite of the learner’s positive ability, opportunity or motivation to learn or acculturate into target society (Selinker & Lamendella, 1978:187, *apud* Long, 2003:488-489). Deste modo, encontram-se diferenças relativamente à definição de 1972, visto que os fenómenos fossilizados afetam, nesta nova perspetiva, todos os níveis da estrutura linguística e do domínio discursivo (1978), alargando a noção original de que fenómenos fossilizados diziam apenas respeito a itens, regras e subsistemas linguísticos (1972).

A constante presença de estruturas próprias da IL na produção de enunciados da LA constitui, para Selinker & Lakshmanan (1993), um importante fator para identificar, nos dados empíricos, estruturas fossilizadas. Por ser um fenómeno que corresponde a um progressivo bloqueio na aquisição/aprendizagem da L2, a fossilização verifica-se de forma permanente e, uma vez que os aprendentes tardios dificilmente alcançam uma

competência e uma proficiência linguística nativas na LA, estes autores consideram-na inevitável (cf. Selinker & Lakshmanan, 1993).

A noção de *fossilização* é complementada através do contributo de Rod Ellis (1986). Segundo este autor, a fossilização assume-se como uma componente necessária da construção da IL, visto que os fenómenos fossilizados, independentemente do grau de coincidência que apresentam em relação aos correspondentes na LA, ocorrem em determinado ponto do desenvolvimento dos sistemas interlinguísticos. Com efeito, para o investigador existem erros fossilizados e formas da LA fossilizadas, ou seja, a fossilização resulta não só em formas incorretas, mas também em formas corretas (Ellis, 1986:48).

Apesar de a fossilização ser um fenómeno muito discutido, a verdade é que grande parte dos investigadores conclui que este termo é pouco operativo. Com efeito, por os fenómenos genuinamente fossilizados serem muito difíceis de identificar, a maioria dos estudos que discutem este fenómeno não se realizam com base em dados empíricos que, verdadeiramente, denotem a existência de estruturas linguísticas fossilizadas. Por essa razão, na medida em que evidenciam o “estádio final” do processo de aquisição/aprendizagem da L2, muitos autores, como Michael Long (2003), consideram que os fenómenos linguísticos fossilizados são impossíveis de verificar empiricamente. O autor prefere, assim, a noção de “estabilização”, estado que antecederá o de uma eventual fossilização (Long, 2003:521).

Para este trabalho reconhecemos, igualmente, a dificuldade em conseguir identificar a presença de verdadeiros fenómenos fossilizados, dadas as características diversificadas dos informantes que participaram na constituição do *corpus* utilizado, nomeadamente no que respeita à idade, ano de início de estudo em PL2, LM e nível de proficiência em PL2. Dessa forma, ainda que admitamos a presença de desvios persistentes quanto à marcação do género nominal num determinado estágio do desenvolvimento dos informantes, optaremos por designar estes mesmos casos como “formas estabilizadas”, que assim contribuem para identificar as principais dificuldades de um aprendente tardio no seu processo de aquisição/aprendizagem.

CAPÍTULO 2

GÉNERO GRAMATICAL

2.1. INTRODUÇÃO E OBJETIVOS

Este capítulo tem como principal objetivo a definição e a descrição da categoria gramatical Género (GEN). Num primeiro momento, apresentar-se-ão as características gerais dos sistemas de GEN das línguas do mundo, definindo-se, assim, esta categoria gramatical e as diversas problemáticas correlacionadas com a sua descrição (secção 2.2).

Depois de assinalar o estatuto singular e particular do GEN, atendendo à forma como atua em vários sistemas linguísticos, procede-se, em seguida, à análise do sistema do GEN nominal do português, análise essa em que se evidenciam as principais características deste sistema, bem como os critérios que subordinam a marcação dos valores de GEN nesta língua (secção 2.3).

2.2. PARA UMA DEFINIÇÃO DO CONCEITO DE GÉNERO GRAMATICAL

A palavra *género* (GEN) tem origem no étimo latino *genus*, *-eris*, termo que se traduz em português por “classe” ou “tipo”. É de notar que, do ponto de vista gramatical, se utiliza esta designação para referir uma categoria com relevância para a classificação dos nomes¹³.

O GEN representa, sem dúvida, uma das categorias gramaticais mais singulares, uma vez que não possui um estatuto universal. Nichols (1999), ao analisar um total de 174 línguas, concluiu que apenas uma pequena parte possuía um sistema de GEN ou outro tipo de classificação nominal. Também Greville Corbett (1991) afirma que, enquanto o GEN é central para algumas línguas, noutras está completamente ausente.

O GEN apresenta, ainda, um estatuto particular, visível pela multiplicidade de critérios que cada língua seleciona e que servem de base à definição da categoria. A este respeito, Mário Vilela (1973) constata que, para além da oposição masculino/feminino, “encontramos, nas diferentes línguas, outros critérios classificadores, como o que supõe a existência ou não existência de vida (animados e inanimados), o que se baseia na hierarquização dos seres (seres humanos, bichos animais, frutos, etc. ou seres superiores e seres inferiores), ou o que se refere ao carácter alienável ou inalienável dos seres e das coisas” (1973:141).

Para além dos sistemas de classes de GEN, é possível reconhecer, nas línguas do mundo, a existência de sistemas de classes nominais. Segundo a proposta de Corbett (1991), a marcação do GEN de um nome está dependente de dois tipos de informação: o significado do item nominal e a respectiva forma. Tais propriedades (semânticas e formais), que servem de base à definição do conceito de GEN gramatical, só podem ser consideradas numa língua se, nela, despoletarem o fenómeno da concordância sintática. Este facto permite distinguir os sistemas de classes de GEN dos sistemas de classes

¹³ Convém salientar que a marcação do GEN pode, nas várias línguas do mundo, afetar diferentes classes de palavras por via da sua necessidade de concordância em GEN com o nome. Em português, por exemplo, a marcação do GEN afeta determinantes, quantificadores, adjetivos e alguns pronomes. Já noutras línguas, o GEN pode ser marcado também em verbos, tal como evidencia o estudo sobre a diversidade linguística de Nichols (1999:161). Também Herculano de Carvalho (2000) verifica que nas línguas semíticas, como o árabe, o verbo ostenta uma marca morfológica de GEN que serve para identificar a classe (de GEN) a que pertence o sujeito, i.e., o nome. Assim sendo, o que estes dados demonstram é que o GEN não deixa de ser, primordialmente, uma propriedade nominal.

nominais¹⁴. Deste modo, e nas palavras de Hockett (1958), as categorias de GEN “are classes of nouns reflected in the behavior of associated words.”(1958:231).

Para além da variabilidade de critérios existente nos diferentes sistemas linguísticos, a verdade é que definir um conjunto de fatores de marcação de GEN aplicável a todos os casos numa única língua pode ser igualmente uma tarefa muito difícil. Assim sendo, vejamos em que medida é possível afirmar o carácter variável e idiossincrático da categoria do GEN.

Defende Corbett (1991), e como se referiu, que a atribuição dos valores de GEN pode depender dos seguintes critérios: por um lado, o critério semântico, em que a marcação do valor de GEN está diretamente associada ao conteúdo referencial do nome e, por outro, o critério formal, dizendo respeito a regras do tipo morfológico e fonológico, associáveis aos valores de GEN.

Tendo em conta os critérios de marcação de GEN que cada língua seleciona, o autor de *Gender* analisa um conjunto de duzentas línguas a partir das quais propõe uma tipologia, postulando a existência de: i) sistemas de marcação de GEN puramente lexicais, visto que a marcação dos valores de GEN se realiza através da aplicação de critérios semânticos¹⁵; ii) sistemas de marcação de GEN simultaneamente lexicais e gramaticais/formais; iii) e a possível existência de sistemas de GEN exclusivamente gramaticais/formais¹⁶.

No que diz respeito aos critérios semânticos, e, no caso dos nomes sexuais por oposição aos nomes não-sexuais, existe, na verdade, uma certa base semântica “natural” na atribuição do GEN aos nomes, dadas as características da entidade designada pelo item nominal.

¹⁴ Nas línguas africanas, por exemplo, predomina o sistema de classes nominais que se caracteriza pela presença dos classificadores, i.e., partículas morfológicas (p.ex. prefixos) agregadas ao nome e que identificam a classe a que estes pertencem (cf. Payne, 2003:109; Ibrahim, 1973). Estes sistemas também se encontram em línguas de outros continentes, como na Austrália, Ásia e América. (cf. Payne, 2003:109).

¹⁵ Um sistema de marcação de GEN estritamente semântico é o que ocorre, por exemplo, em Tamil. Nesta língua, nomes com referentes sexuais recebem o valor de GEN (masculino ou feminino) de acordo com o sexo do referente e os nomes não-sexuais são do GEN neutro. Cf. Corbett, 1991:9.

¹⁶ Ao estudar os sistemas de marcação de GEN, Corbett (1991) conclui que não existe nenhuma língua que apresente um sistema de marcação de GEN estritamente formal, i.e., em que se apliquem somente os critérios formais. Tal conclusão demonstra que os sistemas de GEN possuem, em maior ou menor grau, uma estreita ligação com os critérios semânticos. Cf. Corbett, 1991:63.

Numa fase recuada do primitivo Indo-Europeu, a marcação do GEN nos nomes estava relacionada com o carácter [\pm animado] do seu referente. Os valores de masculino e feminino estavam associados ao traço [+ animado] e o neutro associava-se ao traço semântico [-animado] (Gouveia, 2004). Ainda que se tivesse perdido a associação estreita entre os valores masculino/feminino e o carácter [+ animado] do nome, por um lado, e, por outro, entre neutro e [-animado], subsistiram em Latim e no Grego três géneros distintos (masculino, feminino e neutro) que correspondiam às três classes principais em que os nomes se podiam dividir.

Na transição do Latim Clássico para o Latim Vulgar, o valor de neutro foi-se perdendo gradualmente, podendo tal fenómeno, em grande parte, compreender-se pelas coincidências entre as desinências morfológicas do neutro e do masculino (Gouveia, 2004). Quando se passa para as línguas românicas, o GEN neutro desaparece totalmente¹⁷, e prevalecem os GEN masculino e feminino¹⁸.

A este respeito convém salientar que a distinção entre masculino e feminino que corresponde a uma distinção do *género natural* não deve ser confundida com o conceito de GEN gramatical. O *género natural* corresponde ao sexo das entidades [+ animado] do mundo real, ao passo que o GEN gramatical, além da ocasional correspondência com o dos seres sexuados, representa um sistema de classificação gramatical dos nomes que pode, conforme as línguas, não ter qualquer ligação com o *género natural*.

As evidências para o que se acaba de afirmar são várias. Em primeiro lugar, se existem nomes sexuados do GEN masculino por se referirem a uma entidade do sexo masculino (em português por exemplo: *menino*, *rapaz*) e nomes femininos cujo referente é uma entidade do sexo feminino (*menina*, *rapariga*), existem também nomes cujo valor de GEN não corresponde necessariamente ao sexo dos seus referentes. Em português, há nomes sexuados que não apresentam esta correspondência, como, por exemplo, os sobrecomuns, que são sempre ou masculinos ou femininos, independentemente do sexo dos referentes (*a pessoa*, *a testemunha*) (cf. secção 2.3.2).

¹⁷ Esta mudança é, no entender de Maria Carmen Gouveia (2004:450): “uma das mais importantes ocorridas no Latim Coloquial tardio”.

¹⁸ Numa visão diacrónica do sistema de GEN em português, Vilela (1973) verifica que, com o desaparecimento do GEN neutro, os nomes cuja desinência coincide formalmente com a masculina passaram a ser do GEN masculino e nomes cuja desinência se assemelha formalmente à feminina tornaram-se femininos. Refere o autor: “os neutros, no singular, cujo tema terminava em -o, tomaram o género masculino; os terminados em vogal que não seja -o, ou terminados em consoante, tornaram-se masculinos ou femininos; os nomes provenientes do plural neutro, por causa da terminação em -a, tomaram o género feminino, embora conservando esta noção de reunião ou de ser colectivo, como se nota em *senha*, *boda*, *fruta*, etc.” (Vilela, 1973:146).

Já no alemão, o neutro abrange alguns nomes cujo referente é sexuado: *das Kind* ‘a criança’, *das Weib* ‘a mulher/esposa’, *das Fräulein* ‘a menina, não casada’, facto que, mais uma vez, comprova a ausência de qualquer ligação entre a marcação do GEN gramatical e o *género natural* dos referentes nominais (Bechara, 1999).

Em segundo lugar, nas diferentes línguas, um dado referente pode ser designado por nomes com valores de GEN distintos. Esta divergência pode-se verificar mesmo nas línguas com relação histórica entre si e que pertencem à mesma família, como observa Thomas E. Payne (2003:108): “Romance languages even differ from one another as to the class that particular lexical items fall into, e.g., Italian *il tavolo* (m.), French *la table* (f.); Italian *il mare* (m.), French *la mer* (f.)”, demonstrando que a distribuição dos valores de GEN nas línguas é verdadeiramente inconsistente. Em português, por exemplo, o nome *mar* é masculino, i.e., possui um valor de GEN distinto em relação a nomes correspondentes das restantes línguas românicas. Não será de estranhar, portanto, que a inconsistência na atribuição de valores de GEN ocorra igualmente em línguas que não pertencem ao mesmo tipo: o nome *sol* é masculino em português, ao passo que, no alemão, *Sonne* é do GEN feminino (*die Sonne*); *lua* é um nome feminino em português, mas, em alemão, *Mond* é masculino (*der Mond*) (cf. Bechara, 1999:133).

Também há línguas que perderam, no decurso da sua história, a categoria gramatical de GEN. O sistema de GEN do inglês tem sofrido transformações radicais ao longo da sua evolução e algumas das suas manifestações sintáticas e morfológicas foram-se perdendo (Ibrahim, 1973:24).

Por fim, acrescente-se ainda o facto de o número de categorias de GEN ser variável de língua para língua. Corbett nota que línguas do mesmo tipo (como as que descendem do Indo-Europeu) “show gender (some with three genders, others having reduced the number to two); a few have lost gender, while others, notably the Slavonic group, are introducing new subgenders” (1991:2)”. Todos estes casos são, assim, ilustrativos da não correspondência entre *género natural* e GEN gramatical. Ao mesmo tempo, estes dados confirmam, uma vez mais, a concordância sintática como fator determinante para o estabelecimento de um sistema de classes de GEN¹⁹. Assim sendo, há tantos géneros numa língua quantas as possibilidades de concordância sintática desencadeadas pelo nome.

¹⁹ Também Mário Vilela (1973) reconhece que cabe à concordância e ao artigo o principal papel na identificação do GEN do nome.

Tendo em conta o que ficou referido, poder-se-á concluir que, se nos nomes sexuais existe um critério subordinante na sua associação a um determinado valor de GEN, i.e., o critério semântico, a verdade é que a correspondência entre *género natural* e GEN gramatical nem sempre se verifica.

Além do critério semântico, há línguas que utilizam critérios formais para a marcação dos valores de GEN. Nos nomes não-sexuais, sobretudo, Corbett (1991) reconhece a existência de fatores formais que condicionam a marcação do GEN. Esses critérios são constituídos tanto por regras do tipo morfológico – que têm em conta a estrutura mórfica da palavra e compreendem mecanismos morfológicos tais como a flexão e a derivação –, como também por regras do tipo fonológico.

Todavia, apesar de identificar a existência destas regras em várias línguas, o mesmo autor verifica que nem sempre é fácil distinguir entre o seu estatuto morfológico e/ou fonológico. Assim, associa os critérios de ordem morfológica às situações em que, para a marcação dos valores de GEN, é necessária a referência a mais do que uma forma morfológica do nome. A título de exemplo, Corbett (1991) descreve o sistema de GEN da língua russa e conclui que, em grande parte dos nomes, existe uma relação entre os valores de GEN e a declinação a que pertencem. Por conseguinte, propõe a regra: “nomes que pertencem à declinação α possuem β como valor de GEN”²⁰.

Corbett (1991) salienta que, nestes sistemas, convém considerar o papel que os critérios de natureza semântica têm. Por um lado, os critérios formais atuam quando os critérios semânticos não são suficientes e, por outro, os sistemas de marcação semânticos e morfológicos podem, em alguns casos, ser coincidentes (cf. Corbett, 1991:34). No caso específico da língua portuguesa, a aplicação de regras do tipo morfológico apresenta, em parte, alguma vitalidade. Alguns sufixos derivacionais, por exemplo, possuem informação de GEN e determinam se os produtos nominais serão masculinos ou femininos, o que revela a existência de correlação entre estes itens morfológicos e o valor de GEN dos nomes que os integram (cf. secção 2.3.2).

No que concerne aos critérios do tipo fonológico propostos por Corbett, estes podem traduzir-se através de uma regra do seguinte tipo: “nomes cujo segmento final α possuem β como valor de GEN.” Nestes sistemas existe somente uma forma morfológica do nome e o que determina a diferenciação do GEN é a alteração dos

²⁰ Este critério não é, no entanto, aplicável a todos os casos, porque o sistema de atribuição de GEN da língua russa apresenta, de acordo com o autor, algumas exceções (cf. Corbett, 1991:34).

segmentos fonológicos do item nominal. Além disso, a posição da sílaba tónica é também um fator que pode condicionar o valor de GEN nominal. O investigador analisa alguns sistemas de marcação fonológica, que se podem encontrar em línguas como o *Qafare* a língua *Hausa* (da família das línguas Afro-Asiáticas). Também a língua francesa apresentará um sistema de marcação de GEN fonológico: nomes como *mutualité*, *activité*, *singularité* são sempre do GEN feminino e *voyage*, *courage*, *âge*, *message*, *massage* são masculinos (cf. Corbett, 1991: 51-61).

Sintetizando, compreende-se, através desta breve exposição das principais características dos sistemas de classificação de GEN, o estatuto variável desta categoria gramatical. Tendo em conta o que ficou referido nesta secção, veremos, em seguida, a descrição do sistema de GEN nominal do português, e as diferentes particularidades que apresenta.

2.3. O GÊNERO NOMINAL EM PORTUGUÊS

2.3.1. UMA CATEGORIA DA FLEXÃO?

Em português, a categoria de GEN está presente em todos os nomes e em alguns pronomes. Apesar de ser uma propriedade dos nomes, os valores de GEN não são facilmente dedutíveis a partir da observação das características fonomorfológicas e mesmo semânticas destes. Por isso, importa analisar o sistema em português, sublinhando as suas principais particularidades e as problemáticas relacionadas com a sua descrição.

Existem dois valores de GEN em oposição no português: o masculino e o feminino. A tradição gramatical portuguesa tem tratado o GEN como uma categoria da flexão nominal (Vilela, 1999; Cunha e Cintra, 2005), embora se encontrem posições discordantes sobre esta questão, como a de Herculano de Carvalho (1979) e a de Villalva (1994), que negam a existência da flexão do GEN na classe nominal. Deste modo, é discutível que o GEN se considere uma categoria flexional em português, uma vez que, para tal, é preciso comprovar que obedece a determinadas regras características dos processos flexionais.

Admitindo que a questão da natureza flexional do GEN é exposta de uma maneira incoerente e confusa nas gramáticas tradicionais, Mattoso Câmara (1994:88) defende, em alternativa, que o GEN é uma categoria formal e não semântica, reconhecendo a existência de:

- i) Nomes substantivos com dois valores de GEN e com flexão redundante, i.e., o GEN está marcado tanto no nome como no artigo que o antecede: “*o lobo/ a loba; o menino/ a menina*”;
- ii) Nomes substantivos com dois valores de GEN sem flexão “*o/a modelo; o/a artista*”;
- iii) Nomes substantivos com um único valor de GEN “*a tribo/ o fantasma/ o pente/ o papel/ a dissertação*”.

Portanto, segundo este autor, existe flexão de GEN nos pares de palavras como *menino/menina, professor/professora, mestre/mestra*, visto que, nestes casos, existe

uma forma do masculino a que se opõe uma outra do feminino. Assim, propõe que a marcação do feminino se caracteriza pela presença do morfema de GEN feminino – fonologicamente /a/, somente nas palavras que possuem um correspondente masculino²¹, como *menino/ menina*.

Segundo Mattoso Câmara Jr., o masculino representa a forma não-marcada e o feminino funciona, assim, “como uma particularização mórfico-semântica do masculino, uma forma marcada pela adjunção da desinência /a/”, afirmando, por fim, que se trata “de uma oposição privativa, onde uma forma marcada pela desinência de feminino se afirma em face de uma forma não-marcada, ou de desinência (Ø) para o masculino.” (1966:3). Deste modo, para o gramático, o GEN corresponde à distribuição das palavras em classes, sendo o masculino “uma forma geral, não-marcada, e o feminino indica uma especialização qualquer” (Câmara, 1994:88).

A desinência do feminino já não está, de acordo com a proposta de Câmara (1994), presente em nomes ditos comuns de dois géneros, i.e., que possuem diferente valor de GEN de acordo com o contexto em que ocorrem (*o/a jornalista; o/a artista; o/a estilista*), nem em substantivos que possuem um único valor de GEN (*a carrinha,/ a mesa*), uma vez que estes itens não apresentam uma forma masculina oposta (Câmara, 1966).

No entanto, observando o universo dos nomes da língua portuguesa, verifica-se que, além daqueles que terminam em *-o* ou em *-a* (vogais átonas), existem outros que admitem diferentes terminações e que tanto podem ser masculinos como femininos. Assim, temos em português nomes terminados em *-e* átono (*o pente/ a ponte*); em vogal tónica (*o peru/ a fé/ o sopé*); em consoante (*o chafariz/ a perdiz; o mar/ a colher*); e em vogal ou ditongo nasal (*o coração/ a modificação; o pajem/ a nuvem*) (Martins, 2008).

Acrescente-se que, apesar de todos os nomes em português terem um valor de GEN gramatical, nem todos admitem contrastes de GEN, mesmo aqueles que terminam em *-o/-a*: *a mesa/*o meso; o tronco/*a tronca*. Estes dados demonstram que os únicos nomes que admitem contrastes de GEN são os que possuem o traço semântico [+sexuado], e mesmo esses podem materializar tais contrastes sem recurso à oposição *-o/-a*.

²¹ Segundo Mattoso Câmara (1966:4): “não é simplesmente a presença de um /a/ átono final que assinala o feminino; é a sua presença em face da forma oposta, masculina, sem ela”.

De facto, quando os nomes sexuados participam em contrastes de GEN, os mecanismos que os asseguram podem ser diversos. Assim sendo, a oposição masculino/feminino pode ser assegurada por processos estritamente lexicais: (*o homem/ a mulher*); por contraste de Índice Temático (IT) (cf. p. 30): *o menino/ a menina*; e também por processos morfológicos, como a composição: *o corvo-macho/ o corvo-fêmea*; e a derivação²²: *o imperador/ a imperatriz, o galo/ a galinha*. Nos casos em que o contraste de GEN é assegurado através de um processo derivacional, como em *o conde/ a condessa*, poder-se-á admitir que o feminino representa uma especialização de sentido do nome masculino, visto que o derivado *condessa* representa a mulher do *conde*. Na perspectiva de Mattoso Câmara (1966) é possível admitir que existe variação de GEN nos pares de palavras como *jarro/jarra, barco/barca, fruto/fruta*. Para além da oposição masculino/feminino se processar através da oposição *-o/-a*, essa alteração do valor de GEN implica, simultaneamente, uma alteração semântica do nome, evidenciando-se, assim, um comportamento derivacional.

Além disso, mesmo entre os nomes com referentes sexuados encontramos exemplos no português em que os contrastes de GEN não são possíveis. Existe um pequeno conjunto de nomes sexuados, os sobrecomuns, que denotam indistintamente entidades do sexo masculino e entidades do sexo feminino, embora possuam um único valor de GEN: *o indivíduo, a pessoa, a testemunha, o cônjuge*.

Há igualmente um grupo de nomes, os epicenos, que designam animais e que dispõem de um único valor de GEN independentemente do sexo do referente. Nestes casos, para estabelecer o contraste de GEN, e como atrás de viu, associa-se à base nominal (*águia*) a forma *-macho (águia-macho)* ou *-fêmea (águia-fêmea)*²³ criando-se, assim, um composto sintático por subordinação que especifica, não o valor de GEN do nome, uma vez que o GEN se mantém igual (*a águia-macho; a águia-fêmea*), mas o sexo do referente (Villalva, 2008).

Relembre-se ainda o conjunto de nomes sexuados que possuem dois valores de GEN: os nomes comuns de dois géneros, que apresentam uma forma ambígua quanto ao

²² Na verdade, o facto dos contrastes de GEN se poderem realizar através do recurso a sufixos derivacionais (*o conde/ a condessa; ojogral/ a jogralesca*), confirma a natureza não-flexional do GEN dos nomes em português e, além disso, “sublinha a natureza lexical” (Rio-Torto, 2002) desta categoria gramatical.

²³ Ao estudar o comportamento do género nos nomes que designam animais, Villalva (2008:101) conclui que, ao lado “dos casos mais frequentes (...) dos radicais nominais, masculinos e femininos, que referem indistintamente os machos e as fêmeas de cada espécie e que são chamados epicenos”, existem alguns nomes referentes a animais domésticos ou familiares que dispõem de formas distintas para o masculino e para o feminino (*gato/gata; cão/cadela; leão/leoa*).

valor de GEN, sendo essa ambiguidade resolvida através do contexto sintático: *o/a jovem; o/a estudante; o/a artista*.

Assim como demonstram os exemplos evocados, o tratamento do GEN não pode ser o de uma categoria flexional, uma vez que a flexão, tal como se vai constatar, tem de ser, necessariamente, um sistema coeso e coerente (Câmara, 1994).

A flexão corresponde a um processo morfológico (que diz respeito ao domínio da palavra), obrigatório, quer na sua formação, quer na sua utilização (Villalva, 2008), que obedece a um modelo sistemático e geralmente regular. Este mecanismo caracteriza-se pelo uso de paradigmas coesos, i.e., variações que se repetem quase sempre da mesma maneira, existindo, porém, algumas exceções que evidenciam uma pequena variação alomórfica. Perante esta descrição, poder-se-á admitir que o processo de pluralização dos nomes, que consiste em agregar o morfema de plural *-s* ao tema nominal²⁴, ou seja, ao constituinte composto pelo radical e pelo Índice Temático (IT), claramente se define como um mecanismo flexional.

Quanto ao GEN, exemplos de nomes como *menino/ menina, professor/ professora* poderiam confirmar a existência de um mecanismo flexional, tal como defendeu Mattoso Câmara (1966, 1994). Contudo, à exceção de casos como este, em reduzido número, a variação em GEN dos nomes em português não possui as características acima enunciadas, facto que ficou ilustrado através da análise dos contrastes de GEN possíveis para o português. Além disso, tendo em conta a própria estrutura mórfica do nome, é possível confirmar a natureza não-flexional desta categoria gramatical.

Na verdade, no que respeita à estrutura formal do nome, regra geral associa-se o valor de GEN às vogais átonas (*-o, -a*) que surgem à direita do radical (simples ou derivado), ou seja, é comum associar-se o GEN masculino aos nomes terminados em *-o* e o feminino aos nomes terminados em vogal átona *-a*. No entanto, apesar de apresentarem uma correspondência parcial com o GEN dos nomes que as integram, estas vogais não correspondem a morfemas de GEN. De facto, existem em português nomes terminados em *-a* que pertencem ao GEN masculino e nomes femininos terminados em *-o* (cf. Quadro 2.1). Acrescente-se ainda, e como já foi referido, a

²⁴ No Dicionário Terminológico do Ministério da Educação (DT) o tema nominal tem a seguinte definição: “(...) é um dos constituintes imediatos do nome e domina o radical nominal e o índice temático. O tema nominal é especificado pela flexão em número”.

existência de nomes com diferentes terminações e que podem estar associados aos dois valores de GEN (cf. Quadro 2.1).

Por conseguinte, o papel de estas vogais (-o, -a) à semelhança do que ocorre com as vogais temáticas dos verbos, é o de classificar os nomes. Nesta medida, designar-se-ão Índices Temáticos (IT) ou Vogais Temáticas (VT). Veja-se, em seguida, duas propostas de classificação dos nomes em português, tendo em conta as classes temáticas a que estes itens podem pertencer.

Segundo a proposta de classificação de Mattoso Câmara (1994), a vogal átona final dos substantivos representa uma vogal temática²⁵. De acordo com a classe temática a que pertencem, os substantivos distribuem-se em: i) nomes substantivos de tema em -o (*cogumelo*, *jarro*); ii) nomes substantivos de tema em -a (*programa*, *revista*); iii) nomes substantivos de tema em -e (*mestre*), ou em -e teórico (*professor*) (visto que esta vogal, não sendo visível na forma do singular do nome (*cantor*), aparece, no entanto, no plural (*cantores*) (Câmara, 1994); iv) nomes substantivos atemáticos (*corrimão*).

Por sua vez, Alina Villalva (2003), que prefere o termo “Índice Temático” propõe um sistema de cinco categorias básicas: i) nomes de tema em -a (*mesa*); ii) nomes de tema em -o (*menino*); iii) nomes de tema em -e (*mestre*); iv) nomes de tema \emptyset ²⁶ (morfema zero, em nomes que terminam em consoante: *apresentador*); v) nomes atemáticos (quando terminam em vogal tónica: *chimpanzé*). A partir destas macrocategorias, a autora chega a estabelecer uma tipologia de 23 classes temáticas, através da aplicação de outros critérios, como a oposição de formas variáveis e invariáveis, i.e., nomes que participam em contrastes de GEN e nomes que não admitem tal oposição (cf. Villava, 2003).

Considerando estas propostas, veja-se o seguinte Quadro (Quadro 2.1), em que estão representados os nomes e a relação do IT com o valor de GEN que possuem.

²⁵ Para este autor a única exceção é a vogal -a que corresponde a uma forma do GEN feminino (cf. com o que ficou referido na p. 27 deste Capítulo).

²⁶ Para Mattoso Câmara (1994) estes nomes classificam-se como nomes substantivos de tema em -e (-e teórico).

Índice Temático (IT)	GEN Masculino		GEN Feminino	GEN Masculino/ Feminino
-o	o menino, o cabelo, o rádio (<i>aparelho</i>), o mundo, o tronco, o prédio, o ministro		a tribo, a rádio (<i>estação</i>)	o/a modelo
-a	o dia, o cinema, o panorama		a menina, a rapariga, a banca, a mesa, a história, a fotografia, a leitura	o/a artista
-e	-e	o mestre, o teste	a mente, a localidade, a solidariedade	o/a estudante
	-e [teórico]/ Ø	o cantor, o professor, o compositor	a imperatriz, a luz	o/a mártir
atemáticos	o avô, o cão, o irmão, o pão, o lápis, o coração		a avó, a manhã, a chaminé, a paisagem, a modificação	o/a selvagem

Quadro 2.1 – Distribuição dos nomes em português quanto ao IT

Da análise do Quadro 2.1 depreende-se que o IT não condiciona o valor de GEN do nome, visto que independentemente da classe temática a que pertençam, os nomes podem ser do GEN masculino e feminino.

Em síntese, se, para muitos autores, o GEN em português se caracteriza por ser um mecanismo tipicamente flexional, a verdade é que o sistema de marcação de GEN nominal em português revela características que contrariam esta perspetiva. Subscreeve-se, nesta medida, a proposta de Alina Villava ao afirmar que “o género, nos nomes, não é uma categoria de flexão, mas sim uma categoria morfo-sintáctica cuja especificação é lexicalmente determinada ou resultante da intervenção de um processo morfológico não-flexional.” (1994: 233).

2.3.2. CRITÉRIOS DE MARCAÇÃO DE GÉNERO DOS NOMES EM PORTUGUÊS

Na secção 2.2 foram enumerados os critérios que, de acordo com a análise de Corbett (1991), se podem encontrar na organização dos sistemas de marcação de GEN nas línguas do mundo. Sendo assim, é pertinente verificar a aplicação da proposta de Corbett (1991) no sistema de marcação do GEN nominal em português. Por isso, e também em função dos dados apresentados em 2.3.1., de um lado serão tratados os nomes correspondentes a referentes sexuais e, do outro, os nomes correspondentes a referentes não sexuais.

Em relação aos nomes sexuais, considerem-se, primeiro, os casos em que o valor de GEN gramatical coincide com o valor de *género natural* da entidade designada. Deste modo, a forma do GEN masculino designa uma entidade do sexo masculino (cf. *menino*) e a forma do feminino refere-se a uma entidade do sexo feminino (cf. *menina*). Neste grupo, a marcação de GEN corresponde à aplicação do critério semântico, que pode, em alguns casos, coincidir com o critério formal: o nome sexual *aluno* designa uma entidade masculina e termina com o IT tipicamente associado ao GEN masculino, ou seja, IT *-o*.

Apesar de se reconhecer nos nomes sexuais a aplicação do critério semântico, ocorrem as seguintes restrições²⁷:

- i) Formas nominais masculinas que se referem a mulheres (*mulherão*) e formas nominais femininas que se referem a homens (*bicha*) (Villalva, 2008);
- ii) Nomes sexuais que possuem um único valor de GEN independentemente do sexo do referente (sobrecomens e epicenos) (cf. 2.3.1).

Relativamente aos nomes não-sexuais, a atribuição do GEN, por não estar ligada a um conteúdo referencial específico, é, à luz deste critério semântico, imprevisível e aleatória. No entanto, considerando a proposta de Corbett (1991), seria

²⁷ A este respeito, M. Carmen Gouveia conclui que: “De facto, criar-se-á uma perspectiva muito mais clara para a descrição do género em Português se se tiver em consideração que, embora possa efectivamente haver – nalguns casos – uma relação entre o género gramatical e o sexo, ela não é exclusiva e total.” (2005: 204).

eventualmente possível prever o GEN destes nomes, atendendo à respectiva estrutura formal. Veja-se, assim, em que medida se podem aplicar algumas das regras sugeridas pelo autor aos nomes não-sexuados.

Conforme se disse na secção 2.3.1., é comum associar-se os IT *-o* e *-a* dos nomes aos seus valores de GEN, i.e., masculino e feminino respectivamente (cf. 2.3.1). Portanto, poder-se-ão agrupar os nomes não-sexuados da seguinte forma: de um lado, nomes cujo valor de GEN corresponde ao respectivo IT (*-o/-a*); e do outro, nomes cujo valor de GEN não corresponde ao IT (cf. Quadro 2.2).

1. Nomes não-sexuados cujo IT corresponde ao GEN	
IT -o / GEN Masculino	IT -a / GEN Feminino
<i>o colégio, o estágio, o mosteiro, o caderno</i>	<i>a correia, a mala, a pastilha, a escola</i>
2. Nomes não-sexuados cujo IT não corresponde ao GEN	
IT -a / GEN Masculino	IT -o / GEN Feminino
<i>o mapa, o dia, o fantasma, o telegrama</i>	<i>a foto, a tribo, a moto, a rádio (estação)</i>

Quadro 2.2 - Distribuição dos nomes não-sexuados quanto ao IT e respetivo valor de GEN

Na tentativa de apurar a distribuição relativa dos nomes de GEN masculino e feminino por cada classe temática foram considerados os índices de frequências dos nomes em português, a partir das ocorrências dos lemas recenseados no *corpus* de frequências lexicais do português europeu contemporâneo intitulado “Léxico Multifuncional Computorizado do Português Contemporâneo” (CORLEX).²⁸

Nesta análise teve-se em conta somente os itens nominais com radicais simples e derivados que não são compostos ou possam ser classificados como unidades fraseológicas, i.e., não foram contabilizadas as ocorrências de nomes como: “*pó-de-arroz, à vontade*”.

Para a análise procedeu-se à contagem das frequências dos nomes do GEN masculino e feminino, em função de três subtipos temáticos: *-o*, *-a* e outros.

Assim, a partir do universo total de frequências dos nomes selecionados (3.611.198), calculou-se a percentagem das ocorrências dos itens nominais com IT *-o*, -

²⁸A CORLEX pode ser consultada em <http://www.clul.ul.pt/pt/component/content/article/194-multifuncional-computational-lexicon-of-contemporary-portuguese>.

a e dos itens que pertencem a outras classes temáticas, com base no valor de GEN associado a cada nome. Veja-se o Quadro 3.3 com os valores de frequências lexicais em função das classes temáticas e do GEN que possuem e os Gráficos 2.1, 2.2, 2.3 e 2.4 que apresentam os valores percentuais.

GEN e classe temática dos itens nominais	Frequência de nomes
Nomes do GEN masculino com IT <i>-o</i>	1.156.695
Nomes do GEN masculino com IT <i>-a</i>	56.362
Nomes do GEN feminino com IT <i>-a</i>	971.863
Nomes do GEN feminino com IT <i>-o</i>	1.613
Nomes comuns de dois GEN com IT <i>-o</i>	1.127
Nomes comuns de dois GEN com IT <i>-a</i>	25.982
Nomes do GEN masculino de outras classes temáticas	629.239
Nomes do GEN feminino de outras classes temáticas	706.851
Nomes comuns de dois GEN de outras classes temáticas	61.466
TOTAL	3.611.198

Quadro 2.3 – Frequência dos lemas recenseados na CORLEX em função do GEN e da classe temática dos nomes simples e derivados

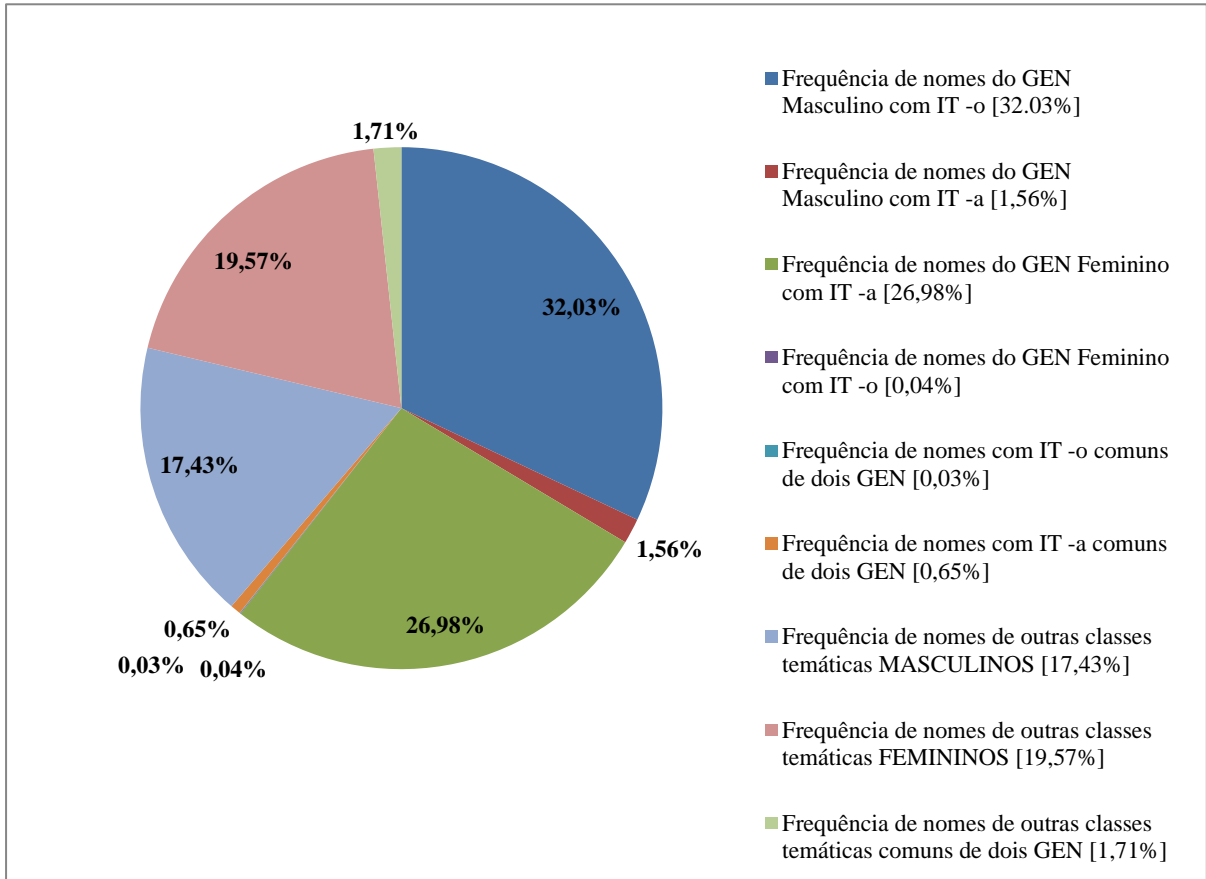


Gráfico 2.1 - Distribuição em percentagem das ocorrências dos itens nominais em função da classe temática e do valor de GEN

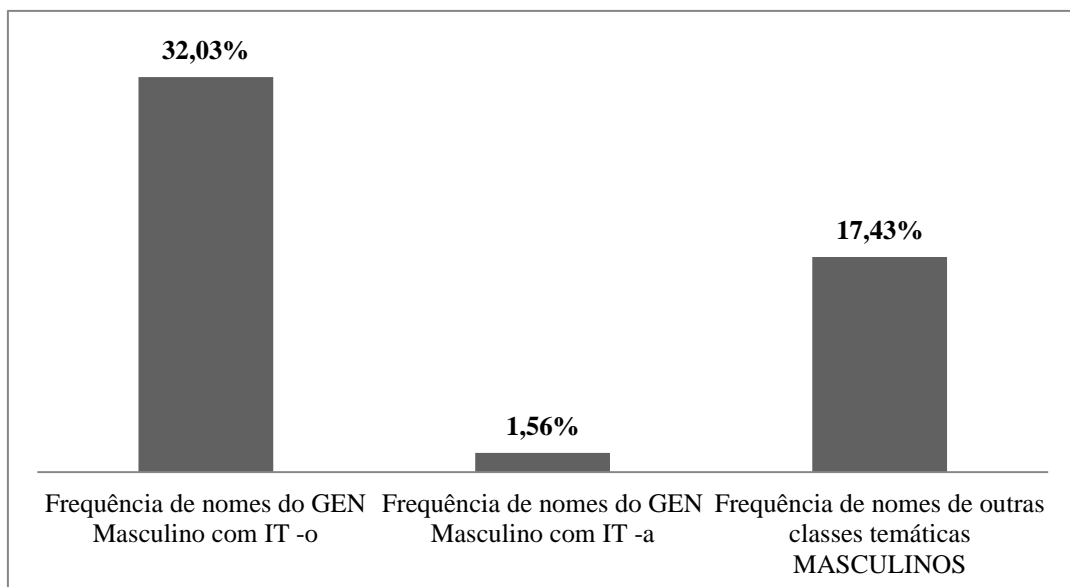


Gráfico 2.2 – Percentagem das ocorrências de nomes masculinos de acordo com a classe temática a que pertencem

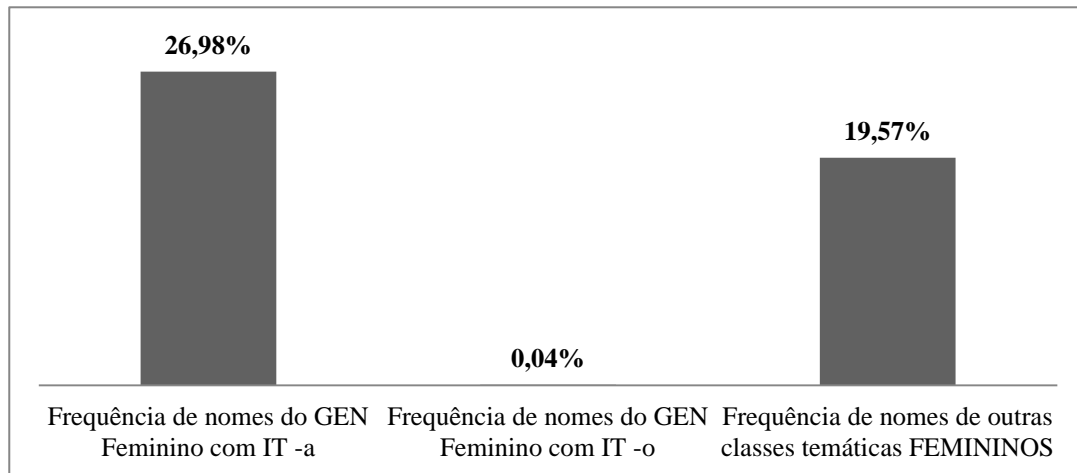


Gráfico 2.3 – Percentagem das ocorrências de nomes feminino de acordo com a classe temática a que pertencem

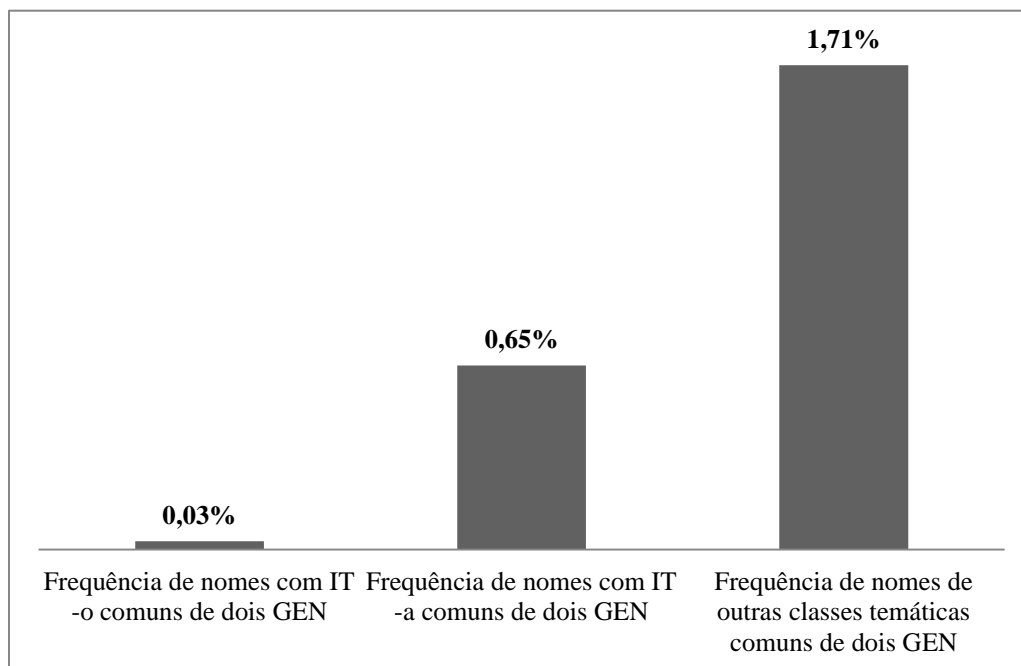


Gráfico 2.4 – Percentagem das ocorrências de nomes comuns de dois GEN de acordo com a classe temática a que pertencem

No Gráfico 2.2 apresenta-se a percentagem da frequência de nomes masculinos de acordo com a classe temática a que pertencem esses itens. Como se pode verificar, a percentagem da frequência de nomes de IT -o do GEN masculino é de 32.03%, e a percentagem da frequência de nomes com IT -a do GEN masculino, é de,

aproximadamente, 1.56%. Já a percentagem das ocorrências dos restantes nomes masculinos, de outras classes temáticas, é de 17.43%.

Em 2.3 está representada a percentagem das ocorrências de nomes do GEN feminino. A percentagem da frequência de nomes de IT *-a* do GEN feminino é de 26.98%, e a percentagem da frequência de nomes com IT *-o* femininos, é de 0.04%. A percentagem das frequências dos restantes nomes femininos é de 19.57%.

Por fim, o último gráfico (Gráfico 2.4) apresenta a percentagem das frequências dos itens nominais simples comuns de dois GEN. A frequência de nomes com IT *-o* com dois valores de GEN é de 0.03%, e com IT *-a* é de 0.65%.

Da análise destes resultados verifica-se que, em cerca de 60% das ocorrências nominais em português se verifica uma relação parcial entre o IT e o respectivo GEN. Logo, para o aprendente de PL2 este pode ser um critério que facilite a marcação do valor de GEN destes itens. Por essa razão, será interessante verificar qual a frequência de desvios produzidos pelos informantes deste estudo em nomes cujo IT corresponde ao seu valor de GEN (cf. Capítulo 3).

Ainda a propósito dos critérios formais, recorde-se que Corbett (1991) defende a eventualidade de também as regras morfológicas poderem condicionar a marcação de GEN dos nomes não-sexuados, sendo este um critério operante em português. Tal como se verificou, o IT é um constituinte morfológico que, apenas até certo ponto, auxilia na marcação do valor de GEN dos nomes. Considere-se, todavia, que existem outros constituintes morfológicos através dos quais se pode recuperar informação sobre o GEN do nome. Esses constituintes são os sufixos derivacionais que, para além de determinarem a categoria sintática do produto final²⁹, possuem informação de GEN quando geram produtos nominais (Villalva, 2003), modificando, em determinados casos, o GEN da base nominal a que se associam: por exemplo, à base *livro* – nome do GEN masculino –, pode agregar-se o sufixo *-aria*, gerando assim um produto nominal do GEN feminino: *livraria*.

²⁹ Ao adicionar o sufixo derivacional [-ção] ao tema verbal [varia] obtém-se um nome: *variação*, e se acrescentar ao mesmo tema o sufixo [-vel] o produto final será adjetival e não nominal: *variável*. Sobre este assunto, veja-se: Rio-Torto, 1998.

Assim, existem sufixos derivacionais responsáveis pela formação de nomes masculinos e sufixos derivacionais que originam nomes femininos, como se pode observar no Quadro 2.3 em que estão registados alguns sufixos³⁰ com os respectivos IT. Note-se que, neste Quadro, apenas estão representados os sufixos que não se associam aos IT *-o* e *-a*, uma vez que, nesses casos, a correlação de GEN e IT é perfeita. A título de exemplo, refiram-se os sufixos *-ari]a*: *a padaria, a livraria, a peixaria*; *-ism]o*: *o simbolismo, o patriotismo*.

Sufixos	Nomes do GEN Masculino	Nomes do GEN Feminino
-agem		<i>a contagem, a lavagem, a miudagem</i>
-ame	<i>o vasilhame, o velame</i>	
-al	<i>o areal, o pantanal, o olival</i>	
-ão	<i>o tropeção, o esfregão</i>	
-dad]e		<i>a irmandade, a atlanticidade, a teatralidade</i>
-ção		<i>a atrapalhação, a animação, a elaboração</i>
-dão		<i>a vermelhidão, a escuridão</i>
-dor	<i>o apresentador, o orientador, o gerador</i>	
-ez		<i>a sensatez, a embriaguez, a solidez, a timidez</i>
-ic]e		<i>a garridice, a palermice</i>
-íci]e		<i>a calvície, a imundície</i>
-im	<i>o farolim</i>	
-it]e		<i>a amigdalite, a bronquite</i>
-or	<i>o amargor, o cantor</i>	
-os]e		<i>a neurose</i>
-tud]e		<i>a altitude</i>
-um]e	<i>o negrume, o azedume</i>	
-ugem		<i>a penugem</i>
-nt]e	<i>o ajudante, o servente, o ouvinte</i>	<i>a ajudante, a servente, a ouvinte</i>

Quadro 2.4 - Relação entre os sufixos derivacionais e respectivo GEN dos produtos nominais

Portanto, em português a atribuição de um valor de GEN aos nomes resulta do cruzamento de critérios semânticos e de critérios formais. Tal como ficou demonstrado

³⁰ O inventário dos sufixos resulta da consulta de Rio-Torto, 1998.

com o estudo efetuado nas ocorrências de lemas nominais recenseados na CORLEX, em cerca 60% das frequências nominais existe correlação entre o IT e o respectivo GEN. Assim, são cerca de 40% as ocorrências de nomes em que este critério formal não funciona e, deste modo, não pode ser ativado pelo aprendente de PL2 no processo de atribuição de GEN a esses itens.

2.4. IMPLICAÇÕES PARA O CONTEXTO DE AQUISIÇÃO/APRENDIZAGEM DE PL₂

Como ficou referido no início deste capítulo, Corbett (1991) define um conjunto de critérios para a determinação do GEN nas línguas do mundo, reconhecendo, portanto, a existência de mecanismos que permitem aos falantes da Língua Materna associar a marcação de GEN dos nomes.

No caso específico do português, com a análise do sistema de marcação de GEN dos nomes, verifica-se que existe um sistema de atribuição de GEN em que se verifica, numa avultado número de casos, a existência de regras nas quais o aprendente possa confiar para este efeito.

Referidas as principais especificidades que envolve o sistema de marcação de GEN do português, veja-se no próximo capítulo o tratamento e análise dos dados apurados neste estudo.

CAPÍTULO 3

TRATAMENTO E ANÁLISE DOS DADOS

3.1. INTRODUÇÃO E OBJETIVOS

Referido o enquadramento teórico do presente estudo nos capítulos anteriores, procede-se, neste capítulo, à apresentação e descrição do estudo empírico e à análise dos resultados obtidos.

Assim, numa primeira parte, descreve-se detalhadamente o método de seleção do conjunto de textos que integra o *corpus* que constitui a base empírica deste estudo, bem como o perfil dos informantes, no que respeita à sua distribuição por idade, início do estudo de PL2, LM e nível de proficiência em PL2 (secção 3.2).

Num segundo ponto apresenta-se as características e a descrição detalhada da tipologia de desvios de marcação de GEN que serve de ponto de partida para a análise dos dados apurados (secção 3.3).

Por fim, procede-se à análise detalhada dos resultados, com vista à identificação de diferentes fases do desenvolvimento da interlíngua (IL) dos informantes (secção 3.4).

3.2. METODOLOGIA

3.2.1. DESCRIÇÃO DO CORPUS

Os textos escritos que servem de base empírica desta investigação foram selecionados a partir de um conjunto alargado de produções que integram o “*Corpus de Produções Escritas de Aprendentes de PL2*” (PEAPL2)³¹ (coord. Cristina Martins), projeto iniciado em Junho de 2008, no *Centro de Estudos de Linguística Geral e Aplicada* (CELGA) da Faculdade de Letras da Universidade de Coimbra (FLUC) (cf. Martins, 2011). Note-se que, sempre que se fizer uma referência ao *corpus* geral que serviu de base à constituição do *corpus* desta investigação, utilizaremos a designação “*Corpus PEAPL2*” e para nos referirmos ao nosso *corpus* usaremos, somente, o termo “*corpus*”.

Para a constituição do *corpus* efetuou-se uma análise rigorosa à totalidade das produções escritas disponíveis no *Corpus PEAPL2*, de modo a verificar que textos apresentavam desvios em relação à marcação do GEN dentro do grupo nominal (GN). Este processo era essencial para, numa fase posterior, se proceder à criação de uma tipologia de desvios. Assim, do conjunto total de **546** produções escritas por 391 aprendentes de PL2 que, entre Maio de 2009 e Maio de 2010, frequentaram cursos de português para estrangeiros na FLUC, foram selecionados **322** textos de 256 informantes. O gráfico 3.1 ilustra a percentagem de textos que, no *Corpus PEAPL2*, possuem casos de desvios de marcação de GEN. Como se pode constatar, 59% dos textos apresentam desvios e é esse conjunto que constitui o *corpus* deste trabalho.

³¹ Todas as produções que compõem o *corpus* PEAPL2, bem como o perfil dos informantes, estão disponíveis no endereço: <http://www.uc.pt/fluc/rcpl2/>. Para obter mais informações sobre o método de recolha dos textos, veja-se Martins, 2011.

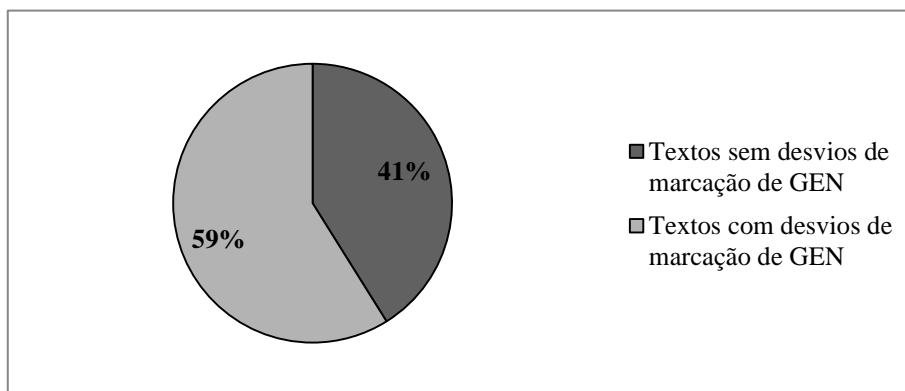


Gráfico 3.1 – Percentagem de textos do *corpus* com desvios de marcação de GEN

Cada produção escrita que compõe o *corpus* representa uma resposta do aprendente a um de nove estímulos previamente apresentados por escrito (cf. Anexo I). O informante deveria responder de acordo com estímulo selecionado, sem recurso a dicionário ou outro auxiliar.

Visto que, para este trabalho, se consideraram os desvios relativos à marcação do GEN dentro do GN, foi relevante determinar a quantidade de nomes por texto, de forma a considerar os desvios em relação ao conjunto total de itens nominais das produções que integram o *corpus* do estudo. Para a contagem consideraram-se todos os itens lexicais que, em cada texto escrito, desempenhavam as funções da classe do Nome. Tal perspectiva pressupôs que se contabilizasse como um único nome: i) os compostos como, por exemplo, *ponto de vista*, *fim-de-semana*; ii) nomes de grupos musicais (*Buraka Som Sistema*), de personalidades históricas (*Inês de Castro*). Além disso, também foram contabilizados os “estrangeirismos”, tanto os que fazem parte do léxico português europeu contemporâneo (é o caso de *karaoke*, *surf*), como os que representam importações dos próprios informantes (por exemplo, nomes de pratos típicos, costumes tradicionais dos países de origem, etc.).

O total de nomes apurados nos textos dos informantes por cada nível de proficiência em PL2 está representado no Quadro 3.1, em que também se estabelece a relação com o número de palavras e número de textos. A média de nomes do *corpus* é de, aproximadamente, 49 itens por texto. Como é possível observar, os dados apresentados no Quadro 3.1 estão organizados em função do nível de proficiência em PL2 dos informantes, uma vez que esta variável vai ser crucial para podermos apurar, durante a análise dos desvios de marcação de GEN, o desenvolvimento linguístico dos aprendentes de PL2.

Nível do QECRL	Número de textos	# Palavras	# Nomes
A1/A1 +	73	11.092	2.432
A2/ A2 +	63	11.064	2.443
B1	150	38.978	8.425
B2	24	6.642	1589
C1	13	3.429	879
TOTAL	323	71.196	15.768

Quadro 3.1 – Distribuição de textos, palavras e nomes por grupos de informantes em função do nível de proficiência em PL2

Na próxima secção (3.2.2) apresenta-se uma descrição detalhada do modo como estes níveis de proficiência foram apurados e como os informantes se organizam por esses mesmos níveis.

Antes de se proceder ao tratamento analítico dos dados, é relevante considerar as informações que dizem respeito ao perfil dos informantes do *corpus*.

3.2.2. PERFIL DOS INFORMANTES

As informações relativas ao perfil dos informantes foram recolhidas a partir dos dados fornecidos pelos próprios, no momento em que respondiam aos inquéritos³². Trata-se de um grupo muito variado, sob múltiplos aspetos, como a idade, ano de início da aquisição/aprendizagem de português, a LM e o nível de proficiência em PL2. Além disso, existem situações muito distintas no que diz respeito ao percurso de formação e aos objetivos dos aprendentes durante a aquisição/aprendizagem de PL2.

Muitos dos informantes são alunos *Erasmus* que frequentam os cursos num contexto de progressão dos estudos e permanecem em Portugal durante um ou dois semestres. Há também um subconjunto de aprendentes imigrantes que, por residirem em Portugal, necessitam de frequentar estes cursos para aprofundar o conhecimento da língua, sem possuírem, necessariamente, formação académica. No gráfico 3.2 pode-se verificar que grande parte dos informantes deste *corpus* (cerca de 69%) são, efetivamente, alunos universitários do programa *Erasmus*.

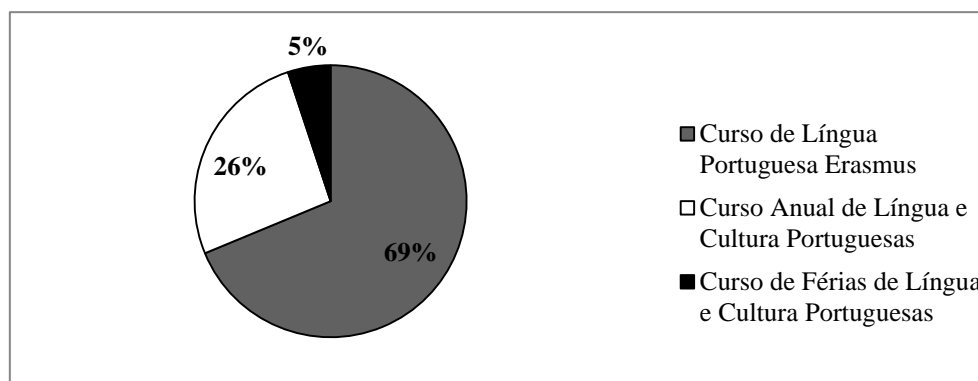


Gráfico 3.2 – Percentagem de informantes por curso de PL2

³² Os dados fornecidos pelos informantes do *corpus* PEAPL2 estão disponíveis para consulta, num documento único *Excel*, em <http://www.uc.pt/fluc/rcpl2/dados/>. Neste mesmo endereço pode-se aceder a informação detalhada sobre os informantes deste estudo.

Idade

A média de idade do grupo dos 256 informantes é de, aproximadamente, 24.40 anos³³. O gráfico 3.3 ilustra a distribuição dos aprendentes pelas diferentes faixas etárias.

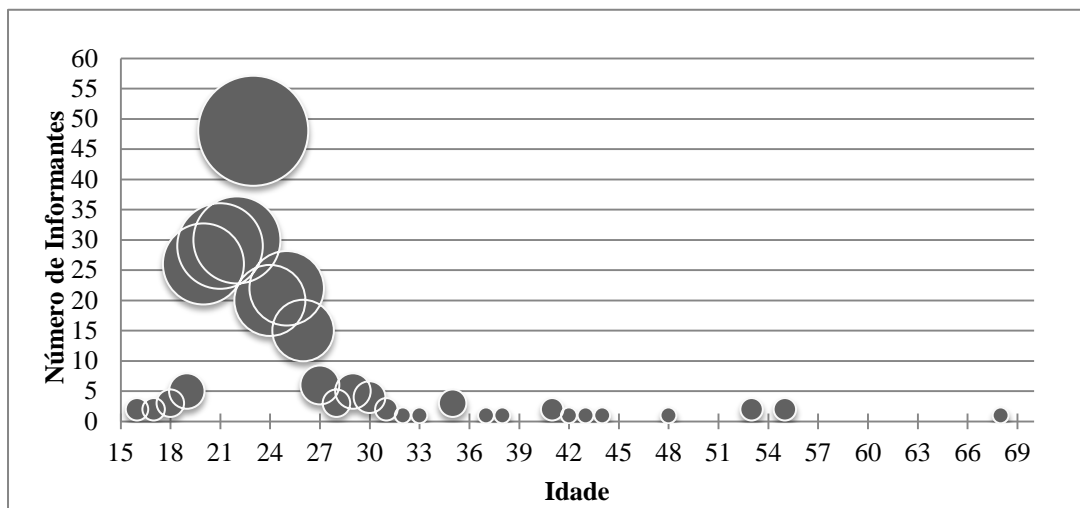


Gráfico 3.3 – Distribuição dos informantes por idade

Início do estudo de PL2

Outra informação concedida pelos aprendentes dizia respeito ao ano em que iniciaram o estudo de PL2 e, de acordo com estes dados, a maioria dos aprendentes iniciou a aquisição/aprendizagem de PL2 em 2009³⁴ (cf. Gráfico 3.4), ou seja, iniciaram o estudo de PL2 numa data muito próxima à recolha dos textos que compõem o *corpus*. Deste modo, e tendo em conta os dados assinalados no Gráfico anterior (cf. Gráfico 3.3), é possível afirmar que este *corpus* foi produzido, em grande parte, por aprendentes tardios de PL2.

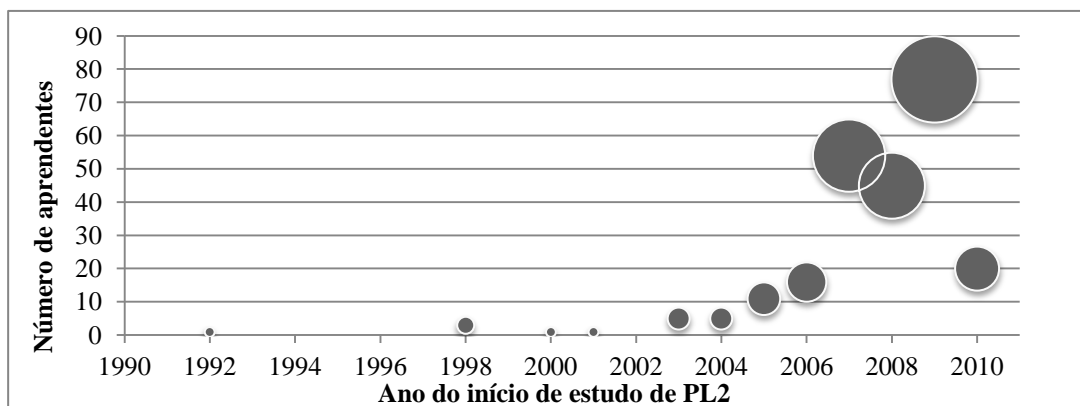


Gráfico 3.4 – Distribuição dos informantes pelo ano em que iniciaram o estudo de PL2

³³ Dos 256 informantes seleccionados, seis não forneceram informação quanto à idade.

³⁴ Dos 256 informantes seleccionados, dezassete não forneceram dados sobre o início de estudo de PL2.

Note-se que este é um dado relevante, uma vez que este trabalho visa, sobretudo, estudar a aquisição/aprendizagem da marcação de GEN nominal por aprendentes que iniciam este processo numa fase tardia do seu desenvolvimento.

Língua Materna

O conjunto das LM dos informantes é, igualmente, muito diversificado. Verifique-se, no Quadro 3.2 a disposição dos informantes por LM.

LM	Número de informantes
Alemão	38
Árabe	1
Basco	1
Búlgaro	2
Catalão	1
Checo	10
Chinês	15
Chinês (Cantonês)	6
Chinês (Mandarim)	1
Coreano	7
Croata	1
Dinamarquês	1
Eslovaco	1
Esloveno	1
Espanhol	16
Farsi	2
Finlandês	2
Francês	8
Galego	3
Grego	3
Hindi	1
Húngaro	2
Inglês	30
Italiano	33
Japonês	9
Letão	1
Lituano	3
Neerlandês	12
Polaco	10
Romeno	8
Russo	7
Sueco	2
Tagalo	1
Tailandês	1
Tétum	1
Turco	4
Ucraniano	4
Alemão/ Turco	1
Ucraniano/ Russo	1
Alemão/ Francês	1
Checo/ Eslovaco	1
Alemão/ Espanhol	1
Espanhol/ Catalão	1
Hindi/ Sindi	1
TOTAL	256

Quadro 3.2 – Distribuição dos informantes por LM

Na verdade, trata-se de um grupo muito heterogéneo no que concerne às LM. Uma vez que neste trabalho se pretende identificar quais os fatores que contribuem para a complexidade da aquisição/aprendizagem da marcação do GEN dentro do GN, independentemente das LM dos aprendentes, não se verificará a influência da LM durante este processo.

Nível de proficiência em PL2

Como se verificou, os informantes deste estudo distribuem-se por diversos cursos de PL2 da FLUC, cujos planos curriculares correspondem a diferentes níveis de proficiência. Assim, e tendo em conta a turma em que estavam inseridos, foi possível estabelecer uma correspondência entre o nível de cada uma e os níveis do *Quadro Europeu Comum de Referência para as Línguas* (QECRL) (cf. Anexo II). Veja-se a distribuição dos informantes por nível, no gráfico seguinte³⁵:

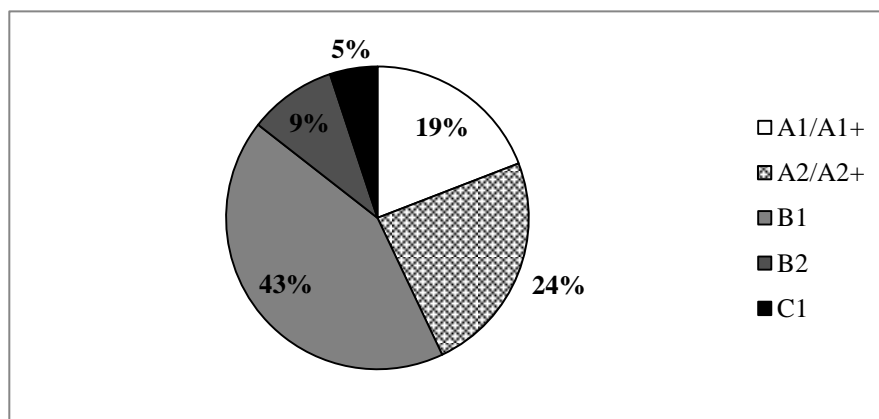


Gráfico 3.5 – Distribuição dos informantes por nível comum de referência do QECRL

Como se pode confirmar, os informantes que produziram os textos que compõem este *corpus* distribuem-se pelos diferentes níveis comuns de referência do QECRL: os níveis elementares (A1 e A2); os níveis intermédios (B1 e B2) e o nível proficiente (C1), sendo o nível B1 o mais representativo, com cerca de 43% dos informantes.

Dado que um dos objectivos deste estudo consiste, precisamente, na análise dos desvios em função dos níveis de proficiência dos aprendentes, importa, em primeiro lugar, referir como são descritos pelo QECRL. No que se refere à competência gramatical que abrange “o conhecimento dos recursos gramaticais da língua e a capacidade para os utilizar” (Alves, 2001:161), na qual se insere a categoria gramatical de GEN, o QECRL propõe os seguintes descritores para cada nível aqui representado (Alves, 2001:163):

³⁵ De modo a avaliar em qual dos níveis de proficiência em PL2 se encontravam e em que turma/nível se ajustariam melhor, os aprendentes que frequentavam os *Cursos Anual e de Férias* deveriam responder a um teste diagnóstico de português. No caso específico dos alunos *Erasmus*, estes eram integrados nas turmas de acordo com o nível de aprendizagem formal do português que frequentavam nos seus países de origem. Todavia convém salientar que não foi possível garantir que todos os alunos estivessem inseridos nas turmas mais adequadas ao seu nível de proficiência.

- A1: “Mostra apenas um controlo limitado de algumas estruturas e formas gramaticais simples (...);”
- A2: “Usa, com correcção, estruturas simples, mas ainda comete erros elementares de forma sistemática (...);”
- B1: “Usa, com uma correcção razoável, um repertório de ‘rotinas’ e de expressões frequentemente utilizadas e associadas a situações mais previsíveis.”;
- B2: “Bom controlo gramatical; podem ainda ocorrer ‘lapsos’ ocasionais ou erros não sistemáticos e pequenos erros na estrutura da frase, mas são raros (...);”
- C1: “Mantém um nível elevado de correcção gramatical de forma constante, os erros são raros e difíceis de identificar”.

Tendo em conta estes descritores, neste trabalho espera-se que, em função do nível de proficiência em PL2 dos informantes, se registem diferentes comportamentos desviantes que, além disso, sejam representativos de distintas fases do processo de construção da *Interlíngua* (IL) dos falantes aprendentes de uma L2 (cf. secção 3.4.2).

Traçado o perfil dos informantes deste estudo no que diz respeito à idade, ano de início de estudo de PL2, LM e nível de proficiência em PL2, veja-se em seguida o tratamento das produções selecionadas, tendo em conta os objetivos delineados para esta investigação.

3.3. TRATAMENTO DOS DADOS

3.3.1. TIPOLOGIA DE DESVIOS

Como se explicou para a seleção dos textos do *corpus* que serviram de base à investigação foram consideradas todas as estruturas que apresentavam desvios quanto à marcação de GEN em português dentro do GN, na medida em que serão reveladoras dos valores de GEN atribuídos aos nomes pelos aprendentes de PL2. Do conjunto total de nomes que compõem o *corpus* (cf. Quadro 3.1) regista-se uma percentagem de ocorrências desviantes na marcação de GEN em cerca de 4% dos itens nominais (cf. Gráfico 3.6). Assim, e apesar da larga percentagem de textos assinalados com desvios de marcação de GEN (cf. Gráfico 3.1), a percentagem de desvios produzidos em relação à totalidade dos nomes no *corpus* é reduzida, o que demonstra que esta estrutura gramatical é uma área que, em grande parte, os aprendentes de PL2 dominam.

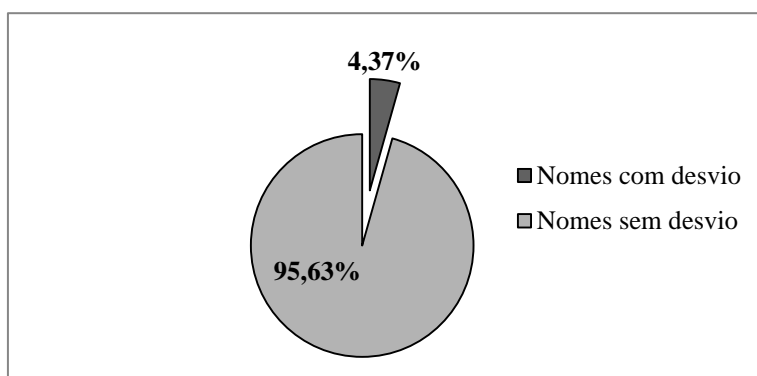


Gráfico 3.6 – Percentagem de desvios de atribuição de GEN nominal

Este valor percentual (4% aproximadamente) corresponde a um total de **689** ocorrências de desvios de marcação de GEN dentro do GN, e foi a partir deste conjunto que se procedeu à criação de uma tipologia dos casos desviantes, de modo a verificar quais os tipos de nomes mais afetados. Esta tipologia baseia-se nas características da estrutura do GEN do português, Língua-Alvo (LA), que foram consideradas no Capítulo 2 deste trabalho.

Desta forma, registaram-se os desvios de atribuição do valor de GEN nominal observáveis quer através da forma morfológica do nome, quer através das demais palavras que com este deverão concordar em GEN, coocorrendo no GN: determinantes, adjetivos e quantificadores (cf. Capítulo 2, p.20).

Para a organização dos dados criaram-se códigos que identificam os vários tipos de desvio, bem como a circunstância de se terem registado comportamentos de hesitação³⁶ do informante.

Veja-se, em primeiro lugar, os casos de hesitação (cf. Quadro 3.3) que correspondem a situações em que o aluno se mostra hesitante quanto à marcação de GEN nominal, riscando um determinado segmento textual e/ou procedendo a alterações.

	Exemplo	Observações
H0 [0]	«Para além disso, vivendo <numa> <u>num cidade</u> , acabamos por <ficar mais> ter mais mobilidade, uma vez que existe uma rede /desenvolvida/ de transportes públicos, graças a qual podemos resolver os nossos problemas de dia-a-dia com /mais/ facilidade.» [UC.CA.S.A.05.09.08.69.3Q] ³⁷	Hesitações que culminaram numa marcação de GEN <u>desviante</u> em relação à LA.
H1 [1]	«Na secunda-feira, <como> na tarde, como acabo de estydar, eu vou sempre no bar da praça a tomar <uma> <u>um coupo</u> de cerveja com os meus amigos, e ai falamos até tarde.» [UC.ER.LPI.A.12.09.17.33.1J]	Hesitações que culminaram numa marcação de GEN <u>correspondente</u> à LA.

Quadro 3.3 – Casos de hesitação

Por conseguinte, os casos de hesitação que originaram uma marcação de GEN desviante (H0) foram contabilizados, enquanto as restantes hesitações, que culminaram numa marcação de GEN correspondente à estrutura de GEN nominal da LA (H1) não foram contabilizadas nesta investigação³⁸.

No que respeita à tipologia de desvios, são cinco os códigos que representam os desvios de marcação de GEN nominal. Assim, o código:

[1] identifica os desvios relativos a nomes cujo referente é [+sexuado] (*o irmão, a irmã*), em que o valor de GEN coincide com o sexo do referente.

³⁶ Foi possível apurar os casos de hesitação, uma vez que as respostas dos alunos foram transcritas segundo determinadas convenções previamente estabelecidas em Leiria, 2006. Tais convenções identificavam as seguintes situações: i) segmentos riscados - <xxx>; ii) segmentos riscados ilegíveis - <(...)>; iii) segmentos acrescentados - /xxx/; iv) leituras conjecturadas - /*xxx/.

³⁷ Cada resposta está identificada com um código que possui um conjunto de informações várias: turma, nível, data da recolha, número do informante, código do estímulo. Para uma descrição detalhada destes códigos, veja-se Martins, 2011.

³⁸ Todos os dados podem ser consultados no Volume II deste trabalho, Anexo III.

[2] identifica os desvios relativos a nomes com referentes extralinguísticos [-sexuados] cujo GEN seria corretamente dedutível a partir do IT: se o IT é *-o*, então o nome é masculino (*o banco, o mundo*); se o IT é *-a*, então o nome é feminino (*a casa, a lata*).

[3] identifica os desvios relativos a nomes com referentes extralinguísticos [-sexuados] cujo GEN não é dedutível a partir do IT: se o IT é *-o*, então o nome é feminino (*a tribo, a rádio* [estação]); se o IT é *-a*, então o nome é masculino (*o mapa, o dia*).

[4] identifica os desvios relativos a nomes com referentes sexuados³⁹ cujo valor de GEN não coincide, necessariamente, com o sexo do referente (*a pessoa, a criança*).

[5] identifica os desvios relativos a nomes com referentes não sexuados cujo GEN somente é inferível a partir dos marcadores sintáticos (determinantes, adjetivos, etc.). É o caso dos nomes de tema em *-e* e dos atemáticos (*o pente, a ponte; a mão, o limão*).

Em cada categoria existem quatro subcategorias que distinguem os diferentes constituintes do grupo nominal nos quais se evidencia(m) a(s) marca(s) do desvio: [a] o desvio é marcado na forma morfológica do nome; [b] o desvio ocorre no determinante; [c] o desvio ocorre no adjetivo; e [d] o desvio ocorre no quantificador.

O Quadro 3.4 apresenta uma descrição detalhada das categorias com exemplos recolhidos no *corpus*.

³⁹ Estes nomes correspondem aos sobrecomuns e aos epícenos. Na verdade, na maior parte destes casos, e sempre que o IT é *-o* ou *-a*, é possível, a partir desse constituinte, inferir o valor de GEN. Por exemplo, os que têm IT *-a* são femininos (*a criança, a pessoa; a águia, o macaco*). Tal facto pode funcionar como uma regra que permita ao aluno reconhecer o valor de GEN do item nominal.

[1]		
	Exemplo	Observações
1a	«Os <u>alemães</u> chegam (frequentemente) em ponto, eles gostam de discutir nos bares durante eles bebem <(...)> cerveja <boa>.» [UC.ER.LPIII.A.12.09.74.50.2L]	Desvios marcados na forma morfológica do nome.
1b	«Gozar muito tempo com <u>meu filha</u> e saver de suas niquietudes.» [UC.CF.EB.02.07.09.04.33.1J]	Desvios que ocorrem quando determinante masculino antecede nome feminino.
	«Nunca imaginava <(...)> <u>uma capitão</u> como ti. <se eu>.» [UC.CA.S.A.05.09.06.6.1B]	Desvios que ocorrem quando determinante feminino antecede nome masculino.
1c	«Tenho um <u>irmã</u> mais <u>novo</u> .» [UC.CA.E.A.12.09.04.1.1A]	Desvios que ocorrem quando um adjetivo masculino co-ocorre com um nome feminino.
	«Porque disso eu estou <(...)> um <u>rapas nova</u> quando a proseima ano commencas.» [UC.CF.EB.02.07.09.07.33.1J]	Desvios que ocorrem quando um adjetivo feminino co-ocorre com um nome masculino.
1d	«Tenho <u>um irmã</u> mais novo.» [UC.CA.E.A.12.09.04.1.1A]	Desvios marcados no quantificador.
[2]		
	Exemplo	Observações
2a	«Depois <a> o <u>junto da freguesia</u> mandou erigir um relógio em frente da casa.» [UC.CF.EB.02.07.09.03.77.3T]	Desvios marcados na forma morfológica do nome. Quando o aluno modifica o IT, i.e., “formata” o nome, colocando-lhe no final a vogal -o.
	«Na 1º piso é uma cozinha, uma <u>quarta</u> <(...)> pai e mãe, uma <u>quarta</u> <(...)> minha irma, grande <u>quarta</u> com televisão <é> e casa de banho.» [UC.ER.LPIA.05.10.09.1.1A]	Desvios marcados na forma morfológica do nome. Quando o aluno modifica o IT, i.e., “formata” o nome, colocando-lhe no final a vogal -a.
	«O <u>equipe</u> que eu gosto mais de chamar-se “Kent C.C.”» [UC.CF.EB.01.07.09.03.33.1J]	Desvios marcados na forma morfológica do nome. Quando o aluno modifica o IT do nome, i.e., “formata” o nome, com recurso ao IT -e.
2b	«Estava <u>um experiência</u> ótimo!» [UC.ER.LPIII.A.12.09.63.6.1B]	Desvios marcados no determinante. Quando determinante masculino antecede nome feminino.
	«Tu ainda lembras <u>danossatempo</u> junto <aqui> em Portugal?» [UC.ER.LPIII.A.12.09.75.6.1B]	Desvios marcados no determinante. Quando determinante feminino antecede nome masculino.
2c	«Por causa disso as oportunidades na minha cidade natal por vezes parecem-me sem fim, lá trabalho numa <u>loja /especializado/</u> <dem> em chocolate <específico> e adoro saber e aprender sempre mais.» [UC.ER.LPIII.A.12.09.12.33.1J]	Desvios marcados no adjetivo. Quando um adjetivo masculino co-ocorre com nome feminino.
	«Mas agora vai mudar tudo, porque o Município<(...)> tem feito um projeto de renovação do todo o bairro: então <vai> aparecerá um parque, novos prédios, e as	Desvios marcados no adjetivo. Quando um adjetivo feminino co-ocorre com nome masculino.

	indústrias <, > será mudadas de lugar, porque este bairro <se> vai ser o primeiro <(…)> “bairro ecologico”, ou seja, construídos só com materiais <ecologicos e com medi> e <u>medidas ecologicos</u> ⁴⁰ .» [UC.ER.LPIII.A.12.09.40.77.3T]	
2d	«Eu gosto muito de sair ás discótecas com os meus amigos e tenho saudades de <minha> /sair a minha/ universidade em Manchester porque há <u>muitos discótecas</u> <(…)> ali e não há muitas /aqui/ em Coimbra.» [UC.ER.LPIII.A.12.09.33.33.1J]	Desvios marcados no quantificador. Quando um quantificador masculino está associado a um nome feminino.
	«Também gosto muito de ler livros, mas é verdade, que depois as dias cheias da leitura <obrigatória> /portuguesa/ <portuguesa> /obrigatória/ (nas aulas da literatura brasileira ou portuguesa) não tenho vontade para ler <u>outras livros</u> mais divertidas paras mim (claro em português), porque na Republica Checa não é possível comprar estes.» [UC.ER.LPIII.A.12.09.19.33.1J]	Desvios marcados no quantificador. Quando um quantificador feminino está associado a um nome masculino.
[3]		
	Exemplo	Observações
3a	Não se registaram desvios desta subcategoria.	Desvios marcados na forma morfológica do nome. Quando o aluno modifica o IT do nome, i.e., “formata” o nome colocando-lhe no final a vogal -o.
3b	«Também foi muito engraçado este dia quando fizemos /este caminho/ <estivemos perdidos> nas montanhas e fomos perdidos porque não usamos uma mapa.» [UC.ER.LPIII.A.12.09.68.6.1B]	Desvios marcados no determinante. Quando determinante feminino antecede nome masculino.
3c	«Quando <prec>preciso de um <u>dia calma</u> , fico em casa e leio um livro ou coso roupa.» [UC. ER.LPIII.F.06.09.28.33.1J]	Desvios marcados no adjetivo. Quando um adjetivo feminino co-ocorre com nome masculino.
3d	«Eu vou fazer <u>um foto</u> e emáil-o para ti!» [UC.ER.LPI.A.05.10.23.6.1B]	Desvios marcados no quantificador. Quando um quantificador masculino está associado a um nome feminino.
	«O meu deus, a outra noite, foi <(…)> tão bêbeda que, tive uma <resaca> /ressaca/ por duas dias!» [UC.ER.LPIII.A.12.09.04.6.1B]	Desvios marcados no quantificador. Quando um quantificador feminino está associado a um nome masculino.
[4]		
	Exemplo	Observações
4a	«A cidade Coimbra tem tradição e ai mesmo tempo não é cheia de <u>turistos</u> e <tamb> a vida aqui também não é muito cara.» [UC.ER.LPIV.A.06.09.39.69.3Q]	Desvios marcados na forma morfológica do nome. Quando o aluno “formata” o item nominal, colocando-

⁴⁰ Registaram-se casos em que numa mesma frase ocorria mais do que um desvio e, em tais situações, os desvios foram igualmente contabilizados e identificados.

		lhe no final a vogal -o ou -a.
4b	«A vida <u>das estudantes</u> <e> são mais estric<ta /do/ que /a vida/ em Portugal; /porque/ tem que ir <as> /a/ todas /as/ suas aulas e tem que fazer todo o seu trabalho.» [UC.ER.LPIII.A.12.09.33.50.2L]	Desvios marcados no determinante.
4c	« <u>Português</u> pessoas são muito simpático.» [UC.CA.E.A.05.09.01.33.1J]	Desvios marcados no adjetivo.
4d	« Eutambem encontro muitos pessoas de país diferentes<(> falo sempre linguas<f> diferentes<,> e gosto muito disso facto porque eu preciso de utilizar a minha cabeça e realizo uma ameliaração todas as semanas.» [UC.ER.LPIII.A.12.09.15.6.1B]	Desvios marcados no quantificador.
[5]		
	Exemplo	Observações
5a	«Agora as minhas <u>examas</u> começou por isso eu tenho de trabalhar.» [UC.ER.LPI.A.05.10.10.6.1B]	Desvios marcados na forma morfológica do nome. Quando o aluno “formata” o item nominal, colocando-lhe no final a vogal -a.
5b	« <u>O viajem</u> preferida esteve à Noruegia.» [UC.ER.LPI.A.05.10.14.75.3S]	Desvios marcados no determinante. Quando determinante masculino antecede nome feminino.
	«Agora eu estudo Gestão na Universidade de Coimbra e tenho de estudar para <u>as exames</u> .» [UC.ER.LPI.A.05.10.15.6.1B]	Desvios marcados no determinante. Quando determinante feminino antecede nome masculino.
5c	«Também apanhe um o autocarro de <Madr> Coimbra à Madrid no mes de Outubro e foi um <u>viagem</u> muito <u>pessado</u> e de muitas horas, ao final amanhã apanho um comboio de Lisboa até Madrid e ainda não se como é que vai ser o viagem!!» [UC.ER.LPII.A.12.09.15.75.3S]	Desvios marcados no adjetivo. Quando um adjetivo masculino co-ocorre com nome feminino.
	«Voltar duma viagem, não é só ter feito experiências e mesmo creixer e ter aberto a cabeça pra <u>novas horizontes</u> .» [UC.ER.LPIII.A.12.09.29.52.2L]	Desvios marcados no adjetivo. Quando um adjetivo feminino co-ocorre com nome masculino.
5d	«Eu gosto muito de <Portugal> <u>todos</u> <u>ciudades</u> em Portugal.» [UC.ER.LPI.A.05.10.06.1.1A]	Desvios marcados no quantificador. Quando um quantificador masculino está associado a um nome feminino.
	«Nos meus tempos livres eu gosto de fazer muitas coisas,, mas, em verdade, não <tehno> tenho muitos tempos livres, porque <ten> tenho <u>muitas exames</u> neste periodo.» [UC.ER.LPI.A.12.09.13.33.1J]	Desvios marcados no quantificador. Quando um quantificador feminino está associado a um nome masculino.

Quadro 3.4 – Tipologia de desvios

Deste modo, apurada a tipologia, procedeu-se à quantificação dos casos desviantes registados em cada categoria. É de salientar a existência de casos em que um único nome origina mais do que um desvio e, nessas situações, cada subcategoria de desvio foi devidamente contabilizada e assinalada com o código que lhe correspondia (cf. Anexo III).

Note-se, ainda, que para o presente estudo não foram contabilizadas as situações de hipercorreção formal, i.e., os casos em que o aprendente procede à correta marcação de GEN do item nominal a partir dos marcadores sintáticos (determinantes, adjetivos e quantificadores) e, além disso, formata o nome de acordo com a regra “se o nome termina em *-o* é masculino; se termina em *-a* é feminino”. Dito isto, os exemplos assinalados no Quadro 3.5, que correspondem a todas as hipercorreções recolhidas no *corpus* (16 casos no total), não foram contabilizados, visto se tratarem de desvios de marcação de GEN que possuem um carácter especial em relação os restantes desvios quantificados nesta investigação.

No entanto, embora estes casos não tenham sido contabilizados de acordo com a tipologia de desvios criada para este trabalho, apresenta-se, em seguida, um conjunto de reflexões a partir destas hipercorreções que, como se depreenderá, são quantitativamente pouco representativas em relação ao total de nomes produzidos no *corpus* (Cf. Quadro 3.8).

Nível de Proficiência do QECRL	Casos de Hipercorreção
A1/A1+	« <u>Az vezes</u> sento na sofá em frente /e vejo/ na televisão.» [UC.CA.E.A.05.09.04.6.1B]
	«Eu <fez os> estudou muitíssimo por os meus exames e <u>testos</u> .» [UC.ER.LPI.A.05.10.02.6.1B]
	«Usámos tantas <u>todas noites</u> e no tivemos programma» [UC.ER.LPI.A.05.10.03.75.3S]
	«Hoje eu tenho <(...)>/ <u>um/ testo</u> e durante <próxima> próximo mês eu tenho de escrever um mais <trabalha> trabalho sobre demografía<da> de Portugal.» [UC.ER.LPI.A.05.10.05.6.1B]
	«Eu gosto de ler os livros no meu tempo livro, gosto de viajar, conhecer novas cidades e países, sair fora com as minhas amigas, ver <u>os filmes</u> interessantes, /ouvir a música,/ e muitas vezes fazer nada.» [UC.ER.LPI.A.05.10.15.1.1A]
	«As pessoas em Coimbra gostam de música muito e / <u>os billeta/</u> dos concertos são baratos / porque o C.M. de Coimbra os suporte/ aos fim-de-semana eu e o Carlos (e, às vezes, /o/ seu filho Brenda) visitaram os castelos ou a praia da Figueira da Foz ou descansaram em Coimbra.» [UC.ER.LPI.A.05.10.23.6.1B]
A2/A2+	«Agora, gosto de viajar de avião, mas antes, há 3 anos, tinha medo

	desste tipo de transporte porque um dia, quando eu e a minha família fomos às férias na Grécia, foi o voo horrível e por causa disso não gostava viajar de avião.» [UC.CA.E.B.11.09.11.75.3S]
	«Eu sei que sou uma parte do sistema , mais queria fazer uma pequena diferença.» [UC.ER.LPII.A.12.09.10.1.1A]
	«Aquila foi a primeira vez que viajei sozinha. Gusto muito viajar, em modo particular com os transportos <pl>publicos.» [UC.ER.LPII.B.01.10.05.75.3S]
B1	«Normalmente tenho de estudar muito durante os semetros .» [UC.ER.LPIII.F.06.09.28.33.1J]
	«<roAchos> pelo contrário, penando no ponto de vista dos residentes, <o que> oferece<(...)> o serviço agradável e promover a amizade entre <os> residentes são os sucessos do senhorio.» [UC.CA.I.B.05.09.12.77.3T]
	«<O> A maioria do meu tempo livre está <co> ocupada com várias festas, porque conheço aqui as pessoas das nacionalidades diferentes, então há o costume bom organizar as festas típicas<(...)> dev< de cada país, com comida <típica> e música típica. [UC.ER.LPIII.A.12.09.19.33.1J]
	«Eu gosto de fazer sobremesas, e sobretudo de fazer uma boa tarta de maçãs.» [UC.ER.LPIII.A.12.09.76.33.1J]
	«As vezes não acho que posa /morar/ num lugar diferente se bem que muitas vezes esteja no norte de Inglaterra, onde estudo.» [UC.ER.LPIII.A.12.09.103.77.3T]
B2	«O país onde nasci chama-se Bulgária e se encontra no lesta da Europa.» [UC.ER.LPIV.A.06.09.07.50.2L]
C1	«Trabalhamos, <estudam> estudamos, ganhamos dinheiro, e depois, mais uma veza , trabalhamos, estudamos e ganhamos!» [UC.CA.S.A.12.09.17.69.3Q]

Quadro 3.5 – Casos de hipercorreção

Em todos os níveis de proficiência se registam casos de hipercorreção, sendo que os níveis A1/A1+ e A2/A2 + apresentam maior número de casos desta natureza. Esta evidência demonstra que, numa primeira fase de aquisição/aprendizagem da marcação do GEN nominal os aprendentes, por reconhecerem corretamente o valor de GEN destes nomes, “formatam-nos” de acordo com a regra de atribuição de GEN que assimilaram (veja-se a p. 87 deste trabalho). Como se verifica a partir destes dados (Quadro 3.5), o que os aprendentes não dominam é a forma morfológica “irregular” do nome cujo valor de GEN conhecem.

Além disso, estes casos podem ser representativos do processo de construção das IL dos falantes aprendentes de PL2 que, até passarem a um novo estágio de desenvolvimento, aplicam a regra de marcação de GEN a todos os itens nominais (cf. Capítulo 1, secção 1.3).

Analisando atentamente os exemplos, verifica-se que os itens mais afetados pela hipercorreção dizem respeito a nomes [-sexuados] cujo valor de GEN somente é dedutível a partir dos marcadores sintáticos (determinantes, adjetivos e

quantificadores). Provavelmente, por não reconhecerem neste tipo de itens nenhum critério que os auxilie na correta marcação de GEN, os aprendentes, tendo interiorizado que a marca final do nome condiciona o seu valor de GEN, modificam o item nominal em consonância com esta regra de atribuição de GEN. Assim, quando se trata de um nome masculino, adicionam-lhe o IT tipicamente associado ao masculino, ou seja, *-o*: **testo*, *testos* (em vez de *teste*, *testes*); **filmos* (em vez de *filmes*); **transporto*, *transportos* (em vez de *transporte*, *transportes*); **semetros* (em vez de *semestres*); **costumo* (em vez de *costume*) (cf. Quadro 3.5)⁴¹.

Quando recorrem a um nome com referente extralinguístico [-sexuado] do GEN feminino, os alunos adicionam-lhe o IT tipicamente associado ao feminino, i.e., *-a*: **veza* (em vez de *vez*); **noitas* (em vez de *noites*); **amizada* (em vez de *amizade*); **tarta* (em vez de *tarte*) (cf. Quadro 3.5).

A sobregeneralização da regra segundo a qual a marca final determinaria o valor de GEN dos nomes é mais evidente nos níveis elementares. Por exemplo, perante o nome *sistema* que, apesar de ter IT *-a* é do GEN masculino, o aluno do nível A2/A2+ formata-o de acordo com a regra: “se o nome termina em *-o* é do GEN masculino; se termina em *-a*, possui valor de GEN feminino” (cf. Quadro 3.5).

Apesar de ocorrerem casos de hipercorreção em todos os níveis, estes são mais notórios nos níveis A1/A1+ e A2/A2+ e tal facto mostra que a interiorização desta “fórmula” faz parte de uma primeira fase do processo de aquisição/aprendizagem da marcação de GEN.

Note-se, ainda, que estes casos de hipercorreção, embora pouco frequentes em relação ao total de nomes do *corpus* (cf. Quadro 3.8), são muito interessantes, pois constituem, como se constatou, um dos indícios do modo como funciona o próprio processo de construção da IL destes aprendentes.

Definida a tipologia de desvios de marcação de GEN, bem como as suas características, na próxima secção deste capítulo apresenta-se os resultados da análise dos dados apurados neste estudo, tendo em conta as categorias e subcategorias de desvios

⁴¹ Refira-se ainda a presença do único caso isolado que consiste na “formatação” do nome masculino *bilhete* em que, ao contrário do padrão seguido, o aprendente optou por colocar o IT *-a*: **billeta* (cf. Quadro 3.5)

criadas para esta investigação e dando particular relevo à distribuição dos desvios por nível de proficiência dos aprendentes, visando, assim, extrair conclusões sobre o padrão de desenvolvimento das respectivas interlínguas.

3.4. QUANTIFICAÇÃO E ANÁLISE DOS DADOS APURADOS

3.4.1. FREQUÊNCIA DE DESVIOS DO *CORPUS*

O Capítulo 2 deste trabalho apresenta um conjunto de critérios (de ordem semântica e formal) que poderá permitir ao aprendente de PL2 identificar, em certos itens nominais, o valor de GEN que lhes está associado. Importa, por isso, analisar os dados de modo a verificar se os desvios mais frequentes dizem respeito aos itens nominais cuja marcação de GEN está associada ao carácter [+ sexuado] do nome ou se, pelo contrário, os casos desviantes são quantitativamente mais representativos em relação aos nomes [- sexuados] em que a atribuição do valor de GEN está em número avultado de casos associada a fatores de natureza formal como, por exemplo, o seu IT.

Ainda no referido Capítulo, procedeu-se ao estudo das ocorrências de todos os lemas recenseados na CORLEX (cf. Capítulo 2, secção 2.3.2), e apurou-se que em quase 60% da totalidade de ocorrências dos itens nominais dessa base de dados se verifica a correspondência entre o IT e o seu GEN, i.e., em grande parte das ocorrências dos lemas nominais em português é possível os aprendentes de PL2 associarem o valor de GEN nominal a partir de regras de carácter formal (cf. Capítulo 2, secção 2.3.2). Sendo assim, os restantes casos em que a identificação do GEN não é visível nem por critérios de ordem semântica, nem por critérios de natureza formal (como o IT) deverão apresentar maiores problemas durante o processo de aquisição/aprendizagem da marcação de GEN e, conseqüentemente, maior número de desvios.

Tendo em conta estas hipóteses, e de acordo com a tipologia criada para este estudo, veja-se no Quadro 3.6 a distribuição das **689** ocorrências de desvios por categoria e no Gráfico 3.7 a representação das percentagens de desvio. Todos os dados que compõem a tipologia de desvios estão disponíveis no Volume II da presente dissertação, no Anexo III.

Categoria de desvio	Frequência de desvios
1 Desvios relativos a <u>nomes [+sexuados]</u> em que <u>o valor de GEN coincide com o sexo do referente.</u>	22
2 Desvios relativos a <u>nomes [-sexuados]</u> em que <u>o valor de GEN coincide com o IT</u> : GEN Masc. = IT -o; GEN Fem. = IT -a.	242
3 Desvios relativos a <u>nomes [-sexuados]</u> em que <u>o valor de GEN não coincide com o IT</u> : GEN Masc. = IT -a; GEN Fem. = IT -o.	60
4 Desvios relativos a <u>nomes [+sexuados]</u> em que <u>o valor de GEN não coincide, necessariamente, com o sexo do referente.</u>	28
5 Desvios relativos a <u>nomes [-sexuados]</u> em que <u>o valor de GEN somente é inferível a partir dos marcadores sintáticos (determinantes, adjetivos, etc.).</u>	337
TOTAL	689

Quadro 3.6 – Distribuição do número de desvios por categoria de desvio de marcação de GEN nominal

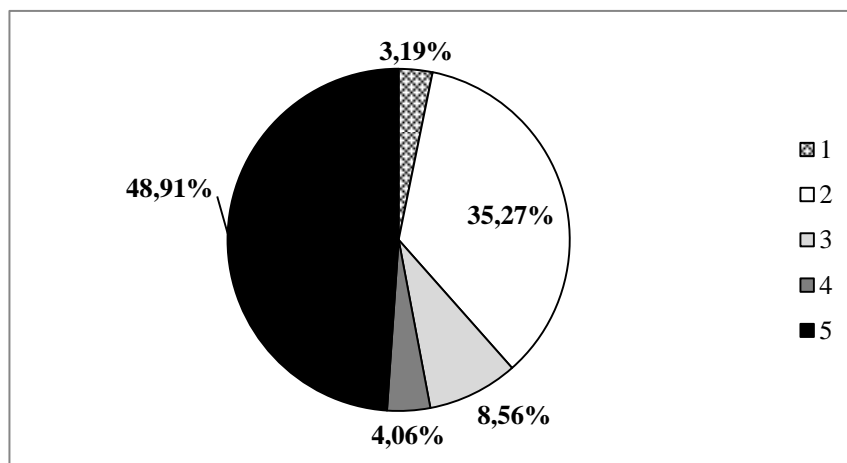


Gráfico 3.7 – Distribuição em porcentagem dos desvios produzidos no corpus por categoria

Como se pode verificar, a categoria de desvio mais frequente no corpus é a categoria 5 que corresponde a desvios relativos a itens nominais [-sexuados] cujo valor

de GEN somente é inferível a partir de marcadores sintáticos (determinantes, adjetivos, quantificadores). Uma vez que o GEN destes itens não é visível através de um conjunto de indícios (semânticos e formais) que permita aos aprendentes de PL2 estabelecer corretamente o seu valor de GEN, a percentagem de desvios é considerável (49% aproximadamente) e traduz-se em **337** ocorrências desviantes. Este resultado está, assim, em consonância com a previsão que postulámos anteriormente.

Também a categoria 2, que identifica os desvios relativos a nomes com referentes extralinguísticos [-sexuados] cujo GEN seria corretamente dedutível a partir do IT, apresenta uma percentagem de casos considerável, representando aproximadamente 35% do total de desvios produzidos, i.e., **242** ocorrências de marcação de GEN nominal desviantes. Tal resultado é bastante surpreendente e, para justificar estes valores, várias hipóteses podem ser equacionadas.

Em primeiro lugar, estes dados parecem indiciar que muitos aprendentes não incorporaram, de modo satisfatório, algumas pistas formais de marcação de GEN em português, não sendo cabalmente capazes de estabelecer, à medida que produzem estes nomes, a correspondência entre os valores de GEN dos nomes e respetivos IT. No entanto, tal poderá não ser o caso. Na verdade, o aluno pode saber as regras de atribuição do valor de GEN aplicáveis em português, mas como não reconhece o GEN de determinado item nominal, é provável que lhe atribua um IT errado ou marque mal o constituinte dentro do grupo nominal (determinante, adjetivo ou quantificador). Sintetizando, o aluno pode saber a regra de marcação de GEN, mas não conhece o valor de GEN que está associado a estes nomes. Sendo assim, estes dados sugerem que o GEN não é uma das primeiras propriedades do nome a serem assimiladas pelos aprendentes de PL2.

Outra hipótese explicativa para o elevado índice de desvios deste tipo pode estar correlacionada com o facto de os aprendentes apresentarem uma certa desconfiança face aos indícios morfológicos, como o IT, para atribuir corretamente os valores de GEN aos nomes. De facto, e como se viu a partir do estudo da base de dados da CORLEX (Capítulo 2, secção 2.3.2), embora este tipo de nomes seja muito frequente no *input* dos aprendentes, não é, todavia, suficiente para os alunos confiarem plenamente nestas pistas de atribuição de GEN.

Tendo todas estas hipóteses em conta, o avultado número de desvios nesta categoria representa um resultado muito interessante, indiciando que o GEN se aprenderá como uma propriedade intrínseca do nome, assimilado item a item.

No que diz respeito às subcategorias de desvios mais afetadas, veja-se o Quadro 3.7, que representa a distribuição da totalidade de desvios por subcategoria e o Gráfico 3.8 com os valores percentuais.

Subcategoria de desvio	Frequência de desvios
a (no nome)	49
b (no determinante)	381
c (no adjetivo)	96
d (no quantificador)	163
TOTAL	689

Quadro 3.7 – Distribuição do número de desvios por subcategoria de desvio de marcação de GEN nominal

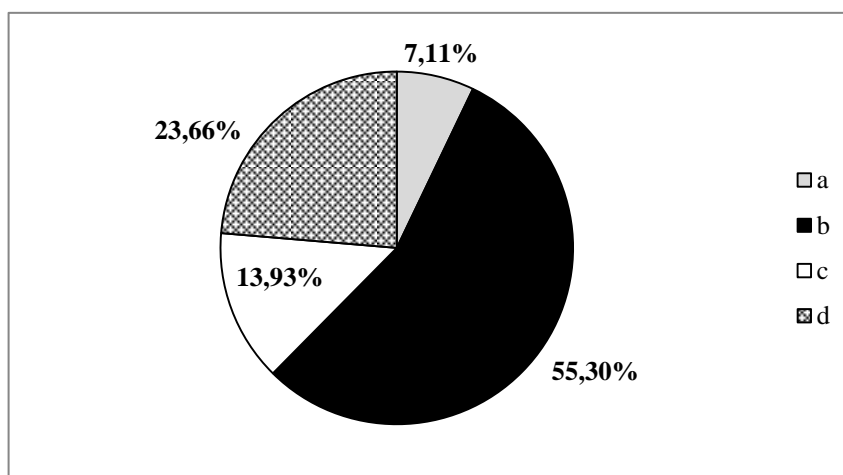


Gráfico 3.8 – Distribuição em percentagem das subcategorias de desvios produzidos no *corpus*

Pela leitura deste gráfico, verifica-se que as subcategorias de desvio mais afetadas são as subcategorias (b) e (d), que correspondem, respectivamente, aos desvios de marcação de GEN ocorridos no determinante – com cerca de 55% dos desvios, a que correspondem **381** casos – e no quantificador – com, aproximadamente, 14% das frequências que se traduzem em **163** comportamentos desviantes de atribuição de GEN nominal.

Trata-se de classes gramaticais que, possivelmente terão, no entender do aluno, pouco valor do ponto de vista comunicativo, pois acaba por constatar que, mesmo

“ignorando” estas marcas, consegue atingir os seus objectivos comunicativos. Com efeito, e como se assinalou no Capítulo 1 da dissertação (cf. secção 1.3), na construção da interlíngua os aprendentes tendem a simplificar o material linguístico da língua-alvo (LA) sobregeneralizando as várias regras que regem o sistema dessa LA (Selinker, 1992).

Definidos os resultados gerais, com a distribuição dos desvios por categoria e subcategoria, na próxima secção (3.4.2) apresenta-se a disposição dos desvios produzidos por nível de proficiência do QECRL dos aprendentes, com vista à identificação de padrões variáveis na aquisição/aprendizagem da marcação do GEN gramatical dentro do GN.

3.4.2 FREQUÊNCIA DE DESVIOS EM FUNÇÃO DO NÍVEL DE PROFICIÊNCIA DOS APRENDENTES

Como ficou referido na secção 3.2.2, no *corpus* desta investigação estão inseridas produções de aprendentes representativos de cinco níveis de aprendizagem formal correspondentes aos estabelecidos pelo QECRL: os níveis elementares (A1 e A2); os níveis intermédios (B1 e B2) e o nível proficiente (C1) (cf. Gráfico 3.5).

Num segundo momento de quantificação dos desvios pretendeu-se verificar a frequência dos casos desviantes em função dos níveis comuns de referência do QECRL representados pelos aprendentes que produziram os dados do *corpus*. Assim, a partir da totalidade dos nomes produzidos pelo conjunto dos aprendentes de um dado nível (cf. Quadro 3.8), procedeu-se ao cálculo da respetiva percentagem de desvios (cf. Gráfico 3.9) de modo a tornar possível a confrontação dos dados em função desta variável.

Nível do QECRL	# Nomes por nível	Número de desvios
A1/A1+	2.432	159
A2/A2+	2.443	166
B1	8.425	293
B2	1.589	45
C1	879	26
TOTAL	15.768	689

Quadro 3.8 – Frequência de desvios produzidos por nível

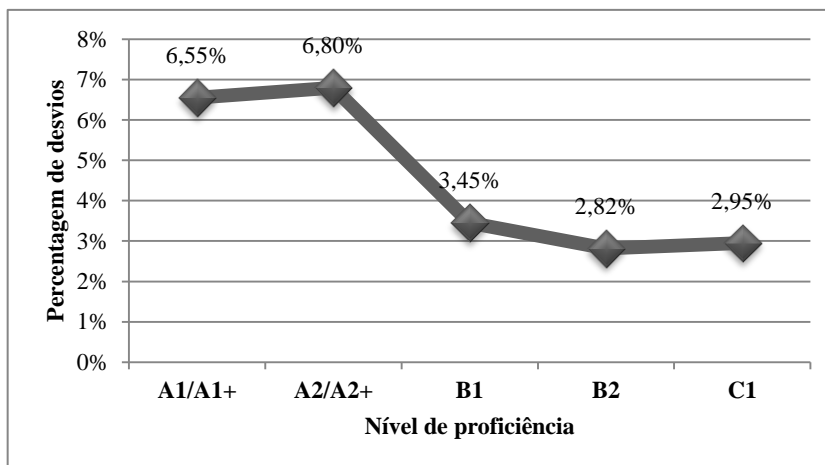


Gráfico 3.9 – Percentagem de desvios em função dos nomes produzidos por nível de proficiência em PL2

Como se pode observar a partir do Gráfico 3.9, os níveis elementares (A1 e A2) apresentam maior percentagem de desvios. Na passagem para os níveis intermédios (B1 e B2) e superior (C1), a percentagem de comportamentos desviantes diminui de forma considerável, o que demonstra que nestes níveis os aprendentes conseguem, em larga medida, contornar as dificuldades que a marcação desta categoria gramatical lhes apresenta. Além disso, observa-se um certo efeito *plateau* a partir do nível B1, o que revela um padrão de estabilização dos comportamentos desviantes, pelo menos em termos quantitativos.

Face a estes resultados, apresenta-se em seguida a análise detalhada dos desvios produzidos por nível de proficiência do QECRL representado neste *corpus*, em função das cinco categorias e subcategorias de desvios da tipologia criada para esta investigação (cf. secção 3.3.1). Além disso, verificar-se-á quais os nomes mais afetados por categoria e em que níveis são mais frequentes.

Veja-se os Gráficos 3.10 a 3.14 com a percentagem de desvios por categoria calculada em função do número total de desvios registados em cada nível de proficiência representado neste estudo (Quadro 3.8).

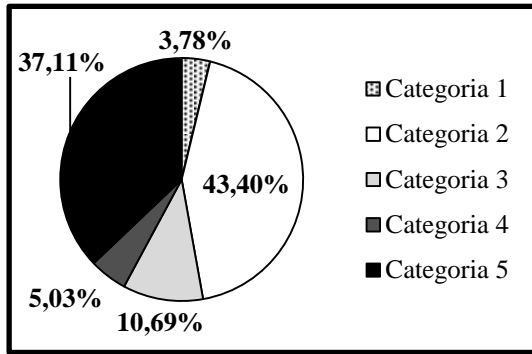


Gráfico 3.10 – Percentagem de desvios por categoria produzidos no nível A1/A1+

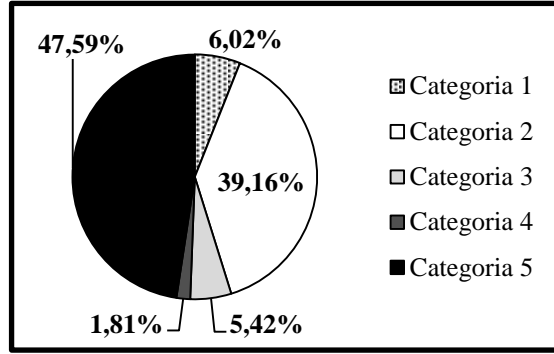


Gráfico 3.11 – Percentagem de desvios por categoria produzidos no nível A2/A2+

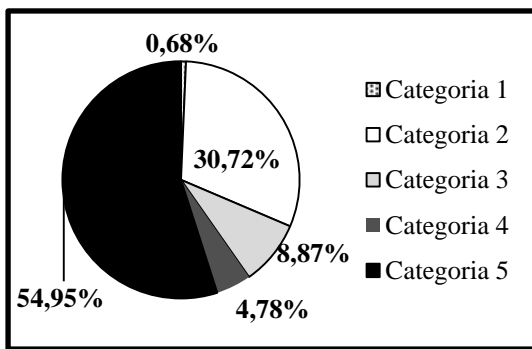


Gráfico 3.12 – Percentagem de desvios por categoria produzidos no nível B1

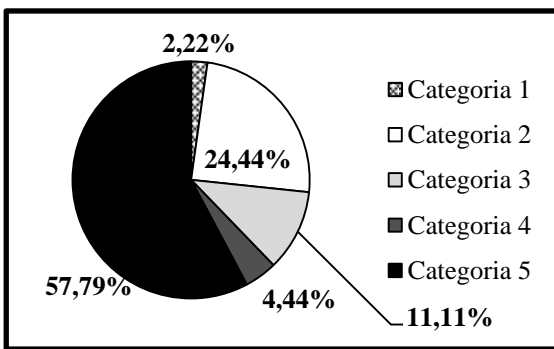


Gráfico 3.13 – Percentagem de desvios por categoria produzidos no nível B2

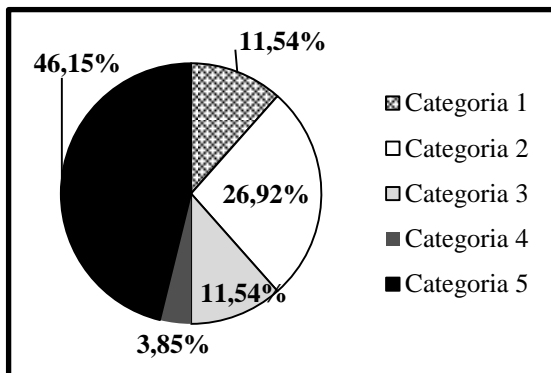


Gráfico 3.14 – Percentagem de desvios por categoria produzidos no nível C1

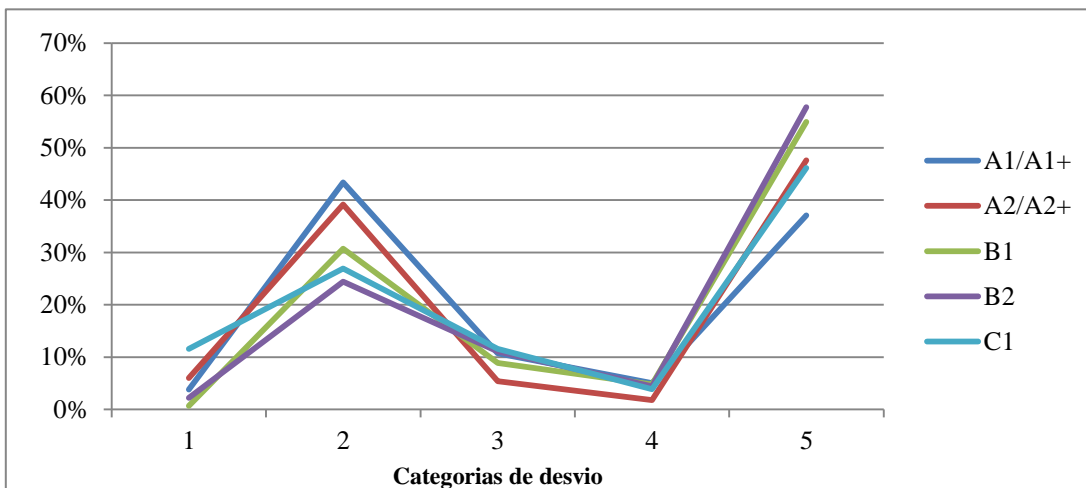


Gráfico 3.15 – Percentagem total de categorias de desvios produzidos por nível de proficiência dos aprendentes

Como é possível averiguar a partir dos Gráficos 3.10 a 3.15, em que figuram as percentagens de desvios calculadas em relação à totalidade de desvios produzidos por nível, os tipos de desvio mais frequentes em todos os casos são, tal como se verificara na análise geral aos dados, os relativos a nomes [-sexuados] cujo valor de GEN somente é inferível a partir dos marcadores sintáticos que coocorrem no GN (determinantes, adjetivos e quantificadores) [categoria 5] e os desvios relativos a itens nominais [-sexuados] em que o valor de GEN corresponde ao IT [categoria 2].

Desta forma, poder-se-á admitir que durante o processo de aquisição/aprendizagem da marcação de GEN nominal existe um padrão regular em relação ao tipo de desvio produzido, verificável em todos os níveis de aprendizagem formal de PL2 dos aprendentes.

No que concerne às subcategorias de desvio mais afetadas, observe-se os Gráficos 3.16 a 3.21 com os valores percentuais calculados em função do total de desvios produzidos por nível de proficiência em PL2.

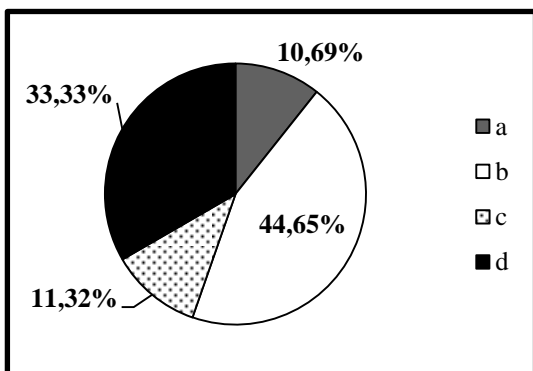


Gráfico 3.16 – Percentagem de desvios por subcategoria produzidos no nível A1/A1+

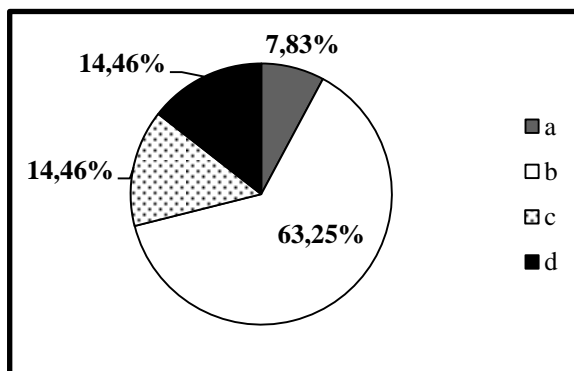


Gráfico 3.17 – Percentagem de desvios por subcategoria produzidos no nível A2/A2+

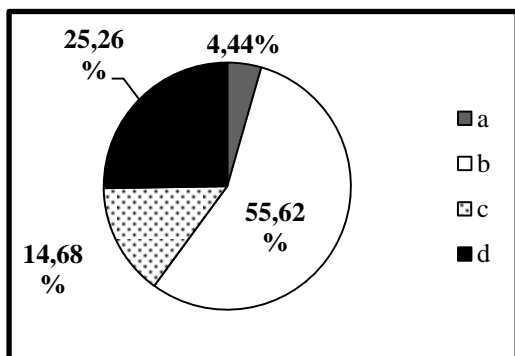


Gráfico 3.18 – Percentagem de desvios por subcategoria produzidos no nível B1

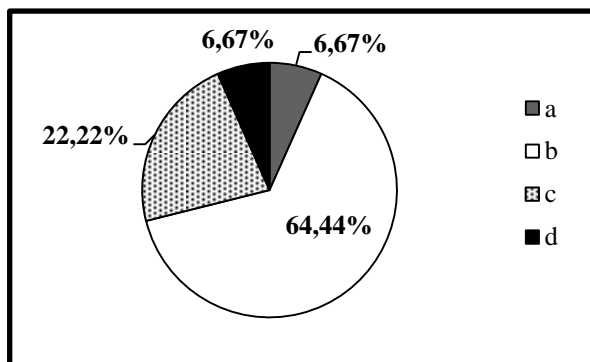


Gráfico 3.19 – Percentagem de desvios por subcategoria produzidos no nível B2

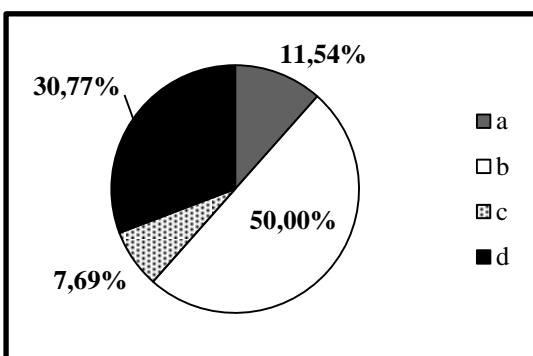


Gráfico 3.20 – Percentagem de desvios por subcategoria produzidos no nível C1

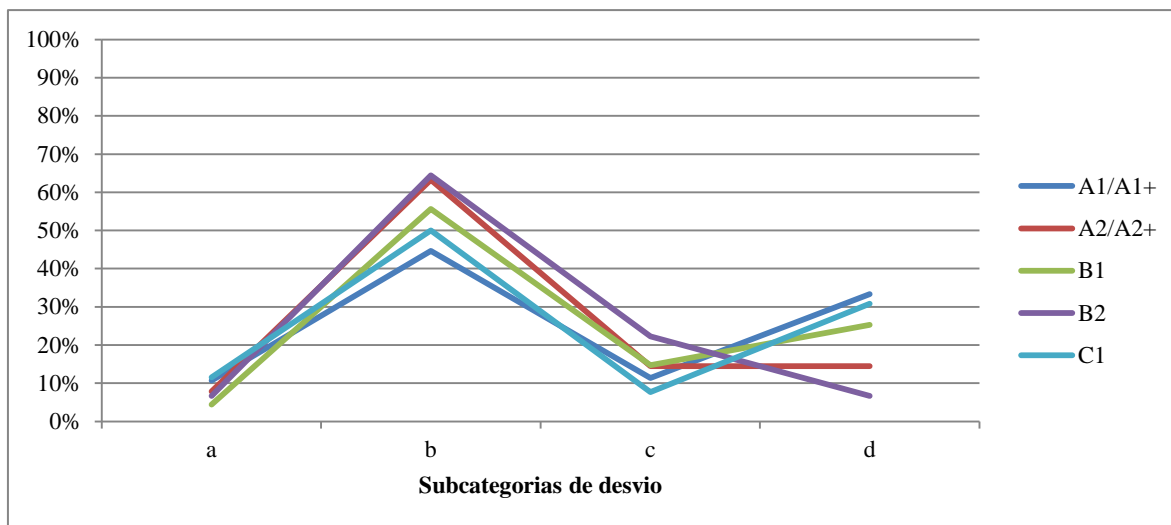


Gráfico 3.21 – Percentagem total das subcategorias de desvio produzidas por nível de proficiência

Os Gráficos 3.16 a 3.21 revelam que em todos os níveis de proficiência em PL2 dos informantes as subcategorias (b) – desvios marcados no determinante –, e (d) – o desvio ocorre no quantificador – são as que registam maior índice de desvios.

Além disso, dado que as subcategorias de desvio mais afetadas são as que dizem respeito à marcação de GEN pelo determinante e quantificador e sendo a atribuição de GEN nominal pelo adjetivo menos afetada, confirma-se que as fases de construção das IL dos aprendentes dão mais prioridade ao processamento dos itens lexicais sobre os gramaticais.

Sendo assim, e uma vez que em todos os níveis de proficiência dos informantes se verificam padrões semelhantes no que respeita aos tipos e subtipos de desvios, é possível considerar que a aquisição/aprendizagem de marcação de GEN corresponde a um processo cíclico, transversal a todos os níveis de proficiência. Assim, e perante os novos nomes a assimilar, os aprendentes parecem apresentar, reiteradamente, os mesmos processos de aprendizagem.

Em seguida apresenta-se a análise detalhada dos resultados obtidos nesta investigação, em função das categorias e subcategorias de desvio estipuladas por níveis de proficiência de PL2 dos aprendentes. Em cada categoria figura um gráfico com as percentagens de desvios calculadas em função do conjunto de desvios produzido por nível de proficiência dos informantes (cf. Quadro 3.8).

Após a análise detalhada dos resultados quantitativos, ilustrada com exemplos retirados do *corpus*, criou-se, para cada categoria, um Quadro com todos os itens nominais que originaram desvios por nível de aprendizagem formal dos aprendentes, de modo a verificar quais os nomes que propiciaram maior número de desvios de marcação de GEN.

3.4.2.1 CATEGORIA 1

No que respeita aos desvios da categoria 1, ou seja, aos desvios relativos a nomes sexuais cujo valor de GEN corresponde ao sexo do seu referente, verifica-se que a percentagem de desvios produzidos por nível é pouco acentuada (cf. Gráfico 3.22).

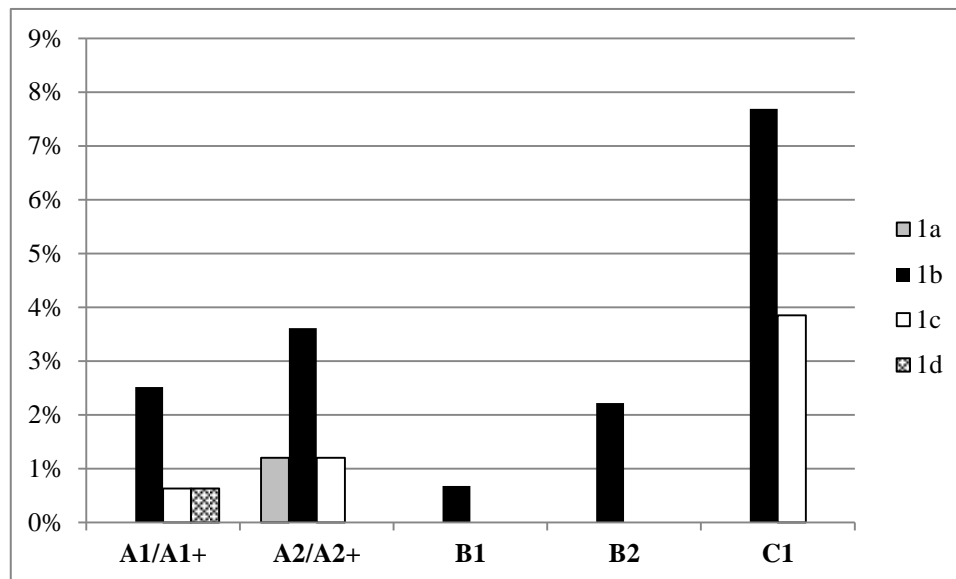


Gráfico 3.22 – Percentagem de desvios da categoria 1 em função do conjunto total de desvios produzidos por nível de proficiência

Sendo assim, estes dados parecem indicar que os aprendentes distinguem mais facilmente o valor de GEN dos nomes cuja distinção de GEN masculino/feminino corresponda a uma diferença de sexo (*o menino/ a menina; o rapaz/ a rapariga*). Além disso, uma vez que os critérios de ordem semântica são muito comuns nas línguas do mundo que possuem a categoria de GEN gramatical (Corbett, 1991:63), os aprendentes facilmente reconhecem o valor de GEN deste tipo de nomes (vid. Capítulo 2, secção 2.2).

Quanto à frequência de desvios por subcategoria, e através da leitura do Gráfico 3.22, verifica-se que os níveis iniciais (A1 e A2) apresentam maiores dificuldades em fazer a correta marcação de GEN destes nomes, principalmente no que diz respeito à marcação dos valores de GEN pelo determinante (1b), como por exemplo: «*Antes fui a Portugal, quando fico na China, gosto de fazer os comprados com **a minha** melhor

amigo.» [UC.CA.E.A.04.10.02.33.1J]; e adjetivo (1c), como ilustra o exemplo seguinte: «*Sou a **irmão** mais **nova**, assim tenho uma irmã maior, de 27 anos.» [UC.ER.LPII.A.12.09.09.55.2M].

Os níveis intermédios (B1 e B2) apresentam dificuldades em estabelecer a marcação de GEN dos nomes sexuados pelo determinante (1b), produzindo desvios como por exemplo: «*A vezes eu gosto de sair com **as minhas amigos** às lojas e pastelarias, e a um bar da noite. [UC.ER.LPIII.A.12.09.25.33.1J].

Por fim, no nível avançado (C1), refira-se, além da presença de desvios da subcategoria (1b): «*Então, ouvi dizer que /agora/ tu és **a capitão** de Universidade de Hokkaido.» [UC.CA.S.A.05.09.06.6.1B], desvios da marcação de GEN pelo adjetivo (1c): «*Mas para mim tanto me faz mesmo que sejas **boa capitão** ou não.» [UC.CA.S.A.05.09.06.6.1B].

Perante estes resultados quantitativos, importa em seguida identificar quais os nomes sexuados afetados e em que níveis foram produzidos. Deste modo, organizou-se no Quadro seguinte os nomes desta categoria por frequência de desvios produzidos em cada nível de proficiência em PL2 dos aprendentes.

Nível	Nomes e Frequência de ocorrência de desvios de marcação de GEN nominal
A1/A1+	<p>AMIGO (2) IRMÃ (2) AMIGA (1) PRIMO (1)</p> <p>Total de nomes afetados:4 Total de desvios:6⁴²</p>
A2/A2+	<p>AMIGA (2) AMIGO (1) FILHA (1) HOMEM (1) IRMÃO (1) NAMORADO (1) RAPAÇ (1)</p> <p>Total de nomes afetados:7 Total de desvios:8</p>

⁴² O total de desvios que figura nos Quadros 3.9 a 3.13 foi apurado a partir da multiplicação dos nomes pela frequência de desvio.

B1	<u>ALEMÃ</u> (2) AMIGO (1) AVÓ (1) Total de nomes afetados:3 Total de desvios:4
B2	<u>AMIGO</u> (1) Total de nomes afetados: 1 Total de desvios: 1
C1	<u>CAPITÃO</u> (3) Total de nomes afetados: 1 Total de desvios:3

Quadro 3.9 – Distribuição de nomes com desvio da categoria 1 e respectivas frequências de ocorrência de desvios por nível de proficiência dos aprendentes

Da leitura do Quadro 3.9 conclui-se que somente 16 nomes desta categoria deram origem a 22 desvios. Os níveis A1 e A2 apresentam maiores ocorrências desviantes, com 9 itens nominais afetados que correspondem, em grande parte, a nomes que denotam relações interpessoais (à exceção de *homem* e *rapaz*) (cf. Quadro 3.9).

A presença de desvios nestes itens nominais pode justificar-se, em parte, pelo facto de estes nomes serem os mais necessários em função dos temas dos textos produzidos para este *corpus* (cf. Anexo I) e, além disso, fazerem parte dos conteúdos programáticos deste nível inicial de aquisição/aprendizagem em PL2.

Nos restantes níveis, a frequência de nomes com desvio é consideravelmente menor, sendo que no nível C1 somente um item lexical deste tipo foi afetado e encontra-se num único texto produzido por um informante (cf. Anexo III).

Em suma, e como mostram os resultados, é possível concluir que em relação aos nomes sexuais cujo valor de GEN corresponde ao sexo do seu referente os aprendentes de PL2 são capazes de, com elevado grau de correção, assinalarem o valor de GEN, apresentando poucas dificuldades, principalmente quando atingem níveis de proficiência linguística superiores.

3.4.2.2 CATEGORIA 2

A categoria agora em análise identifica os desvios relativos a nomes com referentes extralinguísticos [-sexuados] cujo GEN é corretamente dedutível a partir do IT: se o IT é *-o*, então o nome é masculino; se o IT é *-a*, então o nome é feminino. Este tipo de nomes é muito frequente no *input* linguístico dos aprendentes. Com efeito, no estudo das frequências dos lemas nominais recenseados na CORLEX (cf. Capítulo 2), apurou-se que em cerca de 60% das frequências nominais se verifica a correspondência entre o GEN do nome e respetivo IT. Tal conclusão demonstra que em grande parte dos itens nominais os aprendentes de PL2 podem associar, de forma correta, o valor de GEN recorrendo a este critério de natureza formal.

No entanto, os resultados obtidos neste trabalho (cf. Gráfico 3.23) demonstram que os aprendentes, principalmente nos níveis elementares, embora possam saber que a marca final do nome indicia o seu valor de GEN, não são capazes de atribuir corretamente tal valor a cada nome nestas condições. Além disso, e tal como se assinalou anteriormente aquando da análise global dos desvios produzidos no *corpus* (secção 3.4.1), os valores acentuados nesta categoria de desvio parecem indicar que os alunos não dominam o valor de GEN destes nomes e demonstram, igualmente, uma certa “desconfiança” perante os indícios morfológicos. Na verdade, apesar de estes itens serem muito frequentes no *input* do aprendente, não são, ainda assim, suficientemente numerosos para o falante aprendente de PL2 assinalar corretamente os valores de GEN.

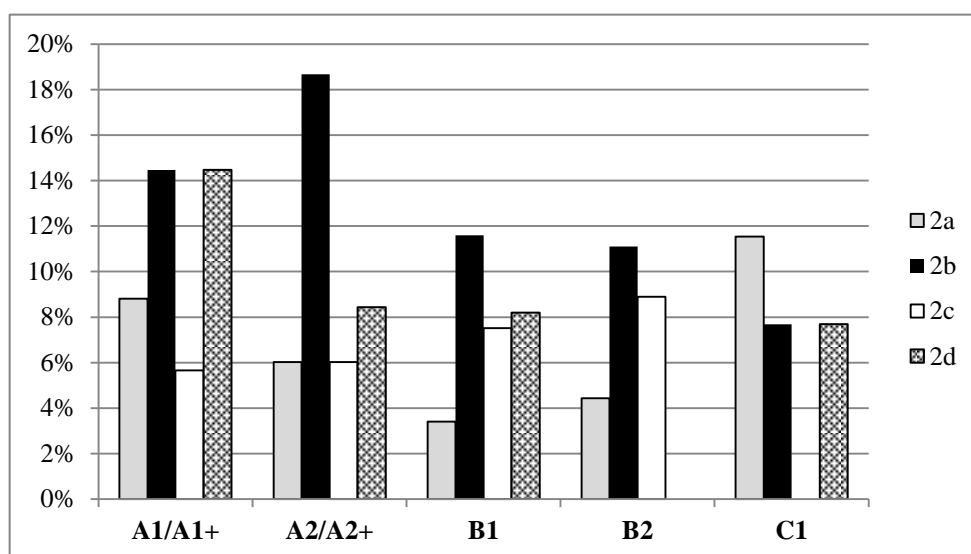


Gráfico 3.23 – Percentagem de desvios da categoria 2 em função do conjunto total de desvios produzidos por nível de proficiência

No que concerne à subcategoria de desvio mais afetada, e através da leitura do Gráfico 3.23, regista-se uma maior dificuldade dos aprendentes dos níveis iniciais (A1 e A2) em assinalar corretamente o valor de GEN deste tipo de nomes a partir do determinante (2b): «*Eu tenho uma desejo para esta noite: depois este test e quero ir para Figueira da Foz porque há uma concerto de Orchesta de Bairos e uma Coró.» [UC.ER.LPII.B.12.09.11.1.1A]; e a partir do quantificador (2d): «*Perto do Lagos estiveram muito praias com os pequenos restaurantes.» [UC.ER.LPI.A.05.10.09.6.1B].

Nos níveis intermédios (B1 e B2), os desvios são quantitativamente menos representativos, verificando-se maiores dificuldades em marcar o valor de GEN pelo determinante (2b): «*Há um Mini-Mercado pequeníssimo que pretende <ao> a um casal <m> muito velho e muito simpática e há a <loga> loja das lembranças /cópias/ internet onde sempre falo com o rapaz <p> brasileiro que faz os meus cópias.» [UC.ER.LPIII.A.12.09.68.77.3T].

Por fim, a subcategoria mais afetada no nível superior (C1) corresponde aos desvios marcados na forma morfológica do nome (2a), ou seja, situações em que os aprendentes tendem a modificar o IT e “formatam” o item nominal, colocando-lhe no final as vogais *-o*, *-a* ou *-e*. O exemplo seguinte ilustra este tipo de desvio: «*O sul de pais sempre faz calor, no oeste é por causa das desertas, mas no <o> este é por causa dos mares tropicais.» [UC.CA.S.A.12.09.08.50.2L]. O aprendente, apesar de não reconhecer o GEN da palavra *deserto*, pois coloca o determinante no feminino (**das desertas*), reconhece, no entanto, a correlação parcial que existe entre IT *-a* e GEN feminino, pois “formata” o nome adicionando-lhe o IT geralmente associado ao feminino: *-a* (cf. Anexo III).

Em seguida, o Quadro 3.10 apresenta os nomes produzidos em cada nível de proficiência dos aprendentes, organizados por frequência de desvio de marcação de GEN que originaram.

Nível	Nomes e frequência de ocorrência de desvios de marcação de GEN nominal			
A1/A1+	<u>COISA</u> (5)	ANO (1)	EXPERIÊNCIA (1)	QUARTA-FEIRA (1)
	<u>QUARTO</u> (5)	APARTAMENTO (1)	FERIADO (1)	RECEITA (1)
	<u>CULTURA</u> (4)	CARACTERÍSTICA (1)	JOGO (1)	SALA (1)
	LÍNGUA (4)	CASA (1)	LOJA (1)	TEMPO (1)
	CASA DE BANHO (3)	CERVEJA (1)	OLHO (1)	TERRA (1)
		COREOGRAFIA (1)	PALAVRA (1)	TOPOGRAFIA (1)
	COMIDA (2)	DESPORTO (1)	PASSEIO (1)	VIDA (1)
	COZINHA (2)	DISCOTECA (1)	PONTO (1)	
	FAMÍLIA (2)	ESCOLA (1)	PRAIA (1)	
	FESTA (2)	ESCRITÓRIO (1)		
	FUTURO (2)			
	MARCA (2)			
	NOTA (2)			
	PISO (2)			
	Total de nomes afetados: 39			
	Total de desvios: 63			
	A2/A2+	<u>COMIDA</u> (4)	ÁFRICA (1)	COPO (1)
<u>DESPORTO</u> (4)		ANO (1)	DESEJO (1)	MEMÓRIA (1)
CASA (3)		APARTAMENTO (1)	DISCOTECA (1)	MOMENTO (1)
EQUIPA (3)		ATMOSFERA (1)	DISTRITO (1)	MÚSICA (1)
PRATO (3)		BEBIDA (1)	ESTATURA (1)	NATA (<i>pastel de</i>) (1)
VIDA (3)		CABELO (1)	FOTOGRAFIA (1)	NATUREZA (1)
ALDEIA (2)		CENTRO (1)	FRANÇA (1)	ORATÓRIO (1)
FAMÍLIA (2)		CERVEJA (1)	INDONÉSIA (1)	PERGUNTA (1)
LÍNGUA (2)		CONCERTO (1)	INSECTO (1)	PRAIA (1)
MORADA (2)		COPA/TAÇA (1)	JUNTA DE	QUARTO (1)
SEMANA (2)			FREGUESIA (1)	TRABALHO (1)
TIPO (2)				
Total de nomes afetados: 44				
Total de desvios: 65				

B1	<u>COIMBRA</u> (9)	ALDEIA (1)	DISTÂNCIA (1)	PALÁCIO (1)
	<u>COISA</u> (8)	ALMA (1)	DITADURA (1)	PASSATEMPO (1)
	LOJA (5)	ANO (1)	EQUIPA (1)	PONTO (1)
	CASA (4)	ARQUITECTURA (1)	ESCOLA (1)	PRAIA (1)
	CULTURA (4)	AUTOCARRO (1)	ESTADO (1)	REGRA (1)
	EXPERIÊNCIA (3)	CAIPIRINHA (1)	ESTRANGEIRO (1)	SÁBADO (1)
	HORA (3)	CERIMÓNIA (1)	EXISTÊNCIA (1)	SERRA DA
	AVENTURA (2)	CERTEZA (1)	FORÇA (1)	ESTRELA (1)
	CAMPO (2)	CONCERTO (1)	HISTÓRIA (1)	SUPERMERCADO
	CERVEJA (2)	CONJUNTO (1)	IDEIA (1)	(1)
	LIVRO (2)	CÓPIA (1)	IMPORTÂNCIA (1)	TEMPO (1)
	SEMANA (2)	CORTEJO (1)	JOGO (1)	TERÇA-FEIRA (1)
	SILÊNCIO (2)	COZINHA (1)	MEDIDA (1)	VOLTA (1)
		DESCULPA (1)	MOMENTO (1)	
		DISCOTECA (1)	MONTANHA (1)	
			MOVIMENTO (1)	
	Total de nomes afetados: 54			
Total de desvios:90				
B2	<u>BONECO</u> (2)	CONFORTO (1)	DIFERENÇA (1)	VOCABULÁRIO (1)
	ANTILHAS (1)	CONTACTO (1)	LADO (1)	
	CAMA (1)	CULTURA (1)	RUA (1)	
Total de nomes afetados: 10				
Total de desvios:11				
C1	<u>DESERTO</u> (2)	CAMPO (1)	SÍTIO (1)	
	ÁREA (1)	DISCOTECA (1)	TEMPLO (1)	
Total de nomes afetados: 6				
Total de desvios:7				

Quadro 3.10 – Distribuição de nomes com desvio da categoria 2 e respetivas frequências de ocorrência de desvios por nível de proficiência dos aprendentes

Numa primeira abordagem ao Quadro 3.10, verifica-se que os níveis elementares (A1 e A2) apresentam maior número de nomes com desvio. Os itens afetados nestes níveis correspondem, em grande parte, a itens nominais simples, sendo que 12 nomes se

repetem nos dois níveis: *ano, apartamento, casa, cerveja, comida, desporto, discoteca, família, língua, praia, quarto e vida.*

Nos níveis intermédios (B1 e B2) regista-se uma maior diversidade lexical, com a presença de 49 itens nominais que não surgiam nos níveis A1 e A2. Verifica-se ainda um maior número de nomes derivados com desvio, tais como *distância, importância* (cf. Quadro 3.10). Além disso, somente um item lexical (*cultura*) se repete nos níveis B1 e B2. Convém assinalar que a frequência de desvios por item nominal é maior em relação aos níveis anteriores.

Por fim, o nível superior (C1) regista frequências de desvio muito reduzidas, em que somente os desvios de marcação de GEN associados aos itens nominais *campo e discoteca* se repetem nos restantes níveis.

Em suma, à medida que os níveis de proficiência em PL2 avançam, os aprendentes contactam com um maior número de unidades lexicais e, em consequência, é nesse novo conjunto de itens que se registam desvios de marcação de GEN nominal. Ainda assim, e através da leitura do Quadro 3.10, verifica-se que os nomes com desvios registados nos níveis elementares (A1 e A2) não apresentam problemas nos níveis seguintes. Este tipo de dados recorrentemente observados sugere que o GEN não é uma das primeiras propriedades do nome assimiladas quando esse nome é aprendido e que o padrão que preside à sua assimilação é reiteradamente ativado sempre que um novo nome é aprendido.

3.4.2.3 CATEGORIA 3

Na categoria 3 encontram-se os desvios de marcação de GEN relativos aos nomes não-sexuados, cujo valor de GEN não é corretamente dedutível a partir do IT, i.e., se o IT é *-o*, o nome é do GEN feminino (*a tribo*); se o IT é *-a*, então o nome possui valor de GEN masculino (*a dia*).

Verifica-se, a partir do Gráfico 3.24, que a percentagem de desvios deste tipo é muito reduzida. Estes resultados podem justificar-se, em parte, por este tipo de nomes ser pouco frequente no universo de itens lexicais em português, como se apurou no Capítulo 2 deste trabalho. Na verdade, e com base num estudo das frequências de lemas nominais recenseados na CORLEX, somente 0.04% da totalidade das ocorrências de nomes femininos possui IT *-o*, e 1.56% do conjunto absoluto das ocorrências de itens nominais masculinos possui IT *-a* (cf. Capítulo 2, secção 2.3.2).

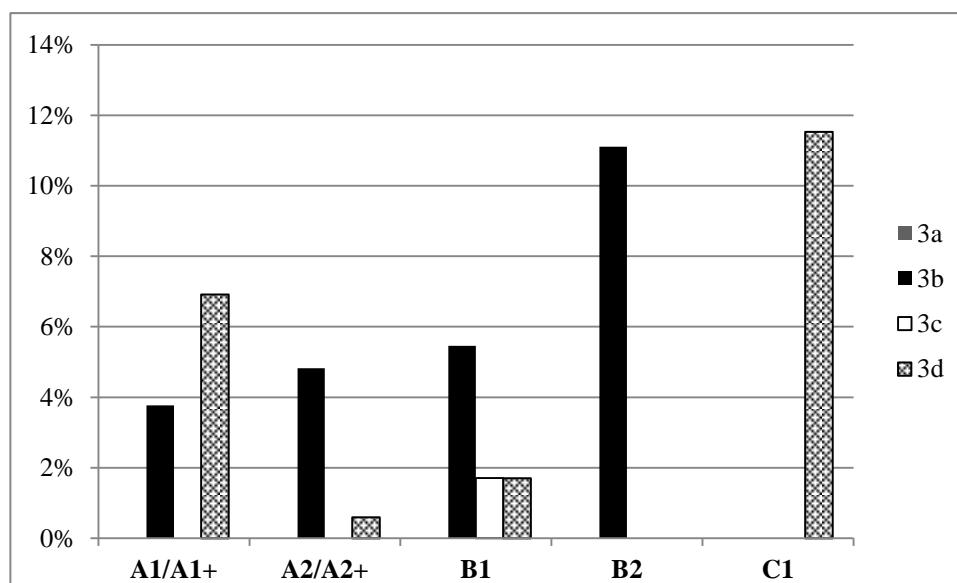


Gráfico 3.24 – Percentagem de desvios da categoria 3 em função do conjunto total de desvios produzidos por nível de proficiência

No que respeita à distribuição de desvios por subcategoria, observa-se que não se registam desvios na subcategoria 3a, ou seja, nenhum desvio marcado na forma morfológica do nome (cf. Gráfico 3.24). Este resultado é, em certa medida, previsível, uma vez que, sendo estes nomes, cujo IT *-a* corresponde ao GEN masculino e o IT *-o* corresponde ao GEN feminino, pouco frequentes no *input*, as formas **dio* ou **triba* seriam estranhas para os aprendentes.

Os níveis iniciais (A1 e A2) apresentam maior percentagem de desvios de marcação de GEN pelo determinante (3b): «*Nós gostamos de fazer comprado na baixa de cidade e vir filme **na cinema**.» [UC.CA.E.A.04.10.02.33.1J] e quantificador (3d): «*O meu deus, a outra noite, foi <(…)> tão bêbeda que, tive uma <resaca> /ressaca/ por **duas dias!**» [UC.ER.LPIII.A.12.09.04.6.1B].

Os aprendentes de níveis de proficiência B1 e B2 têm igual dificuldade em marcar o valor de GEN deste tipo de itens pelo determinante (3b): «*É o tempo de estudante **da programa** Erasmus, e apesar de ter mais tempo livre /aquí/, do que Republica Checa (porque aquí não tenho que trabalhar), <as> parece-me as vezes, que precisaria muito <mas> /mais/ tempo para fazer todas as coisas que gosto <(…)> de fazer.» [UC.ER.LPIII.A.12.09.19.33.1J].

No nível C1 somente se registam desvios de marcação de GEN pelo quantificador (3d): «*Quanto ao vestuário, as regiões têm os seus fatos tradicionais, mas para já, nós já não <tenha> temos o hábito de vestir estes fatos dia-a-dia, embora **alguns tribos** no oeste do país ainda <sejam> os <as> vistam.» [UC.CA.S.A.05.09.01.50.2L] (cf. Anexo III).

Em seguida, veja-se o Quadro 3.11 com a distribuição dos nomes e respectivos índices de ocorrências de desvio produzidos em cada nível comum de referência do QECRL representado neste *corpus*.

Nível	Nomes e frequência de ocorrência de desvios de marcação de GEN nominal
A1/A1+	<p style="text-align: center;"> <u>DIA</u> (10) <u>PROBLEMA</u> (5) CINEMA (1) FOTO (1) </p> <p>Total de nomes afetados: 4</p> <p>Total de desvios: 17</p>

A2/A2+	<u>CHAPA</u> ⁴³ (2) <u>PROBLEMA</u> (2) CINEMA (1) DIA (1) MAPA (1) PROGRAMA (1) TEMA (1) Total de nomes afetados: 6 Total de desvios: 9
B1	<u>DIA</u> (10) <u>PROBLEMA</u> (6) CINEMA (3) PROGRAMA (3) CLIMA (2) MAPA (1) TEMA (1) Total de nomes afetados: 6 Total de desvios: 26
B2	<u>DIA</u> (2) <u>CLIMA</u> (2) CINEMA (1) Total de nomes afetados: 3 Total de desvios: 5
C1	<u>CINEMA</u> (1) <u>PROBLEMA</u> (1) <u>TRIBO</u> (1) Total de nomes afetados: 3 Total de desvios: 3

Quadro 3.11 – Distribuição de nomes com desvio da categoria 3 e respetivas frequências de ocorrência de desvios por nível de proficiência dos aprendentes

Da análise do Quadro 3.11 conclui-se que, e tal como tínhamos referido anteriormente, o conjunto de itens nominais afetados é muito reduzido, correspondendo a um total de 22 nomes que originaram 60 desvios de marcação de GEN. No que se refere à frequência de desvios por nível, são os níveis iniciais (A1 e A2) e o nível

⁴³ O nome *chapa* corresponde, no contexto desta produção escrita, a um moçambicanismo. Veja-se a sua definição: “**chapa (cem)** *n. m. ou f.* transporte colectivo, semiformal; por extensão, qualquer automóvel que transporte pessoas a troco de algum dinheiro (de chapa, “preço único”, de cem meticais)” (informação disponível do endereço da *Cátedra Portuguesa Língua Segunda e Estrangeira*: http://www.catedraportugues.uem.mz/?_target=observatorio no link *Moçambicanismos, um glossário com algumas imagens*: <http://mocambicanismos.blogspot.com/2009/01/s.html>).

intermédio (B1) que registam um maior número de ocorrências desviantes, em relação ao que acontece nos restantes níveis.

Na verdade, os itens afetados são pouco diversificados, i.e., em todos os níveis de proficiência analisados há uma maior propensão para o desvio em nomes como *dia*, *problema* e *cinema*.

3.4.2.4 CATEGORIA 4

A frequência de desvios do tipo 4 – que compreende os nomes sexuais cujo valor de GEN não corresponde, necessariamente, ao sexo do referente (nomes epicenos, sobrecomuns e comuns de dois) – é, como ilustra o Gráfico 3.25, muito reduzida. Numa primeira abordagem, estes resultados evidenciam que este tipo de nomes não levanta muitas problemáticas aos aprendentes de PL2.

Para justificar o baixo índice de desvios neste tipo de nomes, algumas hipóteses podem ser equacionadas. Por um lado, nesta categoria de desvios encontram-se itens que admitem dois valores de GEN (*o/a estudante; o/a jornalista*) e, por isso, em grande parte destes casos não é possível averiguar, nesse contexto, se o aprendente atribuiu corretamente o valor de GEN. Por outro, em muitos destes nomes é possível os aprendentes associarem corretamente os valores de GEN tendo em conta o IT do item (*a pessoa, a criança*).

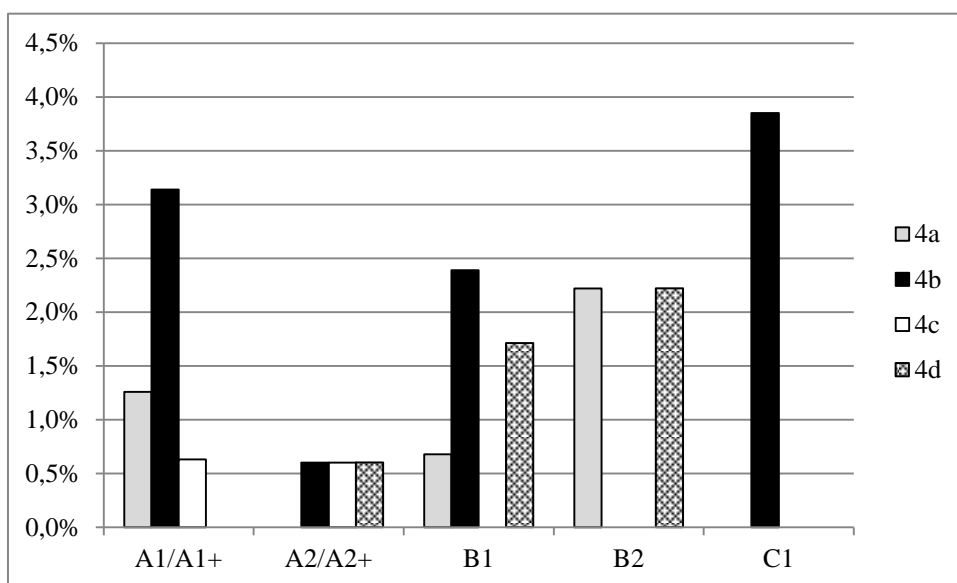


Gráfico 3.25 – Percentagem de desvios da categoria 4 em função do conjunto total de desvios produzidos por nível de proficiência

À exceção do nível de proficiência B2, a subcategoria de desvio mais frequente nos níveis de aprendizagem formal em PL2 dos informantes, é a subcategoria (4b), ou seja, os aprendentes têm dificuldade em marcar o valor de GEN destes nomes a partir do determinante, produzindo frases como «*Na parte norte, fala-se neerlandês; na parte sul,

fala-se francês e na parte este, há <(...)>**uns 70.000 pessoas** que têm a língua alemã como língua materna.» [UC.ER.LPIII.A.12.09.36.50.2L]».

Outra subcategoria mais representativa corresponde aos desvios marcados na forma morfológica destes itens (4a), ou seja, quando o aluno “formata” o nome colocando-lhe no final a vogal *-o* ou *-a*. Nestes casos, os aprendentes “formatam” o nome de acordo com a regra “se o nome termina em *-o*, é do GEN masculino; se o nome termina em *-a*, logo é do GEN feminino”. Veja-se o seguinte enunciado produzido por um aluno do nível A1: «*Fui muitas experiências aqui enquanto viajar, festas com os **novos colegas** e, as vezes, ir para escola.» [UC.ER.LPI.A.05.10.17.6.1B]. Como se pode verificar, o aprendente, além de associar ao nome o adjetivo masculino *novo*, “formata” o nome *colega*, colocando-lhe a marca final associada o masculino: *-o*, pois assim, e no entender do informante, estaria em consonância com o valor de GEN do adjetivo que selecionou (cf. Anexo III) (veja-se a página 18 deste trabalho).

Apresentados os dados quantitativos, verifique-se quais os nomes com desvios de marcação de GEN nesta categoria (cf. Quadro 3.12).

Nível	Nomes e frequência de ocorrência de desvio de marcação de GEN nominal
A1/A1+	<p style="text-align: center;"><u>PESSOA</u> (5) COLEGA (2) ESTUDANTE (1)</p> <p>Total de nomes afetados: 3 Total de desvios: 8</p>
A2/A2+	<p style="text-align: center;"><u>PESSOA</u> (3)</p> <p>Total de nomes afetados: 1 Total de desvios: 3</p>

B1	<p><u>PESSOA</u> (5) <u>TURISTA</u> (5) CICLISTA (1) COLEGA (1) CRIANÇA (1) ESTUDANTE (1)</p> <p>Total de nomes afetados: 5 Total de desvios: 14</p>
B2	<p><u>TURISTA</u> (2)</p> <p>Total de nomes afetados: 1 Total de desvios: 2</p>
C1	<p><u>COLEGA</u> (1)</p> <p>Total de nomes afetados: 1 Total de desvios: 1</p>

Quadro 3.12 – Distribuição de nomes com desvio da categoria 4 e respectivas frequências de ocorrência de desvios por nível de proficiência dos aprendentes

A partir da análise do Quadro 3.12 observa-se que o conjunto de nomes afetados é muito reduzido, correspondendo a um total de 10 nomes com 28 ocorrências desviantes de marcação de GEN.

Nos níveis iniciais (A1 e A2) o conjunto de nomes deste tipo é reduzido, embora se registre um elevado número de desvios por item nominal. O nível B1 regista um maior número de itens e, conseqüentemente, um maior número de desvios. Os restantes níveis de proficiência em PL2 (B2 e C1) registam um item lexical com desvio por nível.

Saliente-se ainda que, à exceção dos sobrecomuns (*a pessoa* e *a criança*), os itens com desvio são comuns de dois GEN (*o/a colega, o/a estudante, o/a turista, o/a ciclista*). Nestes casos, grande parte dos desvios apurados dizem respeito à atribuição do GEN feminino quando os aprendentes se referiam ao conjunto de entidades (*colegas, turistas, estudantes, ciclistas*) (vid. Anexo III).

Em suma, e através da análise destes dados, é possível concluir que, durante a aquisição/aprendizagem da marcação de GEN nominal, os aprendentes apresentam poucas dificuldades com este tipo de nomes.

3.4.2.5 CATEGORIA 5

A última categoria para o estudo dos desvios de marcação de GEN nominal por aprendentes de PL2 compreende os relativos a itens nominais com referentes extralinguísticos [-sexuados] cujo valor de GEN somente é dedutível a partir de marcadores sintáticos (determinante, adjetivo, quantificador). No Gráfico 3.26 figuram as percentagens de desvios desta categoria e, como é possível depreender, este tipo de nomes apresenta maior número de casos de marcação de GEN desviante, em relação às restantes categorias da tipologia.

Por conseguinte, confirma-se que, tal como estipulámos no início da análise dos resultados, quando os aprendentes não se podem socorrer de critérios de marcação de GEN quer de natureza semântica, quer de natureza formal, como o IT, tendem a produzir um maior número de desvios. Todavia, e como a segunda categoria mais afetada é a categoria 2 (cf. secção 3.4.2.2), também não é possível afirmar, com certeza, que o critério formal seja usado com muita confiança pelos aprendentes.

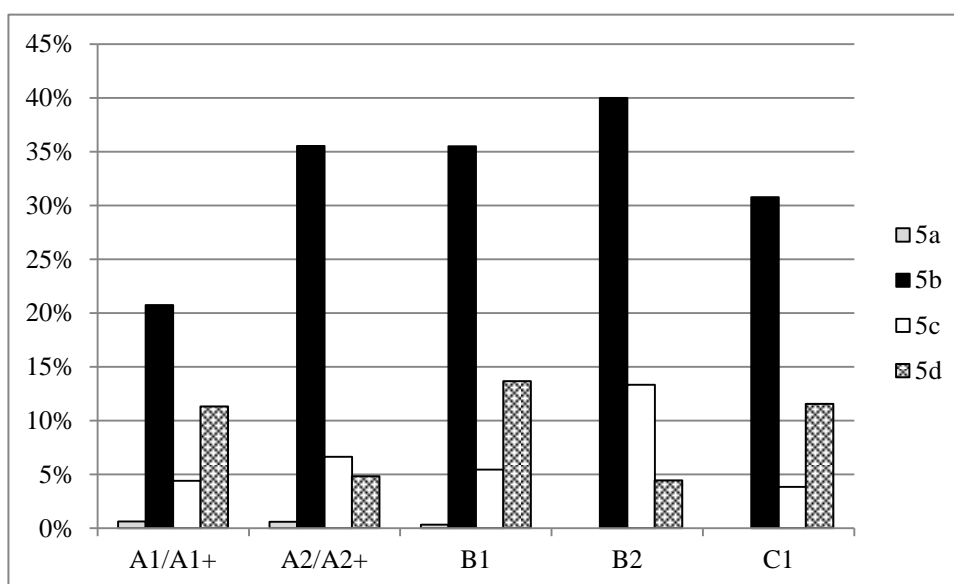


Gráfico 3.26 – Percentagem de desvios da categoria 5 em função do conjunto total de desvios produzidos por nível de proficiência

Quanto à distribuição de desvios por subcategoria, verifica-se em todos os níveis de aprendizagem formal em PL2 dos informantes (cf. Gráfico 3.16) a existência de maiores dificuldades na correta marcação de GEN destes nomes pelo determinante (5b): «*Como o trabalho de investigação força-me ficar <(…)> no laboratório (ou o que

seja) quase o tempo todo, <durante> fora do trabalho gosto **dos atividades** que se desenvolvem na natureza.» [UC.ER.LPIII.A.12.09.64.33.1J].

A segunda subcategoria de desvio mais afetada compreende os desvios de atribuição dos valores de GEN dos nomes a partir do quantificador (5d): «*Certamente há também **alguns desvantagens**.» [UC.ER.LPIII.A.12.09.19.69.3Q].

Referidos os resultados quantitativos, veja-se no Quadro 3.13 quais os nomes que originaram desvios desta categoria, em função do nível de proficiência dos aprendentes de PL2.

Nível	Nomes e frequências de ocorrências de desvio de marcação de GEN nominal		
A1/A1 +	<u>VIAGEM</u> (10) <u>CIDADE</u> (8) VEZ (6) EXAME (4) CAFÉ (3) RESTAURANTE (3) FIM DE SEMANA (2) MÊS (2) SAUDADE (2)	ACTIVIDADE (2) COMPUTADOR (1) FACULDADE (1) FIM (1) GALÃO (1) GARAGEM (1) INGLÊS (1) KARAOKE (1) NACIONALIDADE (1)	NORTE (1) OPORTUNIDADE (1) PÃO (1) PARQUE (1) PROFISSÃO (1) SITUAÇÃO (1) SOFÁ (1) TELEVISÃO (1) UNIVERSIDADE (1) VERÃO (1)
	Total de nomes afetados: 28		
	Total de desvios: 59		
A2/A2+	<u>VIAGEM</u> (23) <u>CIDADE</u> (8) CARÁCTER (7) PARTE (4) ACTIVIDADE (3) PAISAGEM (3) AVIÃO (2) EMOÇÃO (2) FACULDADE (2) INTERNET (2) TORRE (2) UNIVERSIDADE (2) VEZ (2)	ÁRVORE (1) CAPITAL (1) CARNAVAL (1) CATEDRAL (1) CELAS (1) COMPUTADOR (1) ESTAÇÃO (1) LEGUME (1) MUSEU (1) PAÍS (1) PARQUE (1)	PASTEL (1) RESTAURANTE (1) SAÚDE (1) SORTE (1) TRANSPORTE (1) VERÃO (1)
	Total de nomes afetados: 30		
	Total de desvios: 79		

B1	<u>CAFÉ</u> (7) <u>CIDADE</u> (7) ÁRVORE (6) DESVANTAGEM (5) FIM DE SEMANA (5) INTERNET (5) POSSIBILIDADE (5) SAUDADE (5) VANTAGEM (5) ARTE (4) DOR (4) LUGAR (4) REFEIÇÃO (4) SITUAÇÃO (4) VIAGEM (4) ACTIVIDADE (3) ALPES (3) CAPITAL (3) COR (3) PAÍS (3) PARTE (3) TORRE (3) TRADIÇÃO (3) VEZ (3)	CHOCOLATE (2) EXPOSIÇÃO (2) NET (2) NOITE (2) ORIGEM (2) PAISAGEM (2) PONTE (2) RESTAURANTE (2) VERÃO (2) ALGARVE (1) APRENDIZAGEM (1) ATENÇÃO (1) AVE (1) BAR (1) CALOR (1) CANÇÃO (1) CARNE (1) CASAL (1) CATEDRAL (1) CHATICE (1) COMPETIÇÃO (1) CORAÇÃO (1) COSTUME (1) DESORDEM (1) DETALHE (1) DOLCE VITA (1) EMOÇÃO (1)	ESTAÇÃO (1) FACULDADE (1) GENTE (1) GRIPE (1) HIPÓTESE (1) HORIZONTE (1) HOSPITAL (1) LEGUME (1) LIGAÇÃO (1) MÃO (1) MÊS (1) MODIFICAÇÃO (1) NATAL (1) OCASIÃO (1) PAREDE (1) POMBAL (1) REALIDADE (1) RELIGIÃO (1) SORTE (1) TESTE (1) T-SHIRT (1) UNIVERSIDADE (1) VARIAÇÃO (1) VOLUME (1)
B2	<u>REGIÃO</u> (5) <u>CAPITAL</u> (3) CONTRADIÇÃO (2) PAISAGEM (2)	ANIMAL (1) CATEDRAL (1) CIDADE (1) COR (1) HOSPITAL (1) INFORMAÇÃO (1)	LEITE (1) OPINIÃO (1) ORGANIZAÇÃO (1) PARQUE (1) SAUDADE (1) TRADIÇÃO (1) VANTAGEM (1) VIAGEM (1)
C1	<u>CIDADE</u> (3) ACTIVIDADE (1) BALÃO (1) CAPITAL (1)	LUGAR (1) MATERIAL (1) OPORTUNIDADE (1) PARTE (1)	UNIVERSIDADE (1) VEZ (1)

Quadro 3.13 – Distribuição de nomes com desvio da categoria 5 e respectivas frequências de ocorrência de desvios por nível de proficiência dos aprendentes

No Quadro 3.13 figuram 161 nomes que originaram 337 desvios de marcação de GEN. Como se pode depreender, regista-se um elevado número de desvios nos nomes não sexuados cujo valor somente é inferível através de marcadores sintáticos (determinante, quantificador, adjetivo), compreendendo tanto itens nominais simples como derivados.

Nos níveis iniciais (A1 e A2) regista-se uma maior frequência de desvios por nome em relação aos restantes níveis de proficiência dos informantes (cf. Quadro 3.13). Por exemplo, no nível elementar A2 um único nome como *viagem* regista 28 ocorrências desviantes. Por sua vez, nos níveis mais avançados – B1, B2 e C1 – regista-se um maior número de nomes com desvio, embora a frequência de desvios de atribuição de GEN por item nominal seja menos acentuada em relação à que observa nos níveis elementares.

Apesar de os desvios de marcação de GEN em nomes desta categoria afetarem um conjunto diversificado de itens, é possível identificar, e com base nos dados registados no Quadro 3.13, os nomes mais problemáticos. Assim, o nome *viagem* regista nos níveis elementares – A1 e A2 – a maior frequência de desvios, reaparecendo, embora com menos desvios de atribuição de GEN, nos níveis B1 e B2. Em seguida, o item nominal com maior frequência de desvios nos níveis elementares corresponde ao nome *cidade* que regista, ainda, desvios nos restantes níveis de proficiência. De fato, no nível superior (C1) *cidade* apresenta mais desvios em relação aos restantes nomes afetados. Por fim, o nome *vez* regista, igualmente, frequências de desvio acentuadas no nível elementar A1, reaparecendo nos restantes níveis de proficiência, à exceção do nível B2.

Em suma, os nomes que figuram na presente categoria de desvio, por não obedecerem a um critério de marcação de GEN nominal, quer semântico, quer de natureza formal apresentam um elevado índice de desvios em todos os níveis de proficiência dos informantes.

3.4.3 SÍNTESE DOS RESULTADOS APURADOS

No decorrer deste capítulo apresentámos os dados obtidos nesta investigação, bem como a análise detalhada dos desvios de marcação de GEN nominal produzidos pelos informantes que integram o *corpus*. Num primeiro momento, e com base nas características da estrutura do GEN gramatical em português, criou-se uma tipologia de casos desviantes que serviu de ponto de partida para o estudo dos dados selecionados.

Assim, e na tentativa de apurar possíveis padrões durante o processo de aquisição/aprendizagem da marcação de GEN nominal pelos aprendentes de PL2, procedeu-se à análise das frequências de desvio por nível de proficiência dos informantes, tendo em conta as categorias e subcategorias de desvio criadas na tipologia. Verificou-se que os níveis elementares registaram maiores ocorrências de desvios em todas as categorias, embora nos níveis mais avançados também se verificassem casos desviantes de marcação de GEN.

Durante o desenvolvimento do seu sistema interlinguístico os aprendentes apresentam desvios de marcação de GEN variáveis, em função do tipo de nome afetado. Assim sendo, e como se apurou, no decorrer do processo de aquisição/aprendizagem em PL2 os aprendentes vão reconhecendo, em parte, algumas das pistas que lhes permitem identificar corretamente os valores de GEN dos nomes, embora os problemas de marcação de GEN nunca cheguem a ser erradicados.

A análise de desvios contribuiu, ainda, para identificar um conjunto de nomes mais problemáticos, no que se refere à atribuição dos valores de GEN. De facto, nomes cujo valor de GEN masculino/feminino corresponda ao sexo do seu referente apresentam menores casos desviantes, em relação aos itens nominais em que o critério de natureza semântica não atua. É o que ocorre nos nomes [-sexuados] que, como se viu, apresentam maiores frequências de desvios. Na verdade, os nomes [-sexuados] cujo valor de GEN só é inferível a partir de marcadores sintáticos, tais como o determinante, quantificador e adjetivos registaram os maiores índices de desvio.

Foi igualmente interessante verificar que o outro conjunto de itens mais afetados com desvios de marcação de GEN corresponde aos nomes [-sexuados] cujo IT permite identificar corretamente o seu valor de GEN. Tal resultado parece indicar que os aprendentes possivelmente não confiam nos indícios morfológicos, como o IT, para a correta atribuição do valor de GEN.

Por fim, este estudo permitiu demonstrar que a aquisição/aprendizagem da marcação de GEN corresponde a um processo cíclico, transversal a todos os níveis de proficiência em que, perante um novo nome assimilado, os falantes aprendentes de PL2 apresentam reiteradamente os mesmos problemas. Além disso, se o ciclo de aquisição/aprendizagem do GEN é feito item a item, então a regra de atribuição de GEN parece não ter, em parte, muita relevância, uma vez que o aprendente vai fixar o valor de GEN de cada vez que aprende um novo nome.

Considerações Finais

Ao longo das últimas décadas muitos linguistas dedicaram especial atenção ao processo de aquisição/aprendizagem de uma língua não-materna (L2), realçando algumas das múltiplas particularidades que envolvem este processo. No que respeita à aquisição/aprendizagem da marcação do género gramatical (GEN) numa L2, muitos aprendentes, sobretudo os tardios, revelam algumas dificuldades em estabelecer a correta associação dos valores de GEN em todos os itens nominais dessa L2, mesmo em fases posteriores do seu desenvolvimento linguístico. Perante esta evidência, e motivados pelo facto de nenhum estudo, até ao momento, se ter debruçado sobre a aquisição/aprendizagem da marcação de GEN nominal em português como L2, realizámos esta dissertação que, neste momento, se conclui.

Neste trabalho pretendeu-se identificar distintas fases do processo de aquisição/aprendizagem da marcação de GEN nominal em PL2 que, além disso, fossem reveladoras do processo de construção da *Interlíngua* (IL) dos aprendentes. Deste modo, procedeu-se à análise exaustiva dos desvios de atribuição de GEN nominal produzidos pelos informantes que redigiram o *corpus* de produções escritas utilizado nesta investigação.

No primeiro Capítulo, fez-se uma revisão dos conceitos integrados no âmbito da Aquisição de Línguas Segundas, principalmente no que diz respeito à descrição do termo *Interlíngua* e das suas características, tendo em conta a proposta de Larry Selinker (1992; 1993; 2008). Através desta abordagem inicial, foi possível averiguar quais os factores que tipicamente fazem parte do processo da construção dos sistemas interlinguísticos dos falantes aprendentes de uma L2, tais como a simplificação do material linguístico da L2, evidenciando-se, por exemplo, o não emprego dos artigos, a ausência de marcação do plural ou a omissão de formas verbais. Além disso, quando o aprendente assume que atingiu, de forma satisfatória, os objetivos de comunicação da língua-alvo (LA), este pode considerar que a sua IL está apta a desempenhar as funções comunicativas dessa língua e, nesse caso, tende a “estabilizar” determinadas estruturas linguísticas, cessando o desenvolvimento da IL

Em seguida, no Capítulo 2, dedicado à noção de GEN, apresentou-se, num primeiro momento, a definição deste conceito com base no trabalho desenvolvido por Greville Corbett (1991). Tal exposição permitiu realçar as particularidades idiossincráticas desta categoria gramatical, bem como a sua relevância para o contexto de aquisição/aprendizagem de uma L2. Sinteticamente, poderemos dizer que o GEN não é uma categoria presente em todos os sistemas linguísticos do mundo e, mesmo naqueles que possuem esta categoria gramatical, os critérios de atribuição de GEN são muito variáveis. Posteriormente, procedeu-se à descrição detalhada do sistema de marcação de GEN nominal na língua portuguesa, a partir das posições tomadas por autores como Joaquim Mattoso Câmara Jr. (1966; 1994) e Alina Villalva (1994; 2003; 2008). Assim, foi possível apurar que, em português, a atribuição de um valor de GEN aos nomes resulta de um cruzamento de critérios semânticos e de critérios formais. No âmbito destes últimos, é particularmente relevante assinalar a correlação parcial patente em português entre o Índice Temático (-o ou -a) do nome e respetivo GEN.

Ainda neste Capítulo, realizou-se um estudo sobre os índices de frequências das ocorrências dos lemas nominais recenseados na base de dados CORLEX, com o objetivo de apurar a distribuição relativa dos nomes [-sexuados] de GEN masculino e feminino por classe temática. Tal exercício permitiu averiguar que, em quase 60% das frequências nominais da língua portuguesa existe correlação entre os valores de GEN dos nomes e respetivo IT, facto que nos levaria considerar que este poderia ser um critério seguro em que os aprendentes de PL2 se poderiam apoiar.

Por fim, no último Capítulo, procedeu-se à análise dos desvios de atribuição de GEN produzidos por aprendentes de PL2. Na primeira parte, descreveu-se detalhadamente o método de recolha das produções escritas que integram o *corpus* em estudo. Além disso, forneceram-se informações detalhadas a respeito do perfil dos informantes, referindo-se a sua distribuição por idade, ano de início de estudo em PL2, língua materna (LM) e nível de proficiência em PL2.

Em seguida, elaborou-se uma tipologia dos casos desviantes de atribuição de GEN assinalados no *corpus* da investigação. Esta tipologia, baseada nas características do sistema de marcação de GEN nominal do português, serviu de ponto de partida para a análise dos dados apurados. Desta forma, registaram-se os desvios de atribuição do valor de GEN nominal observáveis quer através da forma morfológica do item lexical, quer através das demais palavras que com ele deverão concordar em GEN, coocorrendo no grupo nominal (determinantes, adjetivos e quantificadores). Para cada categoria de

desvio apurada foi atribuído um código. Assim, o código: (1) identificava os desvios relativos a nomes [+sexuados], em que o valor de GEN coincide com o sexo do referente; (2) identificava os desvios relativos a nomes [-sexuados] cujo GEN seria corretamente dedutível a partir do IT: se o IT é *-o*, então o nome é masculino; se o IT é *-a*, então o nome é feminino; (3) identificava os desvios relativos a nomes [-sexuados] cujo GEN não é dedutível a partir do IT: quando o IT é *-o*, o item nominal é feminino; quando o IT é *-a*, o item nominal é masculino; (4) identificava os desvios relativos a nomes [+sexuados] cujo valor de GEN não coincide, necessariamente, com o sexo do referente; (5) identificava os desvios relativos a nomes [-sexuados] cujo valor de GEN somente é inferível a partir de marcadores sintáticos, como determinantes, adjetivos ou quantificadores.

Além disso, em cada categoria de desvio consideraram-se quatro subcategorias que identificavam os diferentes constituintes do grupo nominal nos quais se evidenciava(m) a(s) marca(s) do desvio: (a) o desvio marcado na forma morfológica do nome; (b) o desvio marcado no determinante; (c) o desvio marcado no adjetivo; (d) o desvio marcado no quantificador.

Apurada a tipologia, procedeu-se ao tratamento e análise dos dados do *corpus*, o que permitiu cumprir alguns dos objetivos delineados para esta investigação.

Assim sendo, verificou-se qual a frequência dos casos desviantes em função dos níveis de proficiência dos informantes. Verificou-se que os desvios relativos a nomes [+sexuados] em que o GEN corresponde ao sexo do seu referente apresentam, em todos os níveis de proficiência, menores problemas. Pelo contrário, os desvios mais frequentes em todos os níveis de proficiência de PL2 dos informantes são os relativos a nomes [-sexuados] cujo valor de GEN somente é inferível a partir dos marcadores sintáticos que coocorrem no grupo nominal, bem como os desvios relacionados com itens nominais [-sexuados] em que o valor de GEN é inferível a partir do seu IT.

No que se refere às subcategorias de desvio, foi possível constatar que, em todos os níveis de proficiência, os desvios marcados no determinante (b) e os desvios marcados no quantificador (d) são os mais frequentes.

Embora os índices de desvio sejam menores nos níveis de proficiência mais elevados, a verdade é que os desvios de atribuição de GEN nunca chegam a ser erradicados, ou seja, dificilmente os aprendentes conseguem, de modo pleno, atribuir corretamente os valores de GEN aos nomes.

Além da análise detalhada dos resultados quantitativos, verificou-se, em função das categorias de desvio, quais os nomes mais *propícios* aos desvios de atribuição de GEN em cada nível de proficiência representado no *corpus*. Este trabalho permitiu-nos constatar que, de um modo geral, os nomes mais problemáticos nos níveis inferiores não se repetem nos níveis mais avançados.

Por conseguinte, os resultados obtidos parecem sugerir que, no processo de aquisição/aprendizagem lexical, a assimilação do valor de GEN de um nome é progressiva, relativamente lenta e reiterativa. Assim sendo, estes dados permitem-nos confirmar que existem fases distintas durante a construção dos sistemas interlinguísticos destes informantes no que respeita à marcação de GEN nominal em português. Os dados indiciam ainda, tendo em conta as constatações feitas a respeito de aprendentes de vários níveis de proficiência em PL2, que, a cada novo nome assimilado, todo o sistema se repete, i.e., que o aprendente terá de passar pelas mesmas fases.

Em todo o caso, a aquisição/aprendizagem de novos itens passa por um certo grau de estabilização dos valores de GEN dos nomes já conhecidos e disponíveis no acervo de léxico produtivo do aprendente, já que são poucos os nomes que aparecem persistentemente com desvios de atribuição de GEN nos falantes aprendentes representativos dos diferentes níveis de proficiência em relação a PL2.

Acrescente-se ainda que os resultados obtidos também sugerem uma certa “desconfiança” do aprendente em relação à fiabilidade do IT como marcador de GEN. Esta mesma “desconfiança” é notória nos aprendentes iniciais, mas nunca desaparece de modo completo, visto que, com a expansão do léxico observável em fases posteriores do desenvolvimento linguístico de PL2 dos informantes, há indícios que acusam a sua presença.

Concluída a presente dissertação, resta assinalar que este trabalho pretende, sobretudo, incentivar a realização de trabalhos futuros sobre esta área de investigação. De facto, seria pertinente alargar o estudo dos desvios de atribuição de GEN a um grupo mais vasto de aprendentes de PL2, de modo a apurar se a tendência assinalada aqui se continua reiteradamente a verificar. Seria interessante, por exemplo, elaborar um estudo sobre o papel que a LM dos aprendentes pode ter durante o processo de aquisição/aprendizagem da marcação de GEN. Além disso, este estudo pretende servir de ponto de referência aos professores de português língua não-materna, ao assinalar as principais problemáticas dos aprendentes em relação à marcação de GEN nominal. Na

verdade, seria igualmente interessante, em futuras investigações, explorar com maior detalhe o conjunto de nomes mais propícios aos desvios de atribuição de GEN.

Bibliografia

Alves, José Matias (dir.) (2001), *Quadro Europeu Comum de Referência para as Línguas, aprendizagens, ensino, avaliação*. Lisboa: ASA.

Baddeley, A. (1999), *Memory*. In: Wilson, R.A & Keil, F. C. (Orgs.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, Mass: MIT Press, pp. 514-517.

Bechara, Evanildo (1999), *Moderna Gramática Portuguesa*, Rio de Janeiro: Editora Lucerna.

Bhatia, Tej. K. & Ritchie, William C. (Eds.), *The Handbook of Second Language Acquisition*, San Diego/California: Academic Press, pp.1-46.

Booij, Geert E. (2005), *The Grammar of Words: an Introduction to Linguistic Morphology*, Oxford: University Press.

Carvalho, J. G. Herculano de (1979), *Teoria da Linguagem*, vol. 2, Coimbra: Atlântida Editora.

Carvalho, J. G. Herculano de (2000), *Género*. In: *Enciclopédia Verbo Luso-Brasileira da Cultura*, vol. 13, Lisboa/São Paulo: Verbo, pp. 183-186.

Chomsky, Noam (1975), *Aspectos da Teoria da Sintaxe* (J. A. Meireles e E. P. Raposo, trad.), Coimbra: Arménio Amado.

Corbett, Greville G. (1991), *Gender*, Cambridge: Cambridge University Press.

Corder, S.P. (1992), *La importancia de los errors del que aprende una lengua segunda*. In: Licerias, J.M., *La Adquisición de las lenguas extranjeras: hacia un modelo de análises de la interlengua*, Madrid: Visor, pp. 31-40.

Cunha, Celso e Lindley, L. F. (2005), *Nova Gramática do Português Contemporâneo*, 18ª edição, Lisboa: Editora Sá da Costa.

Ellis, Rod (1986), *Understanding Second Language Acquisition*, Oxford: Oxford University Press.

Ellis, Rod (2000), *The Study of Second Language Acquisition* (7ª edição), Oxford: Oxford University Press.

Ellis, Rod (2003), *Second Language Acquisition* (9ª edição), Oxford: Oxford University Press.

Franceschina, Florencia (2005), *Fossilized Second Language Grammars – the Acquisition of Grammatical Gender*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gass, S. (1996), *Second Language Acquisition and Linguistic Theory: the Role of Language Transfer*. In: Bhatia, Tej. K. & Ritchie, William C. (Eds.), *The Handbook of Second Language Acquisition*, San Diego, California: Academic Press, pp. 317-345.

Gass, S. & Selinker, L. (2008), *Second Language Acquisition: an introductory course* (3ª edição), New York: Taylor and Francis Group, Routledge.

Gleitman, Lila R. & Newport, Elissa L. (1995), *The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language*. In: Gleitman, Lila R. & Liberman, Mark (Eds.), *Language. An Invitation to Cognitive Science*, vol. 1, 2nd ed., Cambridge, MA: MIT Press. Capítulo disponível a 29.09.2011, em http://jan.ucc.nau.edu/~jgc/554/gleitman_newport.pdf.

Gouveia, M. Carmen de Frias e (2004), *Considerações sobre a categoria gramatical de género: sua evolução do latim ao português arcaico*. In: *BIBLOS*, vol. II, pp. 443-475.

Gouveia, M. Carmen de Frias e (2005), *Para uma descrição do género em português*. In: *BIBLOS*, vol. III, pp. 201-246.

Hamers, J.F. & Blanc, M.H.A. (1989), *Bilinguality and Bilingualism*. Cambridge: Cambridge University Press.

Han, Zhao Hong (2004), *Fossilization in Adult Second Language Acquisition*, Clevedond, UK.: Multilingual Matters.

Hockett, Charles F. (1958), *A Course in modern linguistics*, New York: The Macmillan Company, p. 231.

Ibrahim, Muhammad H. (1973), *Grammatical Gender*, Paris/The Hague: Mouton.

Ioup, Georgette (2005), *Age in Second Language Development*. In: Hinkel, Eli (Ed.), *Handbook of Research in Second Language Acquisition Teaching and Learning*, New York: Taylor and Francis Group, Routledge, pp. 419-453.

Krashen, S. D. (1981), *Second Language Acquisition and Second Language Learning*, Oxford: Pergamon Press. Edição online de 2002, disponível a 18.07.2011, em http://www.sdkrashen.com/SL_Acquisition_and_Learning/index.html.

Leiria, Isabel (2004), *Português Língua Segunda e Língua Estrangeira: Investigação e Ensino*. In: *Idiomático. Revista Digital de Didáctica de PLNM*, nº 3, Centro Virtual Camões.

Leiria, Isabel (2006), *Léxico, Aquisição e Ensino do Português Europeu língua não materna*. Lisboa: Fundação Calouste Gulbenkian, Fundação para a Ciência e a Tecnologia, pp. 201-202.

Lenneberg, Eric H. (1967 [1975]), *Fundamentos biológicos del lenguaje*. Madrid: Alianza Editorial. Tradução do original inglês *The Foundations of Language*. [s. l.]: John Wiley and Sons.

Long, M. H. (2003), *Stabilization and Fossilization in Interlanguage Development*. In: Doughty, C. J. & Long., M. H. (Eds.), *The Handbook of Second Language Acquisition*, Malden: Blackwell, pp. 487-535.

Martins, Cristina (2008), *O Módulo “Funcionamento da Língua” dos Programas de Português à luz da nova “Terminologia Linguística para os Ensinos Básico e Secundário*. In: *Os programas de português dos ensinos básico e secundário, Actas das III Jornadas Científico-pedagógicas de português*, 17 e 18 de Novembro de 2005, Coimbra: Instituto de Língua e Literatura Portuguesas.

Martins, Cristina (2011, no prelo), *O Corpus de Produções Escritas de Aprendentes de PL2 (PEAPL2/CELGA). Caracterização e desenvolvimento de uma infra-estrutura de investigação*. In: Flores, Cristina (Ed.). *Português Língua Não Materna: Investigação e ensino*, Lisboa: Lidel.

Mattoso, Câmara Jr. Joaquim (1994), *Estrutura da língua portuguesa*, 22ª edição, Petrópolis: Vozes.

Mattoso, Câmara Jr. Joaquim (1966), *Considerações sobre o gênero em português*. In: *Revista Brasileira da Linguística Teórica e Aplicada*, volume I, n.2, São Paulo: Centro de Linguística Aplicada, p. 19.

Nichols, J. (1999), *Linguistic Diversity in Space and Time*, Chicago, IL: University of Chicago Press.

Paradis, M. (2004), *A Neurolinguistic Theory of Bilingualism*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

Payne, Thomas E. (2003), *Describing morphosyntax: a guide for field linguistics*, Cambridge: Cambridge University Press.

Peres, João Andrade e Mória, Telmo (1995), *Áreas Críticas da Língua Portuguesa*, Lisboa: Caminho.

Rio-Torto, Graça (1998), *Morfologia Derivacional. Teoria e Aplicação ao Português*, Porto: Porto Editora.

Rio-Torto, Graça (2002), *Flexão e derivação: simetrias e assimetrias*. In: *Revista Portuguesa de Filologia*, vol. XXIV, p. 253-289.

Selinker, L. (1992), *La Interlengua*. In: Licerias, J.M., *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*, Madrid: Visor, pp. 79-101.

Selinker, L. & Lakshmanan, U. (1993), *Language Transfer and Fossilization: The multiple effects principle*. In: Gass, S. & Selinker, L. (Eds.), *Language Transfer in Language Learning*, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 197-216.

Ullman, M. T. (2004), *Contributions of Memory Circuits to Language: the Declarative/Procedural Model*. In: *Cognition*, 92, pp. 231-270.

Vilela, Mário (1973), *Considerações sobre o género gramatical*. In: *Revista do Porto da Faculdade de Letras da Universidade do Porto*, vol. I, Porto: Série de Filologia, pp. 139-150.

Vilela, Mário (1999), *Gramática da língua portuguesa*, 2ª edição, Coimbra: Almedina.

Villalva, Alina (1994), *Estruturas morfológicas. Unidades e Hierarquias nas Palavras do Português*, Lisboa: Fundação Calouste Gulbenkian.

Villalva, Alina (2003), *Estrutura mórfica básica*. In: Mateus, Maria H. M., et. *alli*, *Gramática da língua portuguesa* (5ª edição revista e aumentada), Lisboa: Caminho, pp. 919-931.

Villalva, Alina (2008), *Morfologia do Português*, Lisboa: Universidade Aberta.

White, Lydia (1995), *Universal Grammar and Second Language Acquisition*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

White, Lydia (2003), *On the Nature of Interlanguage Representation: Universal Grammar*. In: Doughty, C. J. & Long., M. H. (Eds.), *The Handbook of Second Language Acquisition*, Malden: Blackwell, pp. 19-42.

Sítios consultados na Internet

Cátedra, Português Língua Segunda e Estrangeira, Instituto Camões, Universidade Eduardo Mondlane, Moçambique. Disponível a 29.09.2011, em <http://www.catedraportugues.uem.mz/>.

Centro de Linguística da Universidade de Lisboa (CLUL). Disponível a 29.09.2011, em <http://www.clul.ul.pt/pt/component/content/article/194-multifunctional-computational-lexicon-of-contemporary-portuguese>.

Corpus de Produções Escritas de Aprendentes de PL2 (*PEAPL2*). Disponível a 29.09.2011 em <http://www.uc.pt/fluc/rcpl2/>.

Dicionário Terminológico do Ministério da Educação (DT). Disponível em 29.09.2011, em <http://www.prof2000.pt/users/primavera/>.

Moçambicanismos, um Glossário com muitas imagens. Disponível em 29.09.2011, em <http://mocambicanismos.blogspot.com/2009/01/s.html>.