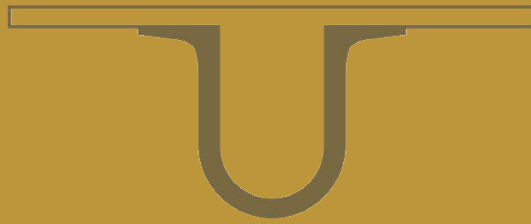




UNIVERSIDADE DE  
COIMBRA



Tânia Santos Ferreira

**AQUISIÇÃO/APRENDIZAGEM DO SISTEMA DE  
ATRIBUIÇÃO DE GÉNERO NOMINAL EM PLNM**

**VOLUME 1**

Tese no âmbito do Doutoramento em Linguística do Português orientada pelas  
Professoras Doutoras Cristina dos Santos Pereira Martins e Graça Maria de Oliveira e  
Silva Rio-Torto apresentada à Faculdade de Letras da Universidade de Coimbra.

Janeiro de 2019







Faculdade de Letras da Universidade de Coimbra

# **AQUISIÇÃO/APRENDIZAGEM DO SISTEMA DE ATRIBUIÇÃO DE GÉNERO NOMINAL EM PLN**

**VOLUME 1**

Tânia Santos Ferreira

Tese no âmbito do Doutoramento em Linguística do Português orientada pelas Professoras Doutoras Cristina dos Santos Pereira Martins e Graça Maria de Oliveira e Silva Rio-Torto e apresentada à Faculdade de Letras da Universidade de Coimbra.

Janeiro de 2019



UNIVERSIDADE DE  
**COIMBRA**





# AQUISIÇÃO/APRENDIZAGEM DO SISTEMA DE ATRIBUIÇÃO DE GÉNERO NOMINAL EM PLNM

VOLUME 1

## Ficha Técnica:

<b>Tipo de trabalho</b>	Tese de Doutoramento
<b>Título</b>	Aquisição/Aprendizagem do Sistema de Atribuição de Género Nominal em PLNM
<b>Autora</b>	Tânia Santos Ferreira
<b>Orientadoras</b>	Professora Doutora Cristina dos Santos Pereira Martins Professora Doutora Graça Maria de Oliveira e Silva Rio-Torto
<b>Júri</b>	Presidente: Professora Doutora Ana Paula dos Santos Duarte Arnout Vogais: 1. Professora Doutora Maria Antónia Ramos Coelho da Mota 2. Professora Doutora Maria de Fátima Henriques da Silva 3. Professor Doutor Jorge Alexandre Loureiro Pinto 4. Professora Doutora Isabel Maria de Almeida Santos 5. Professora Doutora Cristina dos Santos Pereira Martins
<b>Identificação do Curso</b>	3º Ciclo em Linguística do Português
<b>Área científica</b>	Linguística do Português
<b>Data da defesa</b>	05-06-2019
<b>Classificação</b>	Aprovado com Distinção e Louvor, por unanimidade







*À minha família*



# Índice Geral

---

Índice de Figuras .....	v
Índice de Gráficos.....	vi
Índice de Quadros .....	ix
Lista de abreviaturas.....	xv
Agradecimentos .....	xvii
Resumo .....	xix
Abstract .....	xxi
Introdução.....	1
Capítulo 1 .....	7
A categoria de género gramatical .....	7
1.1. Introdução .....	7
1.2. O género gramatical .....	9
1.2.1. Género e Classes nominais .....	12
1.2.1.1. A concordância sintática .....	16
1.2.2. Género e Classificadores.....	19
1.2.3. Considerações gerais sobre a origem e formação dos sistemas de género gramatical.....	23
1.2.4. Critérios de atribuição de género nominal (Corbett 1991) .....	27
1.3. O género nominal em português.....	35
1.3.1. A estrutura formal dos nomes em português .....	35
1.3.2. Género: uma categoria de natureza flexional? .....	38
1.3.3. Critérios de atribuição de género nominal do português .....	43
1.4. O género gramatical nas LM dos aprendentes de PLNM participantes no estudo .....	49

1.4.1.	O sistema de atribuição de género nominal do espanhol .....	51
1.4.1.1.	Sobre a estrutura formal dos nomes em espanhol.....	51
1.4.1.2.	Sobre a natureza (não) flexional do género em espanhol.....	54
1.4.1.3.	Critérios de atribuição de género aos nomes em espanhol .....	56
1.4.2.	O sistema de atribuição de género nominal do italiano .....	59
1.4.2.1.	A estrutura formal dos nomes em italiano.....	59
1.4.2.2.	Critérios de atribuição de género nominal do italiano .....	64
1.4.3.	O sistema de atribuição de género nominal do alemão.....	69
1.4.3.1.	Particularidades do sistema de concordância nominal do alemão (em género, em número e em caso) .....	69
1.4.3.2.	Critérios de atribuição de género nominal do alemão.....	70
1.4.4.	O sistema de classificação nominal do inglês .....	75
1.4.4.1.	Especificidades relativas ao sistema de classificação nominal do inglês... ..	75
1.4.4.2.	O sistema pronominal do inglês.....	79
1.4.5.	O sistema de classificação nominal do chinês (mandarim).....	83
1.4.5.1.	Características da estrutura formal do chinês .....	83
1.4.5.2.	O sistema de classificadores do chinês: características gerais .....	86
1.4.6.	Síntese.....	89
1.5.	O género gramatical no contexto da aquisição linguística: LM vs. LNM.....	91
1.5.1.	Aquisição de género gramatical por falantes nativos.....	91
1.5.2.	Aquisição do género gramatical por falantes não nativos.....	95
1.5.3.	Síntese.....	101
<b>Capítulo 2.....</b>	<b>103</b>	
<b>Para uma avaliação do papel dos conhecimentos linguísticos prévios na aquisição/aprendizagem de uma LNM .....</b>	<b>103</b>	
2.1.	Introdução .....	103
2.2.	Aquisição da LM vs. aquisição/aprendizagem de uma LNM.....	105
2.2.1.	Hipótese do período ‘crítico’ (ou ‘sensível’) para a aquisição linguística .....	107

2.2.2.	Interlíngua .....	113
2.2.2.1.	O papel da Gramática Universal .....	118
2.2.2.2.	O papel do <i>input</i> linguístico .....	121
2.2.2.3.	O papel diferenciado dos subsistemas de memória: o modelo declarativo/procedimental .....	123
2.3.	O papel dos conhecimentos linguísticos prévios para a construção das interlínguas	127
2.3.1.	Para uma definição de transferência linguística .....	127
2.3.2.	Manifestações associadas à transferência linguística .....	135
2.3.2.1.	Transferência positiva .....	135
2.3.2.2.	Transferência negativa (ou interferência) .....	137
2.3.3.	Condições promotoras de transferibilidade do género gramatical .....	140
2.3.3.1.	Proximidade tipológica .....	141
2.3.3.2.	Configuração da gramática da interlíngua .....	142
2.3.3.2.1.	Full Transfer / Full Access Hypothesis .....	142
2.3.3.2.2.	Minimal Trees Hypothesis .....	145
2.3.3.2.3.	Failed Functional Features Hypothesis .....	146
2.3.3.3.	Nível de proficiência e de competência linguística .....	148
2.4.	Síntese .....	149
	<b>Capítulo 3 .....</b>	<b>151</b>
	<b>Metodologia .....</b>	<b>151</b>
3.1.	Introdução .....	151
3.1.1.	Perguntas de investigação e variáveis em estudo .....	151
3.2.	Constituição da base empírica do estudo .....	157
3.2.1.	Descrição das bases de dados consultadas .....	157
3.2.1.1.	<i>Corpus</i> PEAPL2 e Projeto Recolha de Dados de Aprendizagem de Português Língua Estrangeira (PLE) .....	157
3.2.1.2.	<i>Corpus</i> de Aquisição de L2 (CAL2) .....	160
3.2.2.	Seleção do <i>corpus</i> da investigação .....	162

3.3.	Perfil dos informantes.....	169
3.3.1.	Língua Materna.....	170
3.3.2.	Outras línguas não maternas .....	170
3.3.3.	Nível QECRL em português .....	174
3.3.4.	Idade .....	175
3.3.5.	Período de exposição ao português.....	176
3.3.6.	Experiência de imersão .....	178
3.4.	Procedimentos de tratamento dos dados.....	181
3.4.1.	Contabilização de itens nominais .....	181
3.4.2.	Contabilização dos desvios.....	187
3.4.3.	Tipologia de desvios .....	191
	<b>Capítulo 4.....</b>	<b>195</b>
	<b>Resultados e Discussão .....</b>	<b>195</b>
4.1.	Introdução .....	195
4.2.	Resultados.....	196
4.2.1.	Resultados globais.....	196
4.2.2.	Resultados por variáveis de análise.....	197
4.2.2.1.	Distribuição dos resultados por LM.....	197
4.2.2.2.	Distribuição dos resultados por nível QECRL.....	198
4.2.2.3.	Distribuição dos resultados por LM e nível QECRL.....	199
4.2.3.	Resultados apurados em função da tipologia de desvios .....	207
4.2.3.1.	Categorias e Subcategorias de desvios de AGN .....	213
4.2.3.1.1.	Categoria 1 [SEM +, FORM +] ( <i>o menino, a menina</i> ).....	217
4.2.3.1.2.	Categoria 2 [SEM +, FORM -] ( <i>o homem, a mulher</i> ) .....	223
4.2.3.1.3.	Categoria 3 [SEM -, FORM +] ( <i>o carro, a casa, o indivíduo, a pessoa</i> ).....	229
4.2.3.1.4.	Categoria 4 [SEM -, FORM -] ( <i>o pente, a ponte, o coração, a viagem</i> ).....	255
	<b>Considerações Finais.....</b>	<b>287</b>
	<b>Bibliografia .....</b>	<b>299</b>

## Índice de Figuras

---

<b>Figura 1.1</b> – Distribuição do número de valores de género em várias línguas do mundo (fonte: <i>World Atlas of Language Structures Online</i> [WALS], 30A, <a href="http://wals.info/chapter/30">http://wals.info/chapter/30</a> ).....	10
<b>Figura 1.2</b> – Os sistemas de classificação nominal organizados num <i>continuum</i> lexical-gramatical (figura baseada na proposta de Grinevald (2000:61)).....	13
<b>Figura 1.3</b> – Tipos de classificadores (de acordo com proposta de Grinevald (2000)).....	20
<b>Figura 1.4</b> – Critérios de atribuição de valores de género nominal (baseado na proposta de Corbett (1991)) .....	27
<b>Figura 1.5</b> – Aplicação de critérios (semânticos e formais) de atribuição de género nominal do português.....	47
<b>Figura 2.1</b> – Grau de acessibilidade da GU no contexto de aquisição/aprendizagem de uma LNM (figura baseada na proposta de Cook (1994:33)) .....	119
<b>Figura 2.2</b> – Subsistemas de memória (esquema baseado em Baddeley (1999:1)).....	124
<b>Figura 2.3</b> – Distinção dos efeitos positivos e negativos da transferência linguística .....	135
<b>Figura 3.1</b> – Convenções adotadas durante o processo de transcrição dos documentos manuscritos que compõem o <i>Projeto de Recolha de Dados de Aprendizagem de PLE</i> e o <i>Corpus PEAPL2</i> (retirado de Leiria (2006:181)) .....	160
<b>Figura 3.2</b> – Códigos utilizados no cabeçalho dos textos disponibilizados no <i>CAL2</i> .....	161

## Índice de Gráficos

---

<b>Gráfico 3.1</b> – Percentagem de textos consultados nas três bases de dados com desvios de AGN e de CNG em relação à percentagem de textos sem desvios .....	167
<b>Gráfico 3.2</b> – Proporção (numa escala de 0 a 100) de textos com desvios de AGN e de CNG no SN, face ao número de textos sem desvios, consultados nas três bases de dados .....	167
<b>Gráfico 3.3</b> – Distribuição do número de informantes por número de outra(s) LNM conhecida(s) (1, 2 ou mais do que 2).....	171
<b>Gráfico 3.4</b> – Distribuição (em valores percentuais) do conjunto total de informantes por nível QECRL da turma frequentada.....	174
<b>Gráfico 3.5</b> – Distribuição do número de informantes por nível QECRL em português da turma frequentada.....	175
<b>Gráfico 3.6</b> – Distribuição do número de informantes por faixa etária.....	175
<b>Gráfico 3.7</b> – Distribuição do número de informantes por período de exposição à LA .....	176
<b>Gráfico 3.8</b> – Distribuição dos informantes por LM, nível QECRL e período de exposição à LA.....	177
<b>Gráfico 3.9</b> – Distribuição, em valores percentuais, dos informantes em contexto de imersão linguística (SIM) face aos informantes que não se encontram em imersão linguística (NÃO) .....	178
<b>Gráfico 3.10</b> – Percentagem de ocorrências de nomes em SN não reduzidos por oposição à percentagem de ocorrências de nomes em SN reduzidos .....	185
<b>Gráfico 4.1</b> – Distribuição das ocorrências nominais com desvios de AGN e de CNG, por oposição às ocorrências nominais sem desvios .....	196
<b>Gráfico 4.2</b> – Distribuição das percentagens relativas de desvios de AGN e de CNG por LM .....	197
<b>Gráfico 4.3</b> – Distribuição das percentagens relativas de desvios de AGN e de CNG por nível QECRL .....	199



<b>Gráfico 4.4</b> – Percentagens relativas de desvios de AGN e de CNG por segmentos da amostra (LM e nível QERCL).....	200
<b>Gráfico 4.5</b> – Distribuição, em valores percentuais (numa escala proporcional de 0 a 100), dos desvios em função do valor de género do item nominal afetado .....	204
<b>Gráfico 4.6</b> – Percentagens de desvios, distribuídas por categorias .....	208
<b>Gráfico 4.7</b> – Percentagens de desvios, distribuídas por categorias .....	210
<b>Gráfico 4.8</b> – Percentagem de desvios por categoria em função da LM e do nível QECRL da turma frequentada pelos informantes .....	214
<b>Gráfico 4.9</b> – Distribuição das percentagens de desvios por constituinte afetado em função da LM e do nível QECRL da turma frequentada pelos informantes .....	216
<b>Gráfico 4.10</b> – Percentagem relativa de desvios da categoria 1 por segmento da amostra (LM e nível QECRL dos informantes).....	218
<b>Gráfico 4.11</b> – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos nos níveis A1-A2.....	220
<b>Gráfico 4.12</b> – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos nos níveis B1-B2 .....	220
<b>Gráfico 4.13</b> – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos no nível C1.....	220
<b>Gráfico 4.14</b> – Percentagem relativa de desvios da categoria 2.....	226
<b>Gráfico 4.15</b> – Percentagem de desvios da categoria 2, distinguidos por classes de palavras afetadas, nos níveis A1-A2.....	227
<b>Gráfico 4.16</b> – Percentagem de desvios da categoria 2, distinguidos por por classes de palavras afetadas, nos níveis B1-B2 .....	227
<b>Gráfico 4.17</b> – Percentagem relativa de desvios da categoria 3 por segmento da amostra (LM e nível QECRL).....	231
<b>Gráfico 4.18</b> – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos nos níveis A1-A2.....	234
<b>Gráfico 4.19</b> – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos nos níveis B1-B2 .....	234

<b>Gráfico 4.20</b> – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos no nível C1.....	234
<b>Gráfico 4.21</b> – Percentagem de desvios da categoria 3 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM espanhola e italiana.....	248
<b>Gráfico 4.22</b> – Percentagem de desvios da categoria 3 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM alemã...	253
<b>Gráfico 4.23</b> – Percentagem relativa de desvios da categoria 4.....	256
<b>Gráfico 4.24</b> – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, nos níveis A1-A2 .....	258
<b>Gráfico 4.25</b> – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, nos níveis B1-B2.....	258
<b>Gráfico 4.26</b> – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, no nível C1.....	258
<b>Gráfico 4.27</b> – Percentagem de desvios da categoria 4, calculada em função do constituinte temático do item lexical afetado (-o, -a, -e, e aтемáticos (Ø)).....	262
<b>Gráfico 4.28</b> – Percentagem de desvios da categoria 4 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM espanhola e italiana.....	274
<b>Gráfico 4.29</b> – Percentagem de desvios da categoria 4 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM alemã...	284

## Índice de Quadros

---

<b>Quadro 1.1</b> – Síntese das propriedades relativas aos sistemas de classificação nominal baseados em género gramatical e em classificadores .....	23
<b>Quadro 1.2</b> – Distribuição dos nomes em português em função de vogal temática / índice temático e valor de género gramatical.....	37
<b>Quadro 1.3</b> – Frequência e respetivo valor percentual dos lemas recenseados na CORLEX em função do género gramatical e da respetiva classe temática dos nomes simples e derivados (cf. Ferreira 2011).....	45
<b>Quadro 1.4</b> – Relação entre sufixos derivacionais e respetivo valor de género .....	46
<b>Quadro 1.5</b> – Distribuição dos itens nominais em espanhol em função da classe, índice temático e respetivos valores de género (de acordo com a proposta de Harris (1992:65)).....	54
<b>Quadro 1.6</b> – Distribuição dos nomes do italiano, formas variáveis e invariáveis, em função do índice temático e respetivo valor de género .....	63
<b>Quadro 1.7</b> – Paradigmas de concordância nominal do alemão (concordância em género, número e caso) .....	69
<b>Quadro 1.8</b> – Nomes antecedidos de caracteres que permitem a especificação do sexo do referente [+humano] (exemplos recolhidos de Etnner (2002:37)) .....	85
<b>Quadro 1.9</b> – Síntese das particularidades associadas à aquisição do sistema de atribuição de género nominal do português por falantes nativos e por falantes não nativos.....	101
<b>Quadro 3.1</b> – Síntese das características dos sistemas de classificação nominal das LM dos informantes, autores dos textos escritos que compõem o <i>corpus</i> da presente investigação (cf. Capítulo 1, Secção 1.4) .....	155
<b>Quadro 3.2</b> – Acervo total de produções escritas analisadas nas três bases de dados, distribuídas por LM e nível QECRL da turma frequentada pelos informantes .....	163

<b>Quadro 3.3</b> – Acervo textual com desvios de AGN e de CNG que compõem o <i>corpus</i> da investigação.....	166
<b>Quadro 3.4</b> – Distribuição do número de informantes por LM.....	170
<b>Quadro 3.5</b> – Distribuição do número de nomes em SN, de desvios de AGN e de CNG e da respectiva percentagem relativa por segmentos de amostra de informantes, em imersão linguística e sem imersão linguística .....	179
<b>Quadro 3.6</b> – Distribuição do total e média de palavras e de ocorrências de nomes por texto produzido por LM e nível QECRL dos informantes.....	182
<b>Quadro 3.7</b> – Distribuição no <i>corpus</i> de ocorrências de nomes em SN não reduzidos por oposição ao número de ocorrências de SN reduzidos .....	184
<b>Quadro 3.8</b> – Distribuição, por segmentos da amostra (LM e nível QECRL), dos números absolutos de ocorrências de nomes com desvios e do número de desvios, apurados em cada texto .....	188
<b>Quadro 3.9</b> – Proposta de caracterização do sistema de género nominal em português, atendendo aos indícios de atribuição de género nominal .....	192
<b>Quadro 3.10</b> – Tipologia de desvios com exemplos extraídos do <i>corpus</i> .....	193
<b>Quadro 4.1</b> – Distribuição do número total de ocorrências de nomes e de desvios de AGN e de CNG por LM.....	197
<b>Quadro 4.2</b> – Distribuição do número total de ocorrências de nomes e de desvios de AGN e de CNG por nível QECRL .....	198
<b>Quadro 4.3</b> – Distribuição do número total de desvios de AGN e de CNG, e respetivas percentagens relativas, por segmentos da amostra (LM e nível QECRL).....	200
<b>Quadro 4.4</b> – Número absoluto de desvios, e dos valores da média, por categorias de desvio .....	208
<b>Quadro 4.5</b> – Número absoluto de desvios, e dos valores da média, por subcategorias de desvio.....	209
<b>Quadro 4.6</b> – Distribuição do número absoluto e dos valores da média de ocorrências de desvios por categoria apurados por texto produzido em cada segmento da amostra (LM e nível QECRL).....	213

<b>Quadro 4.7</b> – Distribuição do número absoluto e dos valores da média de ocorrências de desvios por constituinte afetado, apurados por texto escrito em cada segmento da amostra (LM e nível QECRL).....	215
<b>Quadro 4.8</b> – Distribuição do número absoluto e dos valores da média de ocorrências de nomes da categoria 1 produzidos no <i>corpus</i> e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QECRL).....	218
<b>Quadro 4.9</b> – Distribuição do número de ocorrências de nomes com desvio da categoria 1 no <i>corpus</i> e respetivo número de desvios por segmentos da amostra (LM e Nível QECRL) .....	222
<b>Quadro 4.10</b> – Distribuição do número absoluto e dos valores de média de ocorrências de nomes da categoria 2 produzidos no <i>corpus</i> e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QECRL).....	225
<b>Quadro 4.11</b> – Distribuição do número de ocorrência de nomes com desvio da categoria 2 no <i>corpus</i> e respetivo número de desvios por segmentos da amostra (LM e nível QECRL) .....	228
<b>Quadro 4.12</b> – Distribuição do número absoluto e dos valores de média de ocorrências de nomes da categoria 3 produzidos no <i>corpus</i> e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QECRL).....	230
<b>Quadro 4.13</b> – Distribuição do número de nomes afetados, do respetivo número de ocorrências no <i>corpus</i> e do número de desvios da categoria 3.....	236
<b>Quadro 4.14</b> – Distribuição de nomes afetados da categoria 3, com a indicação do número absoluto de ocorrências no <i>corpus</i> e do respetivo número absoluto de desvios.....	238
<b>Quadro 4.15</b> – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 3, detetados nas produções escritas por informantes de LM espanhola e italiana, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical.....	247
<b>Quadro 4.16</b> – Distribuição do número de desvios da categoria 3 produzidos nos segmentos da amostra de informantes de LM espanhola e italiana potencialmente atribuíveis à transferência linguística .....	248
<b>Quadro 4.17</b> – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 3, detetados nas produções escritas por informantes de LM alemã, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical.....	252

<b>Quadro 4.18</b> – Distribuição do número de desvios da categoria 3 produzidos nos segmentos da amostra de informantes de LM alemã potencialmente atribuíveis à transferência linguística .....	253
<b>Quadro 4.19</b> – Distribuição do número absoluto e dos valores de média de nomes da categoria 4 produzidos no corpus, e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QECRL).....	256
<b>Quadro 4.20</b> – Distribuição do número de nomes afetados, do respetivo número de ocorrências no <i>corpus</i> e do número de desvios na categoria 4 .....	260
<b>Quadro 4.21</b> – Distribuição dos desvios da categoria 4 produzidos por segmento da amostra (LM e nível QECRL) em função do constituinte temático do item lexical afetado (-o, -a, -e, atemáticos (Ø)).....	262
<b>Quadro 4.22</b> – Distribuição de nomes afetados da categoria 4 de tema em -a, por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no <i>corpus</i> e de desvios por item .....	264
<b>Quadro 4.23</b> – Distribuição de nomes afetados da categoria 4 de tema em -e, por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no <i>corpus</i> e de desvios por item .....	266
<b>Quadro 4.24</b> – Distribuição de nomes afetados da categoria 4 atemáticos (Ø), por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no <i>corpus</i> e de desvios por item .....	268
<b>Quadro 4.25</b> – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 4, detetados nas produções escritas por informantes de LM espanhola e italiana, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical.....	273
<b>Quadro 4.26</b> – Distribuição do número de desvios da categoria 4 produzidos nos segmentos da amostra de informantes de LM espanhola e italiana potencialmente atribuíveis à transferência da LM.....	274
<b>Quadro 4.27</b> – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 4, detetados nas produções escritas por informantes de LM alemã, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical .....	277

**Quadro 4.28** – Distribuição do número de desvios da categoria 4 produzidos nos segmentos da amostra de informantes de LM alemã potencialmente atribuíveis à transferência da LM.....  
.....284





# Lista de abreviaturas

---

- AGN** – Atribuição de valores de Género aos Nomes
- ALNM** – Aquisição de Línguas Não Maternas
- CNG** – Concordância Nominal em Género
- GU** – Gramática Universal
- FFFH** – *Failed Functional Features Hypothesis*
- FTFA** – *Full Transfer Full Access Hypothesis*
- LA** – Língua-alvo
- LAD** – *Language Acquisition Device*
- LM** – Língua Materna
- LNM** – Língua Não Materna
- MCP** – Memória de Curto Prazo
- MLP** – Memória de Longo Prazo
- MTH** – *Minimal Trees Hypothesis*
- N** – Nomes
- PLNM** – Português como Língua Não Materna
- QECRL** – *Quadro Europeu Comum de Referência para as Línguas* (Conselho da Europa 2001)
- SN** – Sintagma Nominal



# Agradecimentos

---

“Aqueles que passam por nós, não vão sós, não nos deixam sós.  
Deixam um pouco de si, levam um pouco de nós.”

*Antoine de Saint-Exupéry*

Concluir uma Tese de doutoramento implica uma longa e difícil jornada. Sabemos que este trabalho exige de nós muita dedicação e empenho, um trabalho individual e solitário que nos desafia constantemente. Sem dúvida que não teria sido possível levar esta investigação a bom porto se não tivesse o apoio de uma rede de pessoas a quem quero deixar aqui os meus sinceros e profundos agradecimentos.

Às minhas orientadores, as Professoras Doutoras Cristina dos Santos Pereira Martins e Graça Maria de Oliveira e Silva Rio-Torto, da Universidade de Coimbra. Agradeço a vós o apoio incondicional a este projeto, a dedicação e a orientação que me concederam ao longo destes anos. Agradeço profundamente toda a confiança depositada em mim e como me foram incentivando durante toda a jornada, para que não desistisse e continuasse a acreditar em mim. Desde o primeiro momento que o vosso apoio foi crucial para o desenvolvimento deste trabalho. Por isso, estou eternamente grata pelos conselhos dados que me permitiram progredir não só como investigadora, mas, principalmente, como pessoa.

Às Professoras do Centro de Linguística Geral e Aplicada (CELGA/ILTEC) da Universidade de Coimbra, em especial às Professoras Doutoras Isabel Santos e Conceição Carapinha pelo apoio e conselhos dados sempre que precisei. Nunca me vou esquecer e guardarei sempre no meu coração a vossa ajuda.

A toda a minha família, pelo apoio incondicional que me concederam durante estes anos. Ao Pedro Santos, pois sem o teu carinho não teria conseguido ultrapassar os momentos mais difíceis. Obrigada por estares sempre ao meu lado.

Às minhas colegas da Faculdade de Letras da Universidade de Coimbra, em especial à Ana Guerra, Sandra Chapouto e Sara Sousa, pelo vosso incentivo ao longo destes anos. Estou grata pela oportunidade de trabalhar com todas vós e de poder contar com as vossas

palavras sempre que precisar. Obrigada por me fazerem sentir em casa e por me ajudarem nos momentos mais difíceis.

À minha amiga Geraldine Garcia, pela tua profunda amizade. Mesmo que estejas a meio mundo de distância, terás sempre um lugar especial no meu coração. Obrigada por me inspirares com a tua força, humildade e bom coração. Muito obrigada pelas tuas palavras motivadoras e pela tua preciosa ajuda e, claro, por toda a paciência que tens e sempre tiveste comigo.

Aos meus eternos amigos da China, Zhang Yufeng e Yiran Liao, pelo espírito de amizade e por me acolherem sempre tão bem. Estou-vos eternamente grata.

Aos meus alunos, por me inspirarem com o vosso empenho e dedicação.

E a todos aqueles que, de forma direta ou indireta, contribuíram para a conclusão desta tese de doutoramento. A todos vós, **MUITO OBRIGADA!**

# Resumo

---

O objetivo primordial da presente tese é averiguar o papel (ou possíveis papéis) que o conhecimento linguístico prévio da língua materna (LM) de um aprendente tardio pode ter no desenvolvimento do processo de assimilação dos valores de género nominal em português como língua não materna (PLNM). Mais especificamente, procura-se avaliar o grau de transferibilidade dos valores de género gramatical no contexto de aquisição/aprendizagem tardia do PLNM, bem como o grau de sensibilidade dos diferentes aprendentes relativamente aos indícios semânticos e formais de associação dos valores de género aos nomes em português, ao longo das diferentes fases do seu desenvolvimento (inter)linguístico.

Tendo em conta os objetivos estipulados, procedeu-se à análise de um conjunto de desvios de atribuição de valores de género e de concordância nominal em género detetados no sintagma nominal (SN) assinalados em produções escritas por aprendentes tardios, falantes nativos de espanhol, italiano, alemão, inglês e chinês, a frequentar turmas de diferentes níveis de aprendizagem do português, e que se encontram disponíveis em diferentes *corpora* de referência. A seleção destes idiomas prende-se, precisamente, com o grau de semelhança e de distância dos mesmos relativamente às características dos sistemas de classificação nominal que exibem face ao sistema de atribuição de género nominal do português, a língua-alvo (LA) de aprendizagem.

Assim sendo, partindo da definição do conceito de género gramatical (Corbett 1991; Aikhenvald 2000), categoria pertinente para a classificação dos nomes, procedeu-se à descrição do sistema de atribuição de género nominal do português (Câmara Jr. 1994; Villalva 2003; Choupina *et al.* 2016; Mota 2016b), bem como à descrição dos demais sistemas de classificação nominal das línguas selecionadas (Ibrahim 1973; Li & Thompson 1981; Zubin & Köpcke 1986; Ambadiang 1994; Dardano & Trifone 1997).

Posteriormente, consideraram-se os contributos teóricos relativos à aquisição/aprendizagem tardia de uma LNM, com especial enfoque para as particularidades relativas à assimilação tardia das propriedades morfossintáticas de um idioma não nativo (Hawkins & Chan 1997; Tsimpli 2003). Para além disso, foram igualmente tidas em conta

as questões teóricas subjacentes ao papel que, no contexto da assimilação das propriedades gramaticais de uma LNM, é atribuído ao conhecimento linguístico prévio proveniente da LM, bem como as manifestações tipicamente associadas ao fenómeno da transferência linguística (White *et al.* 2004; Franceschina 2005; Sabourin *et al.* 2006).

Atendendo ao enquadramento teórico, procedeu-se à análise dos diferentes tipos de desvios assinalados nos vários segmentos da amostra selecionada. Assim sendo, as correlações estabelecidas entre os tipos de desvios produzidos, as propriedades de atribuição de valores de género nominal do português, a LM e o nível de aprendizagem dos aprendentes tardios de PLNМ permitem concluir que: (i) o género gramatical constitui uma área crítica no contexto da aprendizagem tardia, já que em todos os segmentos da amostra, por LM e nível de aprendizagem, se registam desvios; (ii) o facto de a LM do aprendente possuir um sistema de atribuição de género nominal com propriedades similares às do português contribui, favoravelmente, para a assimilação dos valores de género nominal do português, uma vez que o desempenho registado nos aprendentes, cuja LM possui género, é satisfatoriamente melhor do que aquele que é registado nos segmentos da amostra de informantes que são falantes nativos de idiomas sem género; (iii) em todos os segmentos da amostra, ao longo do processo de assimilação dos valores de género nominal, o papel dos indícios semânticos é mais significativo do que o dos indícios formais; e (iv) o número considerável de desvios assinalado em cada segmento da amostra é igualmente revelador de que, durante o processo de aquisição/aprendizagem tardia dos itens lexicais, o género não é uma das primeiras propriedades a ser assimilada, a não ser que estes tenham um correferencial semântico.

**Palavras-chave:** aquisição/aprendizagem tardia do português como língua não materna, sistemas de classificação nominal, valores de género gramatical, transferência linguística, interlíngua

# Abstract

---

The main objective of this thesis is to investigate the role (or possible roles) of a late learner's prior linguistic knowledge (NL) in the development of the assimilation process of nominal gender values in Portuguese as a non-native language (PNNL). More specifically, it seeks to assess the degree of transferability of grammatical gender values in the context of late acquisition / learning of PNNL, as well as the degree of sensitivity of the different learners to the semantic and formal evidence of gender values association to Portuguese names throughout the different phases of their (inter)linguistic development.

Having these targets into account, an analysis was conducted on a set of gender assignment and gender agreement deviations detected in the nominal phrase (NP) marked in productions written by late learners, native speakers of Spanish, Italian, German, English and Chinese who attend classes of different levels of Portuguese learning, deviations which are available in different reference *corpora*. The selection of these languages is related to the degree of similarity and distance between them in relation to the characteristics of the nominal classification systems that they display in relation to the nominal gender assignment system of the target learning language (TL).

Therefore, starting from the definition of the grammatical gender concept (Corbett 1991; Aikhenvald 2000), which is an relevant category for the classification of names, the system of nominal gender attribution in Portuguese was described (Câmara Jr. 1994, Villalva 2003, Choupina (1991), as well as the other nominal classification systems of the selected languages (Ibrahim 1973; Li & Thompson 1981; Zubin & Köpcke 1986; Ambadiang 1994; Dardano & Trifone 1997).

Later, the theoretical contributions regarding late acquisition / learning of a NNL were considered, with special focus on the peculiarities related to late assimilation of the morphosyntactic properties of a non-native language (Hawkins & Chan 1997; Tsimpli 2003). In addition, the theoretical questions underlying the role which, in the context of the assimilation of the grammatical properties of a NNL, is attributed to the prior linguistic knowledge derived from a NL were also considered, as well as the manifestations typically

associated with the linguistic transfer phenomenon (White et al., 2004; Franceschina 2005; Sabourin et al., 2006).

Taking into account the theoretical framework, the different types of deviations noted in the various segments of the selected empirical sample were analyzed. Thus, the correlations established between the types of deviations produced, the nominal gender assignment properties of the Portuguese language, the NL and the learning level of the late learners of PNNL allow to conclude that: (i) grammatical gender constitutes an area critical in the context of late learning, since there are deviations in all segments of the sample, by NL and learning level; (ii) the fact that the NL of the learner having a nominal gender assignment system with similar properties to that of Portuguese contributes favorably to the assimilation of the nominal gender values of Portuguese, since the performance of learners whose NL has gender differentiation is well better than that which is recorded in segments of the sample of informants who are native speakers of languages without gender; (iii) in all segments of the sample, throughout the process of assimilation of nominal gender values, the role of semantic cues is more significant than formal clues; and (iv) the considerable number of deviations noted in each segment of the sample is also revealing that, during the late acquisition / learning process of lexical items, gender is not one of the first properties to be assimilated, unless these have a semantic co-referential.

**Keywords:** late acquisition / learning of Portuguese as non-native language, nominal classification systems, grammatical gender values, linguistic transfer, interlanguage



# Introdução

---

A presente tese constitui um trabalho de investigação sobre o processo de aquisição/aprendizagem da atribuição de género nominal por aprendentes tardios de português como língua não materna (PLNM). Mais especificamente, neste estudo pretende-se averiguar o papel – ou possíveis papéis – do idioma nativo do aprendente tardio ao longo da assimilação dos valores de género nominal em PLNM e, com base nos dados empíricos, avaliar o grau de transferibilidade desta categoria gramatical.

Em função dos objetivos estipulados, procede-se a uma análise cuidada dos comportamentos desviantes relativos à estrutura gramatical em estudo, recolhidos de textos disponibilizados em diferentes *corpora* de referência, e redigidos por aprendentes tardios que se encontravam, no momento da recolha, a frequentar turmas de diferentes níveis de competência e de proficiência linguísticas em português, estipulados de acordo com o *Quadro Europeu Comum de Referência para as Línguas* (QECRL) (Conselho da Europa 2001), a língua-alvo (LA), e que são falantes nativos de espanhol, italiano, alemão, inglês e chinês. A seleção destes cinco idiomas está relacionada com o grau de distância e de similaridade que evidenciam relativamente ao português no que diz respeito às características dos sistemas de classificação nominal que exibem.

O género detém, indubitavelmente, um estatuto especial entre as demais categorias gramaticais (Corbett 1991; Aikhenvald 2000). Trata-se de uma categoria pertinente para a classificação dos nomes de várias línguas, permitindo a sua organização em diferentes valores, sendo que palavras de outras classes, que estabelecem com o item nominal relações de concordância sintática, são igualmente marcadas quanto ao valor de género. No entanto, nem todas as línguas exibem sistemas de classificação nominal baseados em género e, mesmo entre as que exibem esta categoria gramatical, observa-se uma multiplicidade de critérios subjacentes à atribuição dos valores de género, quer sob o ponto de vista semântico quer sob o ponto de vista formal (Corbett 1991, 2013c).

Dados disponíveis sobre a aquisição/aprendizagem de uma LNM revelam que, quanto mais tardia for a exposição à LA, maiores dificuldades se registam na assimilação

do sistema de atribuição de género aos nomes dessa língua e que os desvios de atribuição de valores de género aos nomes (AGN) e de concordância nominal em género (CNG) nunca chegam a ser erradicados por completo, mesmo nas produções de aprendentes de níveis QECRL mais avançados (cf. Oliphant 1997; Dewaele & Veronique 2000, 2001; Franceschina 2003, 2005). A investigação produzida no âmbito da Aquisição de Línguas Não Maternas (ALNM) sobre o processo de aquisição da categoria de género gramatical bem como do mecanismo de concordância decorrente dela por falantes aprendentes tardios em diferentes fases do seu desenvolvimento interlinguístico tem vindo a corroborar que quanto mais tardia é a exposição do aprendente à LA, maiores dificuldades se observam na assimilação dos valores de género nominal, sendo que os desvios são persistentes mesmo nas produções de aprendentes com níveis QECRL superiores (Godinho 2010; Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Martins 2015; Pinto 2015).

Assim, há indícios de que a assimilação desta categoria gramatical poderá estar condicionada por uma ‘idade crítica’, estando, por isso, também dependente de um conjunto de constrangimentos tipicamente associados à assimilação tardia das estruturas gramaticais. Com efeito, inúmeros estudos avançam com hipóteses teóricas várias para compreender a correlação entre a idade de início exposição do falante aprendente à LA e o processo de aquisição/aprendizagem do domínio morfossintático (Hawkins & Chan 1997; Franceschina 2003, 2005; Tsimpli 2003). Para além do peso significativo que, na literatura, tem sido dado aos condicionamentos característicos da aquisição/aprendizagem tardia de um idioma, regista-se igualmente na bibliografia especializada um interesse relativamente ao papel que a configuração do conhecimento linguístico prévio, sobretudo proveniente da LM, acarreta para o processamento das propriedades do sistema de AGN da LA, assim como para o processamento da concordância sintática (Bruhn de Garavito & White 2002; White *et al.* 2004; Franceschina 2005; Sabourin *et al.* 2006).

Tomando como referência os contributos de diferentes estudos sobre a aquisição/aprendizagem tardia do género gramatical, no presente trabalho procura-se avaliar o papel, ou possíveis papéis, do sistema linguístico nativo do falante aprendente ao longo das diferentes fases de desenvolvimento da assimilação da categoria de género nominal em PLNM. Para além disso, no tratamento empírico dos dados recolhidos para esta investigação, tem-se igualmente em conta o papel que as propriedades relativas ao sistema de associação dos valores de género aos nomes em português poderá desempenhar para o desenvolvimento interlinguístico dos aprendentes tardios.

Procura-se ainda apurar se há padrões de comportamento de atribuição de valores de género e de concordância nominal em género que se possam correlacionar, quer com a configuração dos conhecimentos linguísticos prévios do falante aprendente, quer com as propriedades do sistema de classificação nominal baseado em valores de género do português, a LA de aprendizagem.

Refira-se, em seguida, a estrutura do presente trabalho. O **Capítulo 1** centra-se na categoria de género gramatical. Assim, na Secção 1.2., procede-se à descrição do conceito de género gramatical com o intuito de compreender o lugar que, decorrente dos diferentes contributos teóricos, esta categoria ocupa na estrutura gramatical dos idiomas, bem como a relação entre os sistemas de classificação nominal baseados em género e outros sistemas existentes nas línguas do mundo (Ibrahim 1973; Corbett 1991, 2005, 2006b, 2013c; Ritter 1993; Chomsky 1995 [1999]; Comrie 1999; Aikhenvald 2000, 2004; Carvalho 2000; Grinevald 2000; Kibort & Corbett 2008). Partindo da delimitação concetual da noção de género gramatical, analisam-se, na Secção 1.3., as características do sistema de atribuição de género nominal do português, a LA de aprendizagem (Câmara Jr. 1994; Villalva 2003, 2008; Costa *et al.* 2015; Choupina *et al.* 2016; Mota 2016b). Posteriormente, na Secção 1.4., descrevem-se os sistemas de classificação nominal das LM dos aprendentes, autores dos textos escritos selecionados para o *corpus* desta investigação.

Neste trabalho analisaram-se os comportamentos desviantes de atribuição de valores de género aos nomes (AGN) e de concordância nominal em género (CNG), detetados no sintagma nominal<sup>1</sup> (SN), produzidos por falantes nativos de espanhol, italiano, alemão, inglês e chinês. Selecionaram-se estes idiomas, pois, entre eles, é possível distinguir:

- (i) os que possuem um sistema de classificação nominal baseado em género com propriedades muito próximas do português (espanhol e italiano);

---

<sup>1</sup> Convém sublinhar que, por questões de conveniência, neste trabalho optou-se por utilizar a expressão *Sintagma Nominal* (SN) para identificar as estruturas constituídas por um núcleo nominal, o item lexical, e outros constituintes sintáticos, nomeadamente os especificadores (determinantes ou quantificadores) e/ou modificadores (adjetivos) com função atributiva, sendo que todos estes elementos estabelecem com o nome relações de concordância sintática. Note-se ainda que, sempre que for pertinente, é feita a referência à expressão *Sintagmas Nominais não reduzidos* por forma a poder estabelecer a distinção entre as estruturas de SN não reduzidas e os *Sintagmas Nominais reduzidos*, que contêm apenas o núcleo nominal.

- (ii) os que possuem um sistema de atribuição de género nominal e de concordância nominal mais complexo do que o português, apresentando três valores de género, masculino, feminino e neutro, como é o caso da língua alemã; e
- (iii) os que não possuem um sistema de classificação dos nomes baseado em valores de género, como o chinês, ou que possuem pequenos vestígios dessa categoria, como se verifica no inglês.

Na Secção 1.5., procede-se a uma revisão da literatura disponível acerca da aquisição do género gramatical por falantes nativos e por falantes não nativos, a fim de averiguar as principais particularidades associadas a um e a outro processo.

Descrita a categoria gramatical em estudo, o **Capítulo 2** é dedicado à revisão das principais teorias atuantes no âmbito da investigação em ALNM. Num primeiro momento, procede-se à descrição das particularidades relativas à aquisição/aprendizagem tardia de uma LNM (Secção 2.2.). Com vista à identificação do papel que, nos diferentes modelos teóricos, é atribuído aos conhecimentos linguísticos prévios ao longo da aquisição/aprendizagem das propriedades gramaticais de um idioma não nativo (Schwartz & Sprouse 1996; Hawkins & Chan 1997; Franceschina 2003, 2005; Tsimpli & Dimitrakopoulou 2007), na Secção 2.3. analisa-se o conceito de ‘transferência linguística’ e a sua implicação para a aquisição/aprendizagem da categoria de género gramatical, com referência a hipóteses teóricas evocadas na literatura para justificar as condições subjacentes à transferibilidade do género gramatical (Bruhn de Garavito & White 2002; White *et al.* 2004; Sabourin *et al.* 2006).

Após a revisão teórica empreendida nos dois capítulos iniciais, procedeu-se ao tratamento dos dados empíricos. Deste modo, no **Capítulo 3** é descrita a metodologia adotada para o tratamento dos dados selecionados, apresentando-se, num primeiro momento, na Secção 3.1.1., as questões de investigação e as variáveis de análise, relacionadas não só com o perfil dos informantes, aprendentes tardios de PLNM (LM e nível QECRL), mas também com as características do *input* da LA, nomeadamente, as propriedades subjacentes à associação de valores de género aos nomes do português. Deste modo, com a análise dos dados empíricos, neste trabalho pretende-se responder às seguintes perguntas:

- **P1:** O desempenho linguístico do aprendente tardio relativamente à atribuição dos valores de género aos nomes e à concordância nominal em género no português beneficia do facto de ter representada na gramática da sua LM esta categoria gramatical?

---

- **P2:** O valor de género gramatical é uma propriedade transferível da LM para a LNM?

---

- **P3:** Qual o grau de sensibilidade dos diferentes grupos de aprendentes relativamente aos indícios semânticos e formais de atribuição de valores de género aos nomes em português como LNM?

---

- **P4:** Haverá, nos diferentes grupos de aprendentes, uma maior incidência de comportamentos desviantes de atribuição de valores de género aos nomes ou de concordância nominal em género nos especificadores (determinantes e quantificadores) ou nos modificadores (adjetivos) com função atributiva?

Apresentadas as perguntas que norteiam a investigação, na Secção 3.2. expõe-se a constituição da base empírica do estudo, descrevendo-se detalhadamente, em primeiro lugar, as bases de dados consultadas (Secção 3.2.1.) e, em segundo, os pressupostos subjacentes à seleção do *corpus* da presente investigação (Secção 3.2.2.). Posteriormente, em 3.3., apresenta-se o perfil dos informantes, autores dos textos escritos que compõem o acervo textual em estudo. Por fim, a Secção 3.4. é dedicada à explanação dos procedimentos adotados no tratamento dos dados empíricos, com a referência à tipologia de desvios, delineada em função das características do sistema de atribuição de género nominal do português. Para a estipulação dos diferentes tipos de desvios, teve-se em conta o peso dos indícios, semânticos e formais, de associação dos valores de género aos nomes em português, distinguindo-se, assim, os desvios que incidem sobre: (i) nomes cujo valor de género depende única e exclusivamente de critérios semânticos; (ii) nomes cujo valor de género depende de critérios semânticos e formais; (iii) nomes cuja atribuição de género depende unicamente de critérios formais, nos quais se observa uma correlação entre constituintes temáticos *-o* e *-a* e valores de género masculino e feminino, respetivamente; e (iv) nomes cujo valor de género não depende nem de critérios semânticos nem de critérios formais.

Em complemento, para cada categoria, assinalaram-se os constituintes sintáticos que no SN apresentam marcas desviantes de associação de valores de género. Deste modo, identificou-se na análise dos dados empíricos, as manifestações de desvio nos

especificadores, determinantes e quantificadores, nos modificadores e na forma morfológica do item nominal.

O **Capítulo 4** é dedicado à apresentação e discussão dos resultados apurados. Num primeiro momento, são referidos os índices globais de desvios apurados no *corpus* (Secção 4.2.1.) e por variáveis de análise (LM e nível QECRL dos informantes) (cf. Secção 4.2.2.). Já em 4.2.3. é feita a análise detalhada, quantitativa e qualitativa, por segmentos da amostra, dos tipos de desvios apurados e dos respetivos constituintes nos quais se manifestam as marcas de desvios.

Por fim, na última parte da monografia, em **Considerações finais**, é feita uma súmula geral do trabalho empreendido. Assim sendo, procede-se a uma breve revisão dos resultados apurados, procurando-se ainda, em função dos resultados, responder às perguntas colocadas para a investigação.

# Capítulo 1

## A categoria de género gramatical

---

*“Gender is the most puzzling of the grammatical categories”*

(Corbett 1991:1)

### 1.1. Introdução

Este capítulo tem como principal objetivo a descrição e análise dos sistemas de género das línguas maternas (LM) dos aprendentes autores dos textos escritos que compõem o *corpus* da presente investigação. No sentido de se averiguar qual o (possível) papel que a LM desempenha durante o processo de aquisição/aprendizagem do sistema de atribuição de género do português como língua não materna (PLNM), foram selecionados idiomas com sistemas de classificação nominal muito próximos do português — o espanhol, o italiano, o alemão, para além de outros que não apresentam a categoria de género gramatical: o inglês e o chinês. Assim, entre as línguas com género gramatical, encontramos idiomas com dois valores de género, como o espanhol e o italiano, e o alemão com três valores distintos (masculino, feminino e neutro). Entre os idiomas que não possuem género morfossintaticamente expreso, estão o inglês, que ostenta vestígios de um primitivo sistema de género ao nível dos pronomes, enquanto que o chinês não possui uma classificação sistemática e obrigatória dos nomes baseada em género morfossintaticamente expreso, apresentando um sistema de classificação nominal baseado em classificadores.

Refira-se, agora, a estrutura deste capítulo. Num primeiro momento, é feita a descrição do conceito de género gramatical (Secção 1.2.1.), propriedade relevante, em algumas línguas, para a classificação dos nomes e com características que o distinguem de outros sistemas de classificação nominal, nomeadamente os classificadores (Secção 1.2.2.). Na descrição dos sistemas de género analisam-se também as questões relativas à sua origem (Secção 1.2.3.) bem como aos critérios de atribuição de género nominal (Secção 1.2.4.).

Após esta exposição inicial, procede-se à análise do sistema de atribuição de género nominal do português (Secção 1.3.). Em seguida, descrevem-se os sistemas de classificação nominal das LM dos aprendentes, autores dos textos escritos que compõem o *corpus* da presente investigação. Apresenta-se, em primeiro lugar, o sistema de classificação nominal

das línguas românicas, o espanhol (Secção 1.4.1.) e o italiano (Secção 1.4.2.), considerando-se, em seguida, as características do sistema de atribuição de género nominal do alemão e do respetivo funcionamento do sistema de concordância nominal (Secção 1.4.3.). Posteriormente, analisam-se os aspetos relacionados com as particularidades do sistema pronominal do inglês e de como nele se manifestam os últimos vestígios de um primitivo sistema de classificação nominal baseado em valores de género (Secção 1.4.4.). Por fim, procede-se à descrição do sistema de classificação nominal da língua chinesa, baseado em classificadores (Secção 1.4.5.). O último ponto do capítulo é dedicado à revisão dos trabalhos sobre a aquisição do género e da concordância de género, por falantes nativos e não nativos, com destaque para os aprendentes do português (Secção 1.5.).



## 1.2. O género gramatical

Ao longo das últimas décadas, a investigação tem-se debruçado sobre as especificidades relativas ao fenómeno do género gramatical e, independentemente do quadro teórico em que se situam, os trabalhos realizados nesta área (Ibrahim 1973; Corbett 1991, 2005, 2006b, 2013c; Ritter 1993; Chomsky 1995 [1999]; Comrie 1999; Aikhenvald 2000, 2004; Carvalho 2000; Grinevald 2000; Kibort & Corbett 2008) têm colocado em evidência a natureza idiossincrática de que se reveste esta categoria, nomeadamente no que concerne:

- (i) aos critérios que cada língua seleciona e que servem de base à definição desta categoria gramatical;
- (ii) à aplicação, dentro de cada língua, dos critérios (semânticos e formais) de atribuição de género nominal; e
- (iii) à manifestação formal dos valores de género.

O termo ‘género’ provém do étimo latino *genus, -eris* que, do ponto de vista gramatical, é utilizado para designar uma categoria relevante para a classificação dos nomes (por ex., nomes de género masculino, de género feminino, comuns de dois, etc.), sendo que palavras de outras classes, que estabelecem com o nome relações de concordância, são igualmente marcadas quanto ao valor de género. A singularidade desta categoria gramatical evidencia-se, em primeiro lugar, pelo facto de não possuir um estatuto universal, uma vez que nem todas as línguas do mundo apresentam sistemas de classificação nominal baseados em género<sup>2</sup> (cf. Figura 1.1, Corbett 1991:1; Nichols 1999:124) havendo outras que, já o tendo tido em fases históricas anteriores, foram-no perdendo no decurso da sua evolução (Aikhenvald 2000:379). A língua inglesa configura um exemplo típico em que, ao longo do seu percurso evolutivo, se foi assistindo à perda gradual de um sistema de classificação de nomes baseado em valores de género (Ibrahim 1973:86) (cf. Secção 1.4.4.).

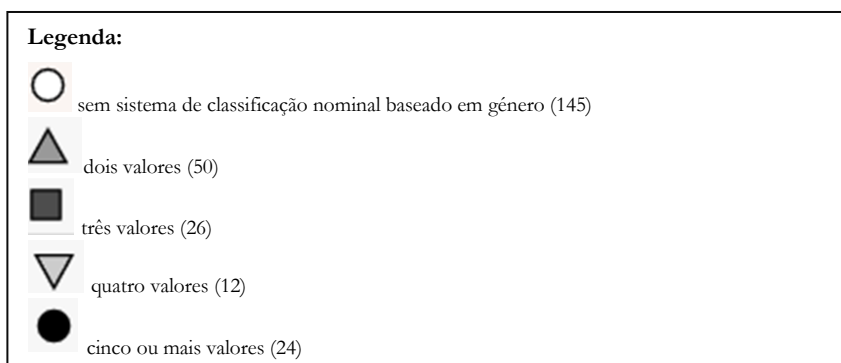
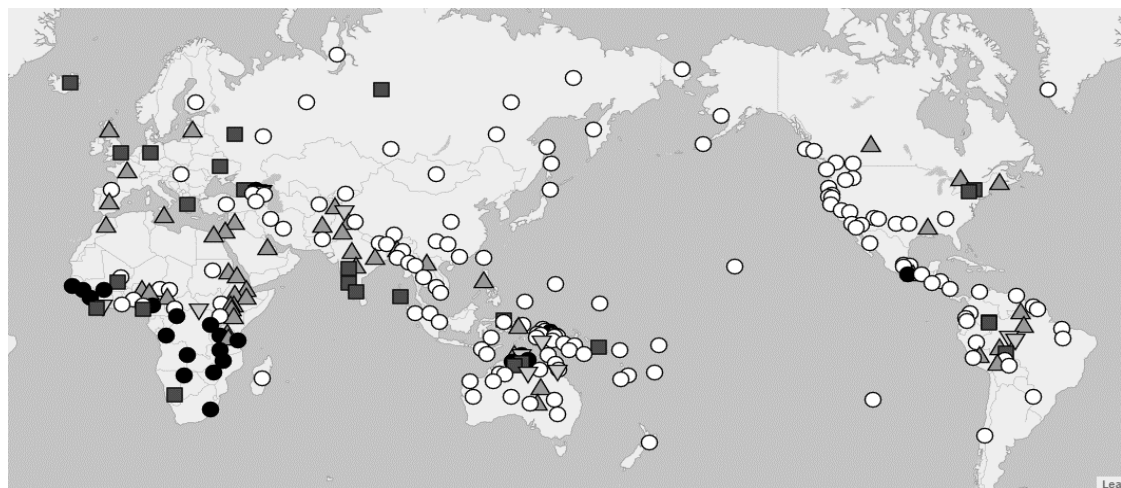
Além disso, há uma multiplicidade de critérios que cada língua seleciona e que servem de base à definição desta categoria (Moravcsik 1978:336; Contini-Morava & Kilarski 2013:271). A este respeito, observa Vilela (1973:141) que, para além da oposição masculino/feminino:

---

<sup>2</sup> Partindo da descrição da estrutura formal das línguas de matriz indo-europeia e da sua comparação com outros idiomas, Ilari (2013:51) conclui que o género não é uma categoria essencial da linguagem humana.

“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”.

O mapa seguinte (Figura 1.1) apresenta a distribuição do número de valores de género existente em várias línguas do mundo. Conforme se verifica, grande parte das línguas representadas na figura (cerca de 145) não exhibe sistemas de classificação nominal baseados em género. Os sistemas de género com dois valores são os mais frequentes, estando presentes em cerca de 50 idiomas, havendo 26 línguas com três valores de género. Sistemas com quatro valores são pouco frequentes, registados apenas em 12 idiomas. Conforme se verifica no mapa, é no continente africano que se localiza a maior parte das línguas com cinco ou mais valores de género.



**Figura 1.1** – Distribuição do número de valores de género em várias línguas do mundo (fonte: *World Atlas of Language Structures Online* [WALS], 30A, <http://wals.info/chapter/30>)

Registam-se na literatura diferentes aceções da categoria de género, nem sempre sendo fácil determinar, com clareza, o lugar que, decorrente destas aceções, esta categoria ocupa na estrutura gramatical dos idiomas. De acordo com a proposta de Greville Corbett (1991, 2005, 2006a, 2006b) o género corresponde a uma categoria inerente aos nomes, não sendo em outras classes de palavras como, por ex., determinantes, adjetivos, verbos, etc. Trata-se, portanto, de uma categoria morfossintática, sendo que cada item nominal possui, tipicamente, um único valor de género intrínseco (Harris 1991:36; Corbett 2006a:126; Mota 2016b:156). De acordo com Corbett (1991), a atribuição de um determinado valor de género ao item nominal pode estar dependente quer do significado deste, quer da sua forma, sendo que tais propriedades (semânticas e formais) só serão consideradas como instanciações de uma categoria de género numa dada língua se nela espoletarem o fenómeno de concordância sintática (cf. Secção 1.2.1.1.). Deste modo, e nas palavras de Hockett (1958:231) os valores de género “are classes of nouns reflected in the behavior of associated words.”. Considera-se, então, que numa língua há tantos valores de género quantas as possibilidades de concordância sintática desencadeadas pelo item nominal.

Modelos teóricos mais recentes descrevem a categoria de género à luz dos contributos da teoria generativa e, mais precisamente, do Programa Minimalista proposto por Chomsky (1995 [1999]). Em linhas gerais, nesta perspectiva teórica é assumido que o género gramatical corresponde a um traço de natureza formal que está disponível no Léxico<sup>3</sup> de uma língua a fim de ser selecionado pelos radicais dos itens nominais em função de diversas operações sintáticas (Ritter 1993). Por conseguinte, um item lexical define-se quer pela seleção do radical, quer pela respetiva associação dos traços formais, sendo certo que numa dada língua a especificação do valor de género nominal é paramétrica, já que está dependente das características estruturais desse idioma (Bruhn de Garavito & White 2002:154; White *et al.* 2004:109; Choupina, Baptista & Costa 2014). Nesta perspectiva, o género pode corresponder a um traço inerente, atribuído no Léxico, como no nome *papel* ou a um traço opcional, atribuído na Sintaxe, como, por ex., em *o cliente, a cliente* (Chomsky 1995 [1999]; Baptista *et al.* 2013:35), já que, em alguns itens, “a especificação

---

<sup>3</sup> De acordo com a perspectiva minimalista “[o] léxico é um conjunto de elementos lexicais, cada um deles um sistema articulado de traços. O léxico especifica, para cada um destes elementos, as suas propriedades fonéticas, semânticas e sintáticas idiossincráticas (...)” (Chomsky 1995 [1999:198]). Ainda nesta perspectiva, é assumido que numa língua, a par do Léxico, existe um sistema computacional, a Sintaxe, que exhibe “três níveis de representação fundamentais, cada um deles constituindo uma «interface» do sistema gramatical com algum outro sistema da mente/cérebro: a Estrutura-D, a Forma Fonética (PF), e Forma Lógica (LF)” (Chomsky 1995 [1999:198]).

não está no radical e verifica-se apenas na Sintaxe com consequente atribuição do valor de género e desencadeamento dos mecanismos de concordância” (Costa *et al.* 2015:332; veja-se ainda Choupina, Baptista & Costa 2014).

Nas secções seguintes, tomaremos em linha de conta as implicações destas várias acções para a descrição dos sistemas de atribuição de género nominal, quer do português, quer das demais línguas seleccionadas para este estudo.

### 1.2.1. Género e Classes nominais

Por se tratar de uma categoria essencial na classificação dos nomes, o termo ‘género’ surge, frequentemente, associado ao conceito de ‘classe nominal’<sup>4</sup> na medida em que estas duas designações são utilizadas para identificar tipos de classificação nominal associados à concordância sintáctica obrigatória e que se distinguem de outros sistemas em que tal obrigatoriedade não se verifica, como, por exemplo, o sistema de classificadores no qual se reconhecem vários tipos e subtipos (nominais, numerais, verbais) (Aikhenvald 2000, 2004; Grinevald 2000, 2004; Contini-Morava & Kilarski 2013) (cf. Secção 1.2.2.).

De acordo com Grinevald (2000), os sistemas de classificação nominal organizam-se em função do grau de gramaticalização em que se encontram e compõem-se num *continuum* lexical-gramatical<sup>5</sup> (cf. Figura 1.2). Num dos extremos desse *continuum* situa-se “a major grammatical system” que, segundo a autora, abrange, quer os sistemas de classificação nominal baseados em classes nominais, quer os sistemas de classificação baseados em género gramatical (Grinevald 2000:55). No extremo oposto, encontram-se os sistemas de classificação nominal estritamente lexicais<sup>6</sup>. Já os classificadores ocupam uma

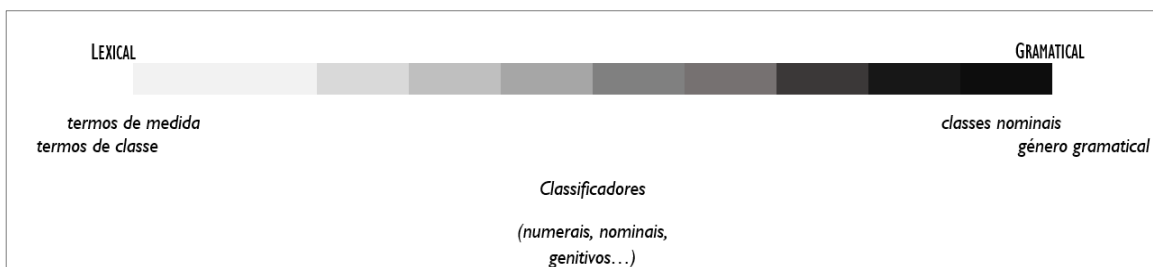
---

<sup>4</sup> Regista-se, na literatura, uma certa tendência para se associarem os dois conceitos, entendidos para muitos autores como termos equivalentes (Dixon 1986; Comrie 1999; Corbett 1991). Aliás, na própria definição da categoria de género, Hockett (1958:231) utiliza o termo ‘classe nominal’.

<sup>5</sup> Na sua proposta, Grinevald clarifica ainda os termos ‘lexical’ e ‘gramatical’ associados aos sistemas de classificação nominal. Assim, e segundo a autora (2000:55), “[l]exical’ (...) means (a) part of the lexicon and its word-building dynamics and (b) semantically compositional, while ‘grammatical’ means part of the morphosyntax of a language.”.

<sup>6</sup> Grinevald (2000) distingue dois tipos de classificação nominal de natureza estritamente lexical: (i) os termos de quantificação e (ii) os termos de classe. O primeiro encontra-se em todas as línguas do mundo e corresponde a itens lexicais de quantificação que acompanham, quer nomes contáveis (em português, por ex., um ‘grupo’ de jovens; uma ‘pilha’ de processos), quer nomes massivos (um ‘copo’ de água, uma ‘cabeça’ de gado). Já os termos de classe correspondem a morfemas classificatórios com uma clara origem lexical e entre os termos de classe, Grinevald destaca os que dizem respeito ao domínio semântico da botânica. Em inglês, por exemplo, a diferença entre nomes de árvores e de frutos é estabelecida através de um processo de natureza morfológica, com a adição de morfemas de classe a

posição intermédia neste *continuum*, ou seja, distanciam-se dos sistemas de classificação nominal estritamente lexicais porque correspondem a constituintes independentes que não estão intrinsecamente ligados ao nome, sendo que múltiplos fatores podem condicionar o uso desses mesmos constituintes. Para além disso, distanciam-se dos sistemas de classificação nominal ditos ‘gramaticais’, como o de género, pelo facto de ainda possuírem vestígios de uma primitiva origem lexical e de, por vezes, se circunscreverem a contextos sintáticos particulares (Grinevald 2000:61).



**Figura 1.2** – Os sistemas de classificação nominal organizados num *continuum* lexical-gramatical (figura baseada na proposta de Grinevald (2000:61)).

Alguns autores (Ibrahim 1973; Dixon 1986; Grinevald 2000) preferem um tratamento diferenciado <sup>7</sup> dos conceitos de ‘género’ e de ‘classe nominal’. Em contrapartida, existem outros investigadores (Moravcsik 1978; Corbett 1991; Nichols 1999; Aikhenvald 2000) que tratam os dois conceitos como sinónimos, como se constata, por exemplo, no trabalho desenvolvido por Corbett (1991) que utiliza a palavra ‘género’ para descrever os sistemas de classificação nominal das línguas que se definem pela necessidade concomitante de concordância sintática. Ainda que de modo muito breve, referem-se, em seguida, os principais argumentos que sustentam, por um lado, o uso diferenciado dos dois conceitos (género e classe nominal) e, por outro, a utilização preferencial de um termo em detrimento do outro.

A palavra ‘género’ é utilizada tradicionalmente para identificar os sistemas de classificação característicos de línguas de matriz indo-europeia e de línguas semíticas que possuem um número muito restrito de valores, essencialmente dois ou três (masculino,

uma base ( $X$ -berry<sub>fruto</sub> vs.  $X$ -tree<sub>árvore</sub>), como em *strawberry*<sub>fruto</sub> ‘morango’; *blueberry*<sub>fruto</sub> ‘mirtilo’ e *apple tree*<sub>árvore</sub> ‘macieira’; *olive tree*<sub>árvore</sub> ‘oliveira’. (cf. Grinevald 2000:58-60).

<sup>7</sup> É preciso clarificar que o tratamento ‘diferenciado’ dos termos não significa que os autores os concebem como noções opostas. Na verdade, e como afirma Grinevald (2000:57): “(...) gender and noun classes are treated as one major system of nominal classification (...). The tradition is to call gender those systems which rely on the feature of sex, and to take the system of Indo-European languages as the prototypical cases, while labelling ‘noun classes’ the larger systems which are considered more ‘exotic’ from a Eurocentric point of view.”.

feminino e neutro), segundo os quais os nomes se organizam. Nestes idiomas, todos os nomes possuem um valor de género, apesar de este nem sempre estar explicitamente marcado na sua forma morfológica. Além disso, e embora se verifique, nos nomes que denominam entidades sexuadas, uma estreita relação entre o género natural do referente (macho/fêmea) e o género gramatical (masculino/feminino), a associação dos valores de género aos nomes é, em grande parte dos casos, idiossincrática (Corbett 1991; Aikhenvald 2000; Grinevald 2000).

Por sua vez, o termo ‘classes nominais’ associa-se, em geral, a sistemas de classificação nominal típicos de línguas africanas<sup>8</sup>. Nestes sistemas, há um maior número de valores de género — que pode ascender aos vinte — estabelecidos em função de diversas propriedades semânticas dos referentes nominais, sendo que nem sempre o critério do sexo das entidades é relevante para a organização dos nomes (Ibrahim 1973; Corbett 1991:48). A título de exemplo considere-se o sistema de classes nominais do suaíli, língua *bantu*, cujas classes se definem de acordo com as várias características semânticas dos referentes dos nomes, distinguindo-se os nomes de entidades [+humanas], os que se referem a animais, a plantas, a frutas, a objetos de diferentes dimensões, entre outros (Corbett 1991:48; Grinevald 2004:1019; Aikhenvald 2004:1031).

No que concerne às manifestações formais, observam-se, todavia, algumas diferenças entre os sistemas de classificação nominal denominados de ‘género’ e os de ‘classes nominais’. Efetivamente, nos sistemas de género, e apesar de todos os itens nominais possuírem um valor inerente<sup>9</sup>, este nem sempre é facilmente dedutível a partir das

---

<sup>8</sup> Para além de serem comuns nas línguas africanas, sobretudo nas línguas *bantu*, as ‘classes nominais’ também são características de alguns idiomas australianos e caucasianos (Ibrahim 1973:64).

<sup>9</sup> É sabido que, de um modo geral, nas línguas com a categoria de género, cada nome possui um único valor. No entanto, em vários idiomas encontram-se casos de nomes que admitem mais do que um valor de género, como, em português, os nomes comuns de dois géneros “o/a *motorista*; o/a *viajante*”. Como se viu, para alguns investigadores (Costa *et al.* 2015:332; Choupina *et al.* 2016:208), o género corresponde, nestes casos, a um traço opcional, selecionado por via da concordância sintática. No entanto, segundo Mota (2016b), os nomes ditos comuns de dois valores de género correspondem a dois lexemas, cada um associado a um valor de género específico, masculino ou feminino: *o artista* e *a artista*. Por conseguinte, segundo esta proposta, cada entrada lexical, masculina ou feminina, vai ser ativada na sintaxe de acordo com o sexo da entidade a que o nome se refere. Há ainda itens, considerados *híbridos* (Corbett 1991) que, apesar de estarem associados a um determinado valor, admitem diferentes padrões de concordância. Um exemplo paradigmático destes casos é o nome alemão *Mädchen* que, embora se refira a uma entidade sexuada, traduzindo-se por ‘menina jovem’, é uma forma de género neutro. Por conseguinte, e respeitando o valor de género gramatical, todas as restantes classes de palavras associadas a este item devem ser igualmente neutras. No entanto, por vezes, este nome é retomado anaforicamente pelo pronome pessoal feminino ‘sie’ em vez do neutro

suas propriedades fonomorfológicas e mesmo semânticas do nome (Corbett 1991). Por exemplo, em português, embora haja uma tendência para palavras terminadas em *-a* serem de género feminino *a casa, a mesa, a lanterna*, encontramos também nomes masculinos com a mesma terminação, tais como *o mapa, o dia, o programa*. O mesmo acontece se se considerar o critério semântico, já que nem sempre, nos nomes cujo referente é uma entidade sexuada, os valores de género (masculino/feminino) correspondem ao sexo dos respetivos referentes. É o caso dos nomes epícenos e dos sobrecomuns, que designam, no primeiro caso, animais e, no segundo, pessoas, cujo valor de género gramatical não corresponde, necessariamente, ao género natural do referente: *o elefante, a criança, a cobra, a pessoa* (cf. Secção 1.3.3.).

Já em línguas de ‘classes nominais’ os nomes estão geralmente marcados de forma explícita, i.e., possuem certos prefixos que, além de identificarem a classe a que pertencem, determinam geralmente as formas de concordância das restantes palavras (Ibrahim 1973:64; Moravcsik 1978:338; Grinevald 2000:57; Dixon 1986:105). Por exemplo, no suaíli os prefixos *ki-/vi-* identificam os nomes da classe 7/8, sendo que outras classes de palavras concordantes também apresentam os mesmos prefixos<sup>10</sup> (Corbett 1991:47).

Admitindo que existe uma certa relação entre sistemas de ‘género’ e de ‘classes nominais’, Ibrahim (1973) afirma que o ‘género’ configura um caso especial dentro dos sistemas de ‘classes nominais’. Assim, este investigador prefere a designação de ‘género’ para se referir aos sistemas de categorização dos nomes presentes em línguas em que o sexo é o critério (semântico) determinante para a constituição dos valores (masculino/feminino/neutro), utilizando o conceito de ‘classes nominais’ para designar os sistemas cujos valores se definem com base em diferentes propriedades semânticas dos referentes (por ex., tamanho, forma e função dos objetos). No entanto, e apesar de adotar dois termos distintos, o autor afirma que tal distinção terminológica é irrelevante para a

---

‘es’ preferindo-se, de certo modo, a classificação semântica em detrimento da gramatical (cf. Corbett 1991:183).

<sup>10</sup> Moravcsik (1978:338-339) chama ainda a atenção para o facto de numa língua como o suaíli, apesar de os nomes apresentarem prefixos que os identificam como pertencentes a uma determinada classe e de tais prefixos se manifestarem em constituintes que estabelecem com o nome relações de concordância sintática, como os adjetivos, a verdade é que o mesmo nem sempre se verifica nos casos de retoma anafórica pronominal. Refere a investigadora que “all nouns referring to human beings (and some referring animals) regardless of their prefixes require anaphoric pronouns of a uniform shape – of a shape that otherwise occurs with nouns belonging to the first nominal prefix class (which (...) includes mostly human nouns)” (Moravcsik 1978:338-339).

descrição gramatical, uma vez que os dois tipos de classificação nominal se definem pela obrigatoriedade do fenómeno da concordância sintática (1973:76).

Ao constatar que a concordância é o critério determinante para a definição destes sistemas, Corbett (1991) considera que as duas designações ('género' e 'classes nominais') resultam, simplesmente, da adoção de diferentes tradições linguísticas e não da presença de efetivas diferenças estruturais<sup>11</sup>, sublinhando que ambas designam um único fenómeno. Consequentemente, este autor utiliza o termo 'género' para se referir ao tipo de categorização nominal que se define pela presença obrigatória da concordância sintática<sup>12</sup>, distinto de outros sistemas em que tal obrigatoriedade não se verifica (Corbett 1991:5; 2005:986; 2006b:750).

Face ao exposto, constata-se, de facto, uma certa flutuação terminológica relativamente à utilização dos termos 'género' e 'classe nominal', e tal como se pretendeu demonstrar, esta polivalência resulta, essencialmente, de questões relativas à tradição linguística seguida pelos investigadores. Neste trabalho, subscreve-se a posição defendida por Corbett (1991) preferindo-se o termo 'género' para designar os sistemas de classificação nominal que se definem pela obrigatoriedade do fenómeno da concordância sintática.

#### **1.2.1.1. A concordância sintática**

Como vimos, a concordância sintática é um elemento fundamental para a definição dos sistemas de classificação nominal baseados em género gramatical (Ibrahim 1973; Corbett 1991). Importa, por isso, refletir, ainda que de forma muito breve, sobre as particularidades associadas a este fenómeno de natureza gramatical bem como sobre os problemas relativos à sua delimitação (Steele 1978; Corbett 1991, 2003; 2006a; Chomsky 1995 [1999]; Siewierska 1999).

A concordância não é um fenómeno universal (Corbett 2006a:1). Por exemplo, em português, a concordância é uma estrutura linear, i.e., consiste na colocação reiterada de

---

<sup>11</sup> Observa Aikhenvald (2004:1034) que os conceitos de 'género' e de 'classe nominal' são, muitas vezes, utilizados indistintamente.

<sup>12</sup> Há ainda um uso muito particular do termo 'género'. Refere Aikhenvald (2000:9) que "[i]n the Athabaskan linguistic tradition the term 'gender' is used to refer to verbal classifiers which mark agreement with intransitive subject or transitive object, and characterize the referent noun in terms of shape and form (...)".



marcas de concordância em vários elementos que constituem a frase. Nem todas as línguas admitem, no entanto, o estabelecimento de relações de concordância sintática.

De acordo com Steele (1978:610) a concordância corresponde a:

“some systematic covariance between a semantic or formal property of one element and a formal property of another. For example, adjectives may take formal indication of number and gender of the noun they modify.”

Portanto, e em linhas gerais, a concordância define-se como uma relação sintática que se estabelece entre dois elementos, no qual a forma de um determina a forma do outro (Foucart 2008:26). Trata-se de um fenómeno que implica a interseção dos diferentes planos estruturados de uma língua, uma relação entre as palavras estabelecida, sobretudo, ao nível sintático, semântico e morfológico (Corbett 2006a). Na descrição deste fenómeno, Corbett (1998, 2006a) observa que a concordância se manifesta sob diferentes condições que tipicamente envolvem a relação formal entre dois elementos linguísticos: o elemento controlador e o elemento controlado, o alvo. Estes dois elementos partilham, por sua vez, um traço de natureza formal, tais como de género, de número e de pessoa, com um determinado valor, como masculino ou feminino para o género, por exemplo. Desta forma, o controlador determina o valor a ser assumido pelos traços formais dos elementos que compõem o domínio sintático, sendo que o elemento controlado, ou seja, o alvo, se encontra sintaticamente relacionado com o elemento controlador (Corbett 1998:191, 2006a:11).

A concordância é, portanto, uma relação sintática que se manifesta morfológicamente o que, por sua vez, implica: (i) a existência de valores opositivos (número: singular vs. plural; género: masculino vs. feminino; pessoa: 1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup> do singular e do plural); (ii) a presença de morfemas flexionais nas classes de palavras variáveis (nomes, pronomes, especificadores e modificadores), sendo certo que nem todas as palavras variam<sup>13</sup>; e (iii) a presença reiterada de traços formais nas palavras que sustentam as relações de concordância entre si (cf. (1-2)):

---

<sup>13</sup> Num sintagma com nome e adjetivo, este último deve concordar em género e em número com o nome. Por exemplo, o adjetivo ‘pequeno’ associado a um nome masculino toma a forma de masculino *um carro pequeno* e quando associado a um nome feminino, toma a forma do feminino *uma casa pequena*. Há, contudo, alguns adjetivos que não apresentam qualquer variação formal, como é o caso dos adjetivos uniformes ‘grande’ e ‘simples’. Assim, temos *um quarto grande* e *uma janela grande, um texto simples* e *as canções simples*.

- (1) Nesta rua, há vários prédios altos e uma casa grandiosa.
- (2) Eles não sabem onde vive a Cátia.

No que concerne à natureza dos fenómenos linguísticos que estão sob o escopo da concordância sintática, registam-se, na literatura, diferentes aceções, não havendo, contudo, uma resposta clara a este respeito (Corbett 2003:112). Por um lado, numa visão mais lata de concordância, assume-se que fenómenos linguísticos como a retoma anafórica pronominal são igualmente determinados pelos mecanismos de concordância (Moravcsik 1978:338; Corbett 2006a:21). Por outro lado, numa aceção mais restrita, considera-se que a concordância é simplesmente um fenómeno de natureza sintática, não se devendo incluir nele a retoma anafórica pronominal (Ibrahim 1973; Aikhenvald 2000). Na perspectiva de Corbett (1991, 2003, 2006a) não existem evidências claras que sustentem esta última posição, pelo que o autor adota uma visão mais abrangente de concordância<sup>14</sup>. Também Siewierska (1999:225) defende que “there is no good basis for distinguishing between agreement and antecedent-anaphora relations”.

No que à concordância de género gramatical diz respeito, a aceção mais lata do conceito de concordância proposta por Corbett (1991) leva o investigador a assumir que o facto de numa língua os pronomes apresentarem traços formais de género é suficiente para admitir que essa mesma língua exibe um sistema de atribuição de género gramatical. Por conseguinte, o investigador defende que o inglês exibe a categoria de género (Corbett 1991:169). Também Aikhenvald (2000:30) observa que, decorrente da estrutura formal de concordância, algumas línguas:

“have different noun class/gender agreement systems depending on the domain of agreement (...) and on the morphological class of the agreeing element.”

Não obstante os argumentos que sustentam esta posição (cf. Corbett 2003, 2006a) consideramos que a simples existência de processos de retoma anafórica pronominal não é suficiente para se admitir que uma língua exibe traços formais de concordância e, conseqüentemente, um sistema de atribuição de género gramatical. Nesta medida, entendemos que a concordância se define pela obrigatoriedade, sistematicidade e reiteração dos traços formais e, portanto, em linha com a proposta de Aikhenvald (2000:29), implicará

---

<sup>14</sup> Corbett (2006a:22) conclui que “agreement covers feature covariance in a range of domains, from within the noun phrase to antecedent-anaphor relations.”

sempre “a systematic covariance between the grammatical meanings of grammatical morphemes”. Deste modo, assumimos neste estudo que a língua inglesa não exhibe atualmente um sistema de classificação nominal baseado em valores de género. Com efeito, os dados disponíveis sobre a evolução deste idioma revelam uma perda gradual dos traços de concordância sintática, prevalecendo, atualmente, vestígios de um primitivo sistema de classificação nominal baseado em valores de género sobretudo na marcação dos pronomes de 3ª pessoa do singular (cf. Secção 1.4.4.).

Do ponto de vista estrutural, a concordância constitui uma vantagem para o processamento cognitivo das categorias gramaticais e ativa, por seu turno, a identificação de relações sintáticas e semânticas entre as unidades do léxico<sup>15</sup> (Franceschina 2005:81; Corrêa & Augusto 2017). Com efeito, e como veremos mais adiante (Secção 1.5.), relativamente à aquisição do género por falantes nativos, estudos vários revelam a importância do reconhecimento das relações de concordância sintática entre elementos da frase para a assimilação dos valores de género dos nomes (Augusto & Corrêa 2005).

### 1.2.2. Género e Classificadores

Como vimos, os sistemas de classificação nominal baseados em género definem-se pela presença obrigatória da concordância sintática, sendo esse o critério que os distingue de outros sistemas como, por exemplo, o dos classificadores. Conforme se verificou na Figura 1.2, os classificadores ocupam uma posição intermédia num *continuum* lexical-gramatical, distinguindo-se dos sistemas de classificação nominal estritamente lexicais, por um lado, e, por outro, dos sistemas de classificação dos nomes considerados ‘gramaticais’ (Contini-Morava & Kilarski 2013:266).

---

<sup>15</sup> Num trabalho desenvolvido por Vigliocco & Frank (1999), as autoras analisaram um conjunto de hipóteses relativas às implicações relacionadas com a natureza concetual, i.e., semântica, dos elementos lexicais para o processamento das relações de concordância sintática. Segundo as investigadoras, em modelos de descrição puramente formais enfatiza-se a ideia de que é necessário separar dois tipos de processamento: o gramatical e o concetual. Nesta perspetiva, a concordância é vista puramente como uma operação sintática que consiste no transporte de traços formais (de número e de género, por ex.) de um controlador para um alvo. A favor desta hipótese está o facto de em grande parte das línguas se verificar uma preferência pela concordância sintática, i.e., puramente gramatical em detrimento da concordância concetual ou semântica. As autoras ilustram esta posição com o seguinte exemplo (1999:458): “[i]n British English (...) collective nouns may take a plural verb as in example (6). However, noun-specifier agreement has to be syntactic, and therefore (7) is ungrammatical.

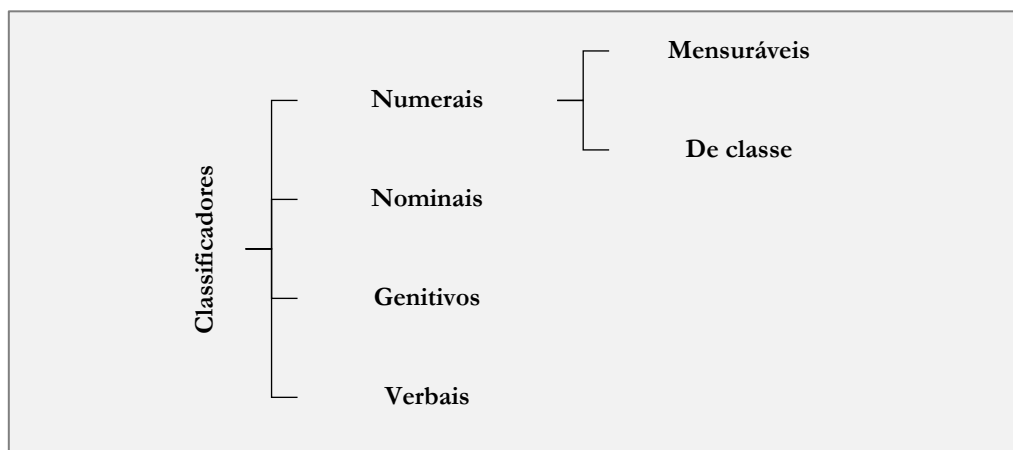
(6) The committee are voting themselves a raise.

(7) \*These committee voted for a raise.”

Para além disso, observam as autoras que as crianças adquirem as relações de concordância sintática entre os itens muito antes de serem capazes de compreender a natureza concetual dos traços formais de concordância.

Os classificadores correspondem a formas livres, i.e., independentes do nome (e, em alguns casos, do verbo) que não obrigam ao estabelecimento de operações de concordância sintática (Dixon 1986:106; Corbett 1991:137). O inventário de classificadores numa dada língua pode ser muito variável, nem sempre sendo possível determinar, com exatidão, o número de classificadores nela existente. Com efeito, se há línguas que possuem, por exemplo, cerca de 100 classificadores, outras há que apresentam mais de 400 (Dixon 1986:106). Além disso, nem todos os nomes admitem a presença de classificadores, por um lado, e, por outro, há nomes que admitem a associação a vários tipos de classificadores (Dixon 1986; Aikhenvald 2000).

Grinevald (2000) propõe, entretanto, uma tipologia de classificadores que atende às suas propriedades morfossintáticas e semânticas. Deste modo, a autora identifica quatro principais tipos de classificadores: (i) numerais; (ii), nominais; (iii) genitivos; e (iv) verbais (cf. Figura 1.3). Apresenta-se, em seguida, uma breve explicitação das características de cada um dos tipos de classificadores propostos.



**Figura 1.3** – Tipos de classificadores (de acordo com proposta de Grinevald (2000))

Os classificadores numerais, característicos das línguas asiáticas, tais como o chinês e o japonês, são os mais frequentes e surgem associados a expressões numéricas ou de quantificação sendo, geralmente, obrigatórios dentro destas expressões. Além disso, podem ainda estar ligados a demonstrativos e a adjetivos, embora este último uso seja pouco frequente (Aikhenvald 2000:118; Grinevald 2000:63). A principal função desempenhada por este tipo de classificadores é a de organizar os nomes contáveis em diferentes classes semânticas (Contini-Morava & Kilarski 2013:273; Gil 2013).

Na língua chinesa<sup>16</sup>, por exemplo, os classificadores numerais ( $cl_{num}$ ) coocorrem, prototipicamente, com quantificadores (por ex., *yí* ‘um’, *sān* ‘três’, *zhěng* ‘todo’), podendo ainda coocorrer com demonstrativos (*nèi* ‘aquele’, *zhèi* ‘este’, por ex.) (Li & Thompson 1984:104; cf. Secção 1.4.5.2.), e surgem antes do nome, conforme ilustram os exemplos seguintes recolhidos de Li & Thompson (1984:105):

- (3) *sān ge (cl<sub>num</sub>) rén* ‘três pessoas’  
 (4) *nèi liù běn (cl<sub>num</sub>) shū* ‘aqueles seis livros’

Entre os classificadores numerais distinguem-se ainda dois subtipos semânticos: os classificadores numerais mensuráveis e os classificadores numerais de classe, embora nem sempre seja fácil estabelecer uma clara distinção entre eles<sup>17</sup> (Grinevald 2004:1020).

Os classificadores nominais<sup>18</sup>, típicos, sobretudo, nas línguas australianas (Aikhenvald 2000:82-96), correspondem a morfemas livres, intimamente correlacionados com o item, e não estão dependentes de outros elementos que compõem o sintagma nominal. Estes classificadores têm, essencialmente, uma função anafórica, servindo para retomar o nome com o qual se relacionam (Grinevald 2004:1021). No exemplo que a seguir se apresenta, retirado de Aikhenvald (2000:87), o nome *duguur* ‘casa’ surge numa pergunta e, na resposta, o classificador nominal *bulmba* ( $cl_{nom}$ ) é usado para retomar esse mesmo item nominal:

- (5) Pergunta: *nyundu duguur-um gada-any* ‘Vieste de casa?’  
 Resposta: *(yiyi) ngayu bulmba (cl<sub>nom</sub>)-m* ‘(Sim) Vim do acampamento.’

<sup>16</sup> Na Secção 1.4.5. deste trabalho, será feita uma descrição mais detalhada relativa às características do sistema de classificadores do chinês.

<sup>17</sup> De um modo geral, os classificadores numerais mensuráveis correspondem a unidades que especificam a quantidade relativa a nomes contáveis e não-contáveis sendo, por isso, equivalentes a expressões utilizadas em línguas que não possuem classificadores (Aikhenvald 2000:115; Grinevald 2000:64, 2004:1020) como, por exemplo, em português: “um *pedaço* de pão” ou “uma *mão-cheia* de laranjas”. Os classificadores de classe correspondem a partículas que especificam determinadas propriedades semânticas dos referentes nominais, identificando, por exemplo, os traços relativos à sua natureza [ $\pm$  animada] ou [ $\pm$  humana] ou a outras características relativas à forma e função das entidades e objetos designados pelo nome (Gil 2013). Por este motivo, este subtipo de classificadores é considerado, em grande parte dos casos, semanticamente redundante. Veja-se ainda em Aikhenvald (2000:286-293) uma análise exaustiva das propriedades semânticas dos classificadores numerais existentes em diversos idiomas.

<sup>18</sup> Nos trabalhos realizados no âmbito dos sistemas de classificadores, Grinevald (2000, 2004) constata que nem sempre este tipo de classificadores é considerado e tal acontece porque, frequentemente, se entende que os *classificadores nominais* correspondem a *classificadores numerais*.

Os genitivos, característicos de línguas da Micronésia, correspondem a morfemas livres ou presos utilizados em construções possessivas. A principal função deste tipo de classificadores é a de marcar o possuidor e classificar, em termos semânticos, o termo possuído. Na seguinte expressão de uma língua micronésia (Ponapean) — *kene-i mwenge* “a minha comida”, o classificador genitivo ‘*kene*’ classifica o termo possuído *mwenge* ‘comida’ como sendo ‘comestível’ (cf. Grinevald 2004:1022).

Os classificadores verbais, que se encontram em línguas indígenas norte-americanas, possuem uma forma distinta dos restantes classificadores referidos, correspondendo a afixos associados a uma forma verbal, e cuja função é a de especificar a função sintática do respetivo argumento nominal (veja-se Aikhenvald 2000; Grinevald 2000, 2004).

Como se depreende desta breve descrição, os diferentes tipos de classificadores distinguem-se, substancialmente, dos sistemas de género nominal e, por esse motivo, cada um destes sistemas de classificação se encontra associado a línguas tipologicamente distintas (Corbett 1991). Com efeito, idiomas de morfologia rica são, do ponto de vista histórico, propícios ao desenvolvimento de sistemas de classificação nominal baseados em género gramatical — como é o caso de muitos dos idiomas de matriz indo-europeia — enquanto que línguas do tipo isolante tendem a apresentar sistemas de classificadores, como, por exemplo, as línguas do sudoeste asiático (Dixon 1986:109)<sup>19</sup>.

No Quadro 1.1 apresenta-se uma síntese das principais propriedades dos sistemas de classificação nominal baseados em género gramatical e dos sistemas baseados em classificadores.

---

<sup>19</sup> Corbett (1991:137) constata que, se em grande parte dos casos, é possível afirmar que existe uma estreita relação entre o tipo de sistema de classificação nominal e a respetiva estrutura morfológica da língua, a verdade é que nem sempre esta relação se verifica. De facto, e embora seja um caso excecional, há um idioma, pertencente ao ramo das línguas da Nigéria, que exhibe, em simultâneo, um sistema de classificação nominal baseado em género e um sistema de classificadores possessivos.

Propriedades	Sistemas de classificação nominal	
	Gênero gramatical	Classificadores
Concordância sintática	Sim.	Não.
Escopo da marcação	Marcado no nome, nem sempre de modo explícito.	Independente do nome.
Crítérios de atribuição	Semânticos e formais (fonológicos e morfológicos).	Semânticos.
Número de valores	Relativamente pequeno, tratando-se de sistemas fechados.	Relativamente grande, tratando-se de sistemas abertos.
Variabilidade	Salvo raras exceções, os membros associados a cada valor são mutuamente exclusivos (ou seja, se masculino não pode ser feminino, por ex.)	Determinados nomes exigem um classificador específico, embora esta situação nem sempre se verifique, visto que há nomes: (i) que admitem mais do que um classificador; e (ii) que não exigem classificador.

**Quadro 1.1** – Síntese das propriedades relativas aos sistemas de classificação nominal baseados em gênero gramatical e em classificadores

### 1.2.3. Considerações gerais sobre a origem e formação dos sistemas de gênero gramatical

Uma das questões que mais tem intrigado os investigadores ao longo dos séculos prende-se, precisamente, com a origem da categoria de gênero gramatical. Em boa verdade, procura-se compreender a natureza do desenvolvimento da categoria de gênero atendendo, por um lado, à influência que a perceção dos falantes em relação ao mundo poderá ter tido para a classificação dos nomes em diferentes valores de gênero e, por outro, à dinâmica que se observa entre esta categoria gramatical e as demais formas de classificação nominal existentes nas línguas do mundo (Corbett 1991; Aikhenvald 2000; Foundalis 2002; Kilarski 2007).

Da análise das propostas, é possível identificar duas grandes linhas de pensamento: por um lado, admite-se que os sistemas de categorização nominal refletem o modo como os falantes conceptualizam na língua a realidade envolvente e, por outro, postula-se que os valores de gênero resultam, predominantemente, de fatores de natureza formal.

As primeiras reflexões em torno das especificidades do sistema de classificação nominal baseado em gênero do indo-europeu remontam ao século XIX. Autores como Humboldt e Grimm, citados por Wheeler (1899), consideravam que o sistema de classificação nominal baseado em gênero do indo-europeu resultou, sobretudo, de um processo de ‘personificação’ dos objetos por parte do homem primitivo. Como observa Wheeler (1899:529), nesta perspetiva é assumido que:

“[n]atural objects were viewed as persons, and, as sex afforded the most prominent characterization and classification of persons, objects were not only personified, but also freely sexualized. Grammatical gender as a classification of word-forms is then a conventionalized, crystallized of this primitive sexualization.”

Aikhenvald (2000:372) analisa o desenvolvimento dos sistemas de classificação nominal de género à luz do processo de gramaticalização, observando que, numa língua, um sistema de categorização lexical pode, progressivamente, tornar-se mais ‘gramatical’, ficando, do ponto de vista semântico, menos ‘transparente’.

Segundo Mendes (2013:249), a gramaticalização é muito frequente e produtiva nas línguas naturais e corresponde a:

“um processo progressivo e contínuo de passagem de certas unidades linguísticas de uma classe lexical para uma classe gramatical ou de uma classe menos gramatical para uma classe mais gramatical.”

Trata-se, portanto, de um processo no qual se assiste à reconfiguração dos elementos linguísticos que, gradualmente, vão perdendo as suas características semânticas plenas e passam a expressar determinadas noções de natureza gramatical. Dada a sua natureza gradual, é possível identificar diferentes estádios no decurso da gramaticalização (cf. Mendes 2013).

Segundo Brugmann (1897), o género resulta, sobretudo, de alterações de natureza formal. Apesar de este autor admitir que no indo-europeu existe uma relação, ainda que parcial, entre a categoria de género gramatical e o género natural, rejeita a possibilidade de os sistemas de género se terem desenvolvido por um processo de personificação da língua e, por conseguinte, considera que fatores de natureza formal, tais como a presença de determinados afixos, estão na base do desenvolvimento dos sistemas de género gramatical (Brugmann 1897:27-28).

Do mesmo modo, Ibrahim (1973) defende que, no primitivo indo-europeu, o aparecimento do género gramatical resultou, sobretudo, de alterações no plano formal, sendo que a classificação semântica dos itens com referentes sexuais, estabelecida em função do sexo dos respetivos referentes nominais, terá surgido posteriormente (1973:50). Este investigador considera que fatores extralinguísticos, i.e., de natureza social ou



psicológica, tenham tido pouca influência no aparecimento de valores de género gramatical, concluindo, assim, que no indo-europeu esta categoria “was an accident of linguistic history” (1973:50).

Relativamente ao desenvolvimento dos sistemas de classificação nominal baseados em género, Corbett (1991) começa por observar que, apesar de estes sistemas possuírem propriedades distintas dos sistemas de classificação nominal baseados em classificadores – na medida em que estes últimos são, em certa medida e do ponto de vista semântico, mais transparentes do que os de género –, há alguns pontos de contacto entre os dois tipos de sistemas. Para este investigador, os classificadores podem, em parte, estar correlacionados com o desenvolvimento dos sistemas de género gramatical, constituindo-se, assim, como “a source for gender agreement markers” (1991:137). Mais especificamente, Corbett, baseado em dados de natureza diacrónica, constata que o facto de em algumas línguas os classificadores passarem a ter uma função essencialmente anafórica e, por conseguinte, surgirem cada vez menos associados ao nome, pode ter potenciado o aparecimento de marcas formais de concordância de género (Corbett 1991:137-143; veja-se também Aikhenvald 2000:372). Esta perspetiva encontra-se intimamente relacionada com a proposta de Greenberg (1978) que, ao tratar a origem dos demonstrativos, observa que estes resultam, sobretudo, de classificadores (Greenberg 1978:78). Em última instância, Corbett (1991:312) considera que:

“the ultimate source of gender systems is nouns, more specifically nouns with classificatory possibilities such as ‘woman’, ‘man’, and ‘animal’”

Todavia, é também de salientar que algumas línguas foram, gradualmente, perdendo os traços formais de concordância sintática, como é observável, por exemplo, na língua inglesa<sup>20</sup>, com consequências para as propriedades dos sistemas de classificação nominal que exibem (Ibrahim 1973; Aikhenvald 2000; veja-se também a Secção 1.4.4. do presente trabalho). Aikhenvald (2000:380) associa o desaparecimento dos sistemas de classificação nominal baseados em valores de género a fenómenos de natureza formal, ao afirmar que:

---

<sup>20</sup> A este respeito, observa Aikhenvald (2000:380) que “[m]orphological causes may contribute to the restructuring and loss of gender classes. Old English lost gender classes partly due to the convergence of inflections (genitive singular *-es*, originally only masculine and neuter, spread to feminine; and nominative-accusative plural *-as*, originally only masculine, spread to feminine and neuter) (...)”.

“[n]oun classes can get reduced, and even lost due to the coalescence of paradigms for morphological and/or phonological reasons. This is typically found in fusional languages. Loss of noun classes is often accompanied by loss of number or case distinctions with which noun classes interacted.”

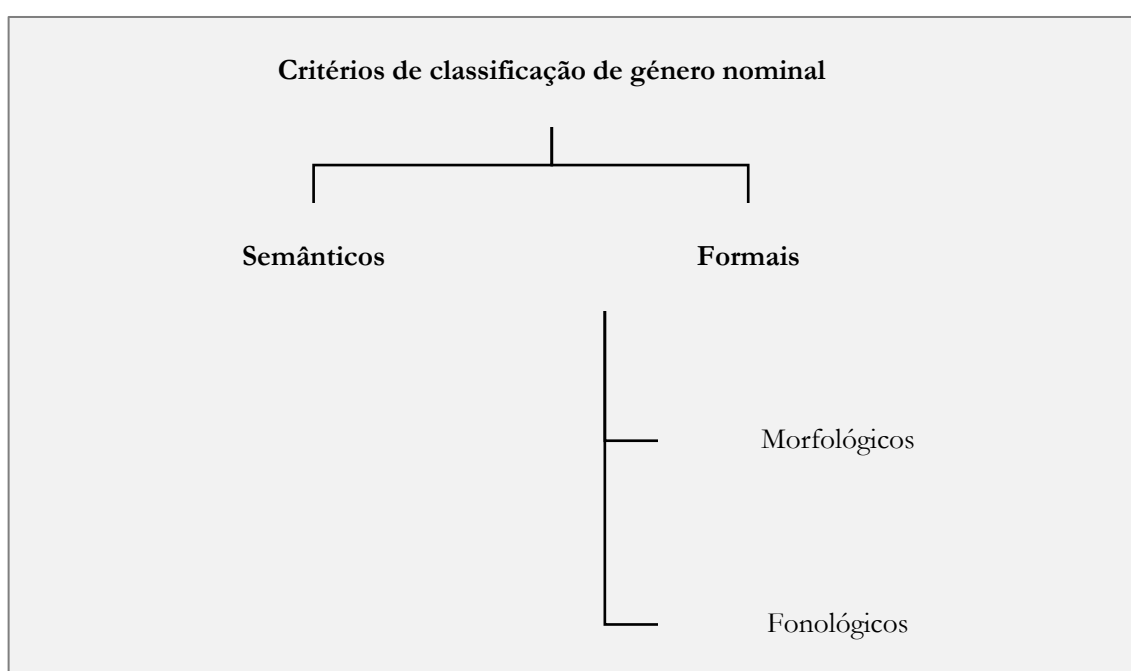
Neste processo, observa-se que, tipicamente, as marcas de concordância associadas aos itens lexicais, nomes e adjetivos, desaparecem primeiro do que as marcas de concordância associadas aos pronomes (Corbett 1991:143). Por esta razão, Aikhenvald (2000) considera que, apesar de muitas línguas de matriz indo-europeia terem perdido um sistema de género tripartido (masculino, feminino e neutro), a verdade é que em alguns idiomas, como o português, se observam vestígios de um primitivo valor de género neutro ao nível do sistema pronominal, como são exemplo as formas *isto*, *isso* e *aquilo*. Deste modo, conclui a investigadora (2000:399) que “agreement appears to be more stable on pronouns and demonstratives than on other targets.”

Greenberg (1978:79) distingue, entretanto, a origem dos sistemas de género da sua respetiva formação, assumindo que “[t]he way in which gender arises needs not to be the same as that by which the system can expand by the development of new genders”. Ou seja, é admitido que a expansão de um sistema de classificação nominal baseado em género não depende, necessariamente, do modo como essa categoria surgiu na língua, podendo antes depender de processos formais que resultam das propriedades do próprio idioma (Corbett 1991:313).

Em suma, e como se depreende desta breve exposição, o surgimento e expansão de valores de género podem depender de diferentes fatores, apesar de as perspetivas teóricas disponíveis assumirem, preferencialmente, que o género resulta, sobretudo, de processos de natureza formal.

#### 1.2.4. Critérios de atribuição de género nominal (Corbett 1991)

Assumindo, e como se viu, uma definição da categoria de género mais lata baseada no critério da concordância sintática obrigatória, Corbett (1991) defende que a atribuição dos valores de género aos nomes se pode basear, por um lado, em critérios semânticos, nos casos em que a associação de um valor de género está diretamente associada ao conteúdo referencial do nome e, por outro, em critérios de natureza formal que correspondem a regras do tipo morfológico e fonológico associáveis à atribuição dos valores de género dos itens nominais (cf. Figura 1.4).



**Figura 1.4** – Critérios de atribuição de valores de género nominal (baseado na proposta de Corbett (1991))

Em função destes critérios, o autor analisa um conjunto de duzentas línguas e como resultado dessa análise propõe uma tipologia, postulando a existência de:

- (i) sistemas de género estritamente lexicais, em que a atribuição dos valores de género é realizável através da aplicação exclusiva de critérios semânticos;
- (ii) sistemas de atribuição de género simultaneamente lexicais e formais; e
- (iii) a possível existência de sistemas de atribuição de género estritamente formais.

Uma vez que Corbett (1991) não encontrou nenhum sistema de género exclusivamente formal, considera que existe sempre, em maior ou menor grau, uma certa base semântica em todos os sistemas de classificação nominal baseados em género<sup>21</sup> (Corbett 1991:63, 2006b:751-752).

Em sistemas estritamente semânticos, o valor de género gramatical de um nome infere-se a partir das propriedades semânticas do respetivo referente. Este tipo ocorre, por exemplo, no tamil e no canarês, dois idiomas dravídicos falados no sul da Índia (Corbett 1991:8-9, 2006b:751, 2013b). Nestas línguas os nomes são divididos em dois grandes grupos, distinguindo-se aqueles que designam entidades dotadas de “racionalidade” — como os seres humanos e as divindades —, de outros considerados como nomes de referentes “não racionais”. Os itens nominais que integram o primeiro grupo são de género masculino ou feminino de acordo com o sexo do respetivo referente, enquanto os do segundo estão associados ao género neutro (Corbett 1991:9, 2013b).

O conceito de animacidade encontra-se intimamente relacionado com a noção de género, dado que, e como vimos, existe sempre um núcleo semântico nos sistemas de atribuição de valores de género nominal (Dahl 2000; Cruse 2000:273; Croft 1990:161). Na literatura é geralmente consensual a ideia de que as propriedades gramaticais de uma língua obedecem, e salvo raras exceções, a uma escala de animacidade (humano > animado > não animado) e, no que ao género diz respeito, Dahl (2000:101) observa que:

“in many Indo-European languages, humans and also, to a varying extent, higher animals, are assigned masculine or feminine gender on the basis of their sex, while inanimates and lower animals get their gender by lexeme-specific or formal criteria.”

No primitivo indo-europeu, a associação dos três valores de género (masculino, feminino e neutro) relacionava-se com o traço [ $\pm$ animado] do seu referente. O masculino e o feminino associavam-se ao carácter [+animado] dos itens nominais e o neutro ao traço [- animado]<sup>22</sup> (Meillet 1921: 211; Kurzová 1993:61; Gouveia 2004:449). O latim e o grego

---

<sup>21</sup> Também Aikhenvald (2000, 2004) defende que os sistemas de género possuem uma base semântica intrínseca, embora as línguas se distingam no modo como tais propriedades semânticas condicionam a marcação dos valores de género dos nomes.

<sup>22</sup> Quanto à forma morfológica das palavras, observa Meillet (1921:212) que no primitivo indo-europeu “la flexion de masculin ne se distinguait en rien de celle du féminin. Le caractère masculin ou féminin d'un substantif ne se reconnaissait donc en indo-européen qu'à la forme masculine ou féminine des adjectifs qui éventuellement s'y rapportaient. (...) Sans l'accord de l'adjectif, la distinction du masculin et du féminin n'existerait pas en indo-européen.”

preservaram os três valores de género que correspondiam às três principais classes nas quais os nomes se dividiam, embora se tenha perdido a estreita ligação entre os valores de género (masculino e feminino) e o carácter [+animado] do nome, por um lado, e entre o neutro e o traço [-animado], por outro (Meillet 1921).

Em latim, e independentemente do valor de género, os nomes (e adjetivos) distribuíam-se por cinco declinações distintas (Ernout 1914). Na 1ª declinação, de tema em *-a*, encontravam-se maioritariamente nomes femininos (*rosa*, *-æ* ‘rosa’), registando-se também algumas formas masculinas que se referiam, essencialmente, a profissões, como *agricola*, *-æ* ‘agricultor’ ou *poeta*, *-æ* ‘poeta’. Na 2ª declinação, de tema em *-o*, agrupavam-se nomes masculinos (*dominus*, *-i* ‘senhor’), bem como alguns femininos (*figus*, *-i* ‘figueira’) e neutros (*bellum*, *-i* ‘guerra’). Os nomes de tema em consoante e em *-i* constituíam a 3ª declinação, da qual faziam parte nomes masculinos (*rex*, *-is* ‘rei’), femininos (*soror*, *-is* ‘irmã’) e neutros (*vulnus*, *-is* ‘ferida’). Na 4ª declinação encontravam-se nomes de tema em *-u* que eram, sobretudo, de género masculino (*carrus*, *-us* ‘carro’), e, embora pouco frequentes, também faziam parte desta declinação formas do feminino (*manus*, *-us* ‘mão’) e do neutro (*genu*, *-us* ‘joelho’). Por fim, na 5ª declinação, de tema em *-e*, encontravam-se, quase de modo exclusivo, os femininos<sup>23</sup> (*res*, *rei* ‘coisa’), embora também se encontrasse o nome masculino *meridies*, *-ei* ‘meio-dia’ (Gouveia 2004:448).

Como se pode verificar, e apesar de a maioria dos nomes da 1ª e 5ª declinação que, nos termos apresentados, correspondem a outra classe temática, serem do género feminino, e os da 2ª e 4ª serem, essencialmente, masculinos, não existia um valor de género que estivesse, única e exclusivamente, associado a uma declinação.

É na transição do latim clássico para o latim vulgar que se verifica uma perda gradual do género neutro, justificável, em parte, pelo facto de existir uma forte correspondência entre os morfemas do neutro e do masculino<sup>24</sup> (Ernout 1914:3; Ibrahim 1973:85; Nunes 1975:220; Gouveia 2004:450). Na evolução para as línguas românicas, o valor de género neutro desaparece, dando lugar a sistemas em que prevalecem os valores

---

<sup>23</sup> É também da 5ª declinação o nome *dies*, *-ei* ‘dia’. No entanto, no singular este nome podia ter dois géneros, ou seja, tanto podia ser considerado como um nome masculino ou feminino. Já no plural, era sempre masculino (Gouveia 2004:448, n.21).

<sup>24</sup> Já no indo-europeu, a diferenciação dos valores de género neutro e masculino existia, apenas, nos casos nominativo, acusativo e vocativo, sendo que nos restantes casos as desinências das duas categorias eram praticamente idênticas (Ernout 1914:4).

de género masculino e feminino<sup>25</sup> (Barbosa 1822; Meillet 1921). As formas nominais provenientes do neutro foram integradas, maioritariamente, no masculino, embora haja algumas exceções, como é o caso dos nomes provenientes do plural neutro terminados em *-a* que se tornaram femininos, preservando ainda essa noção de coletivo ou de reunião<sup>26</sup> (Ibrahim 1973:86). A título de exemplo, refira-se o nome latino «*fructa*» (neutro plural de ‘*fructu-*’) que passou a ser feminino no português ‘(a) fruta’, no espanhol ‘(la) fruta’ e no italiano ‘(la) frutta’. José Joaquim Nunes (1975:223) verifica que, na transição do latim vulgar para o português, os neutros do singular terminados em *-o* “tomaram (...) o género masculino” e os provenientes do plural neutro terminados em *-a* passaram a ser do género feminino conservando, porém, “a ideia de *reunião, ajuntamento*, o que bem se nota em *lenha, braça, senha, folha, ova, fruta, boda, virilha*, etc.”. Decorrente do facto de estas formas apresentarem um singular terminado em *-o* que adotou o género masculino e uma forma do plural neutro que passou a ser feminina resultou “a mesma palavra achar-se representada por duas formas, uma masculina, correspondente ao singular, outra feminina, representante do plural (...) *lenho, lenha (...), braço, braça, ovo, ova, fruto, fruta*, etc.” (Nunes 1975:223). Por analogia, palavras que inicialmente não se enquadravam nestes casos adotam duas formas, uma masculina e outra feminina, que não apresentam grandes diferenças ao nível da sua significação, tais como *barco/barca; jarro/jarra; rio/ria*, etc. (Nunes 1975: 223).

Também há casos em que um mesmo neutro latino passou a masculino em algumas línguas românicas e feminino em outras, uma vez que já no latim se registava alguma ambiguidade no tratamento das formas correspondentes que, por vezes, tanto eram associadas ao neutro como ao masculino (Nunes 1975:220). O neutro «*mare* ‘mar’» tornou-se masculino em italiano (*il mare*) e em português (*o mar*), sendo feminino na língua francesa (*la mer*). Já em espanhol *mar* pode estar associado a dois valores de género<sup>27</sup> (Ambadiang 1999:4857).

---

<sup>25</sup> Sobre a passagem dos nomes do latim para o português, observa Barbosa (1822:127-128) que “[n]o uso presente de nossa Língua não ha nome algum substantivo de genero *incerto*, isto é, de que se possa usar arbitrariamente, ou com o genero masculino, ou com feminino. Todos são ou masculinos ou femininos.”

<sup>26</sup> O italiano preservou esta característica. De facto, alguns nomes masculinos terminados em *-o* (*braccio* ‘braço’; *osso* ‘osso’; *uovo* ‘ovo’) possuem uma forma de plural, dita ‘regular’, em *-i* (*bracci* ‘braços’, *ossi* ‘ossos’, *uovi* ‘ovos’) e uma outra, em *-a* (*braccia, ossa, uova*). Esta última forma, para além de ser feminina, acarreta ainda uma clara diferença semântica, i.e., abrange um sentido de ‘coletividade’ (Lausberg 1974:261). Veja-se, sobre este assunto, a Secção 1.4.2.1.

<sup>27</sup> De facto, em espanhol há algumas formas nominais [-animadas] ambíguas quanto ao género. Ambadiang (1999) sublinha que dentro deste tipo de itens distinguem-se (i) os nomes que apresentam

No que respeita à relação entre género gramatical dos nomes e respetivo género natural (sexo) dos referentes, se existem nomes de entidades sexuadas de género gramatical masculino por designarem entidades do sexo masculino (em português: *menino, rapaz*; em espanhol: *niño, muchacho*) e nomes femininos com referente de sexo feminino (em português: *menina, rapariga*; em espanhol: *niña, muchacha*), há também itens cujo valor de género não corresponde, necessariamente, ao sexo da entidade designada. Efetivamente, na língua portuguesa, e em outras línguas da mesma família, existem subconjuntos de nomes, os sobrecomuns e os epicenos, que apesar de terem o traço semântico [+sexuado], possuem um único valor de género gramatical independentemente do género natural da entidade designada (*a testemunha / o indivíduo / a serpente*) (cf. Secção 1.3.3.).

A arbitrariedade e assistemática da associação dos valores de género aos nomes é também visível pelo facto de haver, em línguas da mesma família, formas nominais com o mesmo referente, mas com valores de género distintos. Por exemplo, em português o nome *árvore* é feminino, não o sendo em italiano (*albero*), em espanhol (*árbol*) e em francês (*arbre*). Já *leite* é masculino em português, em italiano (*latte*) e em francês (*lait*), embora seja feminino em espanhol (*leche*) (Villalva 1994:236-237).

A não correspondência entre género natural e género gramatical também se evidencia pelo facto de o número de valores de género ser muito variável de língua para língua. A propósito da evolução dos sistemas de género nas línguas de matriz indo-europeia, Ibrahim (1973:70) constata que se em algumas esta categoria gramatical se extinguiu por completo (como, por exemplo, na língua persa), houve idiomas que preservaram alguns traços desta categoria, sobretudo ao nível dos pronomes (como o inglês). Entre as línguas que mantiveram o sistema de género, observa-se que, para além das que conservaram o mesmo número de valores, i.e., masculino, feminino e neutro, como o alemão<sup>28</sup>, outras apresentam sistemas de género bipartidos (masculino e feminino). Também Corbett refere que o número de valores de género numa língua pode ser muito diverso<sup>29</sup> (Corbett 1991:5, 2013a; Kibort & Corbett 2008:5).

---

um mesmo significado quer estejam associados ao masculino quer estejam associados ao feminino; e (ii) os nomes em que se verifica uma alteração ao nível do significado tendo em conta o valor de género que lhes está associado. Há ainda formas nominais que admitem os dois valores de género no singular, embora no plural surjam associadas a um único valor (1999:4857). Veja-se ainda a Secção 1.4.1.

<sup>28</sup> Ibrahim (1973) salienta que apesar de haver idiomas que preservaram os três valores de género, a verdade é que tais sistemas apresentam configurações distintas do sistema de atribuição de género oriundo do indo-europeu.

<sup>29</sup> Após analisar cerca de 258 idiomas, Corbett (2013a) constatou que entre as que possuíam a categoria de género, a que correspondem cerca de 112 línguas, 50 exibiam sistemas de género bipartidos, havendo

As evidências assinaladas comprovam que, para além de o género gramatical não corresponder de modo sistemático ao género natural, a concordância é, e como vimos, um critério importante para a definição dos sistemas de género (Hockett 1958; Corbett 1991).

Para além de critérios semânticos, Corbett (1991) distingue ainda um conjunto de princípios de natureza formal que atuam na atribuição dos valores de género nominal. Segundo o investigador, tais princípios são constituídos por regras de tipo fonológico e morfológico, embora nem sempre seja possível distingui-las com clareza.

Nos sistemas em que atuam regras fonológicas, verifica-se que a distinção de género se pode estabelecer da seguinte forma:

- (i) “nomes cujo segmento final é  $\alpha$  possuem  $\beta$  como valor de género”

Nestes casos, existe uma única forma morfológica do nome a que se associa determinado valor de género. Além disso, também a posição da sílaba tónica no nome pode associar-se a um determinado valor. Segundo Corbett (1991, 2006b), na língua *qafare* (da família das línguas afro-asiáticas) a associação dos valores de género aos nomes com o traço semântico [-sexuado] baseia-se em princípios de natureza fonológica, uma vez que todos os itens terminados em vogal tónica são femininos e os restantes são masculinos (Corbett 1991:51). Também numa larga maioria de nomes do francês a atribuição dos valores de género assenta, sobretudo, em princípios fonológicos. Corbett (2006b:752) constatou que, num subconjunto de 938 nomes terminados em [ɛ̃], cerca de 99% dos itens são masculinos «(*le*) *pain* [pɛ̃] ‘(o) pão’» (Corbett 2006:752).

Os critérios de natureza morfológica atuam nas situações em que, para a atribuição dos valores de género, é necessária a referência a mais do que uma forma morfológica do nome. De acordo com Corbett (1991), o sistema de género gramatical da língua russa ilustra um exemplo claro em que atuam regras morfológicas, pois existe uma estreita relação entre a declinação do nome e os respetivos valores de género. Por conseguinte, o autor propõe a regra:

- (ii) “nomes que pertencem à declinação  $\alpha$  possuem  $\beta$  como valor de género”

---

26 idiomas com três valores de género distintos. Já em 24 línguas encontraram-se sistemas com mais de quatro valores. Veja-se ainda a Figura 1.1 no presente trabalho.



Em alemão, por exemplo, para a associação dos valores de género aos nomes aplicam-se critérios de natureza morfológica, já que alguns sufixos derivacionais condicionam o valor de género do produto nominal que originam. O sufixo ‘-lein’ forma diminutivos e quando adicionado a uma base nominal, como, por exemplo, o nome masculino «(der) Tisch ‘(a) mesa’» produz um diminutivo de género neutro «(das) Tischlein ‘(a) mesinha’» (Donaldson 2007:41; cf. Secção 1.4.3.). Na língua portuguesa verifica-se igualmente que os sufixos derivacionais determinam o valor de género dos produtos nominais (Villalva 2003; Rio-Torto *et al.* 2016) (cf. Secção 1.3.3.).

Corbett (1991) refere ainda que a aplicação de critérios morfológicos pode, em determinados casos, estar condicionada por critérios semânticos. Em russo, por exemplo, a associação do valor de género gramatical de alguns itens nominais estabelece-se em função do sexo do seu referente, independentemente da declinação a que pertencem<sup>30</sup>.

Referidas algumas das particularidades relativas aos critérios de atribuição dos valores de género nominal, em função da proposta de Corbett (1991), na próxima Secção procede-se à descrição e análise do sistema de género nominal do português, a LA dos aprendentes. Posteriormente, na Secção 1.4., apresentam-se os sistemas de classificação nominal das línguas maternas dos informantes, autores das produções escritas que tomámos como base empírica do presente estudo.

---

<sup>30</sup> A título de exemplo, Corbett (1991:37) refere que as formas nominais *djadja* ‘tio’ e *deduska* ‘avô’ são masculinas, apesar de pertencerem à declinação de nomes femininos.



## 1.3. O género nominal em português

### 1.3.1. A estrutura formal dos nomes em português

Na língua portuguesa, todos os nomes e alguns pronomes são especificados quanto ao género gramatical, existindo dois valores em oposição: masculino e feminino. Além do item nominal, a atribuição de género afeta outras classes de palavras pela necessidade de com ele estabelecerem relações de concordância sintática. Distinguem-se quanto ao valor de género as palavras concordantes localizadas à esquerda e à direita do nome, ou seja, especificadores (determinantes e quantificadores) e modificadores (adjetivos). Os elementos localizados à esquerda do nome compõem a estrutura funcional do sintagma nominal (Brito 2003:345).

Verifica-se que, no interior dos sintagmas nominais, a concordância se estabelece entre:

- (i) o núcleo nominal e as expressões de determinação e de quantificação que o antecedem (especificadores) – os artigos definidos e indefinidos, os determinantes demonstrativos e possessivos, e os quantificadores; e
- (ii) o núcleo nominal e os adjetivos integrados em sintagmas adjetivais que funcionam como seus modificadores ou elementos apositivos (Brito 2003:328).

Embora seja uma propriedade intrínseca dos itens nominais, nem sempre é possível inferir os valores de género tendo em conta as suas características semânticas e fonomorfológicas. No que respeita à estrutura formal do nome, tipicamente associam-se os valores de género masculino e feminino às vogais átonas *-o* e *-a* que surgem à direita do radical nominal, simples ou derivado. Porém, não se pode considerar que estas vogais correspondam a morfemas de género, uma vez que encontramos nomes masculinos terminados em *-a* (*o dia*) e femininos terminados em *-o* (*a tribo*) (Rio-Torto & Rodrigues 2016:158). Além disso, existem itens nominais com diferentes terminações que tanto podem ser de género masculino como feminino (por ex., *o sofá, a pá, a maré, o banzé, o coração, a canção, a manhã, o papel, a cascavel, o animal, a catedral*). Por conseguinte, a principal função das vogais *-o* e *-a* é a de classificar os nomes, à semelhança do que acontece com as vogais temáticas dos verbos, e são designadas *índices temáticos* (Villalva 2003) ou *vogais temáticas* (Câmara Jr. 1994).

Considere-se, em seguida, as propostas de classificação dos nomes em português tendo em conta as classes temáticas a que tais itens podem pertencer. Para Mattoso Câmara Jr. (1994), e à exceção da vogal *-a* presente em nomes com referentes sexuados como *menina* que corresponde, na perspectiva do autor, a um morfema de género feminino a que se opõe uma forma do masculino *menino* (cf. 1.3.2.), as vogais átonas finais dos nomes substantivos correspondem a *vogais temáticas*. Logo, em função da classe temática a que pertencem, os itens classificam-se como:

- (i) nomes de tema em *-o* (*dedo, martelo*);
- (ii) nomes de tema em *-a* (*mapa, garrafa*);
- (iii) nomes de tema em *-e* (*pente*) ou *-e* teórico (*cantor*)<sup>31</sup>; e
- (iv) nomes atemáticos (*coração*).

Alina Villalva (2003) propõe a designação alternativa de *índices temáticos*, sendo que os nomes se encontram distribuídos por cinco classes temáticas básicas, distinguindo-se os:

- (i) itens nominais de tema em *-o* (*aluno, livro*);
- (ii) nomes de tema em *-a* (*mapa, mosca*);
- (iii) nomes de tema em *-e* (*abutre, dente*);
- (iv) nomes de tema  $\emptyset$  (um morfema zero, nos nomes terminados em consoante como, por ex., *cantor, mar*<sup>32</sup>); e
- (v) formas nominais atemáticas, i.e., as que terminam em vogal tónica (*chaminé*), e em vogal ou ditongo nasais (*manã, diapasão*).

Em função destas cinco classes básicas, a autora chega a estabelecer um conjunto de 23 classes temáticas, ao aplicar outros critérios, tais como a oposição de formas variáveis e invariáveis, ou seja, os nomes que admitem contrastes de género gramatical e os que não admitem essa variação (cf. Villalva 2003:924).

---

<sup>31</sup> Mattoso Câmara Jr. classifica os nomes terminados em consoante como itens com *-e* teórico porque, embora esta vogal não surja na forma do singular do nome (por exemplo, *leitor*), vai aparecer no plural (*leitores*) (Câmara Jr. 1994:87).

<sup>32</sup> Estes casos correspondem aos que Mattoso Câmara Jr. identifica como sendo de tema em *e* teórico.

Assim, e tendo em conta as duas propostas apresentadas, veja-se o Quadro 1.2 em que se apresentam os nomes, quer os que designam entidades sexuadas, como os que possuem o traço semântico [-sexuado], distribuídos em função de índice temático / vogal temática e respetivo valor de género:

Vogal Temática Índice Temático	Valor de género		
	Masculino	Feminino	Comum de dois (masculino e feminino)
-o	<i>o dedo, o menino, o caderno, o mundo, o campo, o ministro, o apartamento, o carro, o rádio (aparelho eletrónico)</i>	<i>a libido, a tribo, a rádio (a estação)</i>	<i>o/a modelo, o/a piloto<sup>33</sup></i>
-a	<i>o mapa, o dia, o programa, o quilograma, o cinema</i>	<i>a menina, a rapariga, a mochila, a bicicleta, a telenovela, a revista, a notícia</i>	<i>o/a motorista, o/a jornalista, o/a dentista, o/a economista</i>
-e	<i>o dente, o pente, o ventre, o ambiente</i>	<i>a ponte, a mente, a crise</i>	<i>o/a estudante, o/a ajudante, o/a presidente, o/a mestre</i>
	<i>-e (teórico)/ Ø</i>	<i>o professor, o tradutor, o cantor, o leitor, o rapaz</i>	<i>a imperatriz, a codorniz, a avestruz</i>
atemáticos	<i>o avião, o avó, o café, o galão</i>	<i>a manhã, a canção, a avó, a paisagem</i>	<i>o/a personagem</i>

**Quadro 1.2** – Distribuição dos nomes em português em função de vogal temática / índice temático e valor de género gramatical

Como se pode verificar a partir dos dados que constam no Quadro 1.2, não é possível inferir o valor de género de um nome tendo em conta o respetivo índice temático<sup>34</sup>, visto que, independentemente da classe temática a que pertençam os itens nominais, estes podem ser de género masculino ou feminino e até admitir os dois valores de género. Por conseguinte, e tal como observam Choupina *et al.* (2016:210), em português:

“classe e género não são traços do mesmo tipo: os traços de classe permitem inserir um Item de Vocabulário numa classe de nomes em que todos os elementos partilham o mesmo funcionamento morfológico; os traços de género promovem a organização dos nomes em grupos ou classes que determinam diferentes padrões de concordância sintática.”

<sup>33</sup> Nem todos os dicionários do português consideram *piloto* um nome comum de dois géneros. Tendo em conta a aceção “aquele que regula a direção de uma embarcação, aeronave ou automóvel de corrida”, o Dicionário da Língua Portuguesa da Porto Editora *online* (Infopedia) categoriza ‘piloto’ como um ‘nome de 2 géneros’, ao passo que no Dicionário Priberam, este nome é classificado como ‘substantivo masculino’, i.e., com um único valor de género gramatical, a que se opõe o feminino *pilota*.

<sup>34</sup>A este respeito observa Mota (2016a:172) que o valor de género, masculino ou feminino, encontra-se especificado na entrada lexical da maioria dos nomes, e, por conseguinte, “não tem relação constante com a forma do índice temático”.

### 1.3.2. Género: uma categoria de natureza flexional?

Na tradição gramatical portuguesa, o género gramatical tem sido maioritariamente tratado como uma categoria relativa à flexão nominal (Vilela 1999; Cunha & Cintra 2005), apesar de alguns autores negarem a existência de flexão de género dentro da classe nominal (Carvalho 1979; Villalva 1994; Mota 2016a). Por conseguinte, para se comprovar que o género é uma categoria flexional em português é necessário verificar se esta obedece a determinadas características dos processos flexionais.

A flexão corresponde a um processo de natureza morfológica, relativa ao domínio da palavra, que se caracteriza pela existência de paradigmas coesos, de variações de palavra que se repetem quase sempre do mesmo modo, ainda que possam existir algumas exceções que evidenciam uma pequena variação alomórfica (Mota 2016a). Deste modo, poder-se-á admitir que o processo de pluralização das formas nominais, que consiste em agregar o morfema de plural *-s* ao tema nominal (um constituinte formado pelo radical e índice temático), se define claramente como um processo flexional.

Assim, se atendermos a pares como *menino / menina*, *doutor / doutora*, poder-se-ia admitir que existe um mecanismo flexional em português, tal como o defende Câmara Jr. (1966, 1994). Assumindo que as gramáticas tradicionais apresentam de forma confusa e incoerente a questão da natureza flexional do género, o autor (1994:88) defende que o género gramatical deve ser considerado como uma categoria de natureza formal e não semântica, existindo em português:

- (i) nomes substantivos com dois valores de género, com flexão redundante, visto que nestes casos o género é marcado, quer no item nominal, quer no determinante que o antecede: “*o gato / a gata*”;
- (ii) nomes substantivos que possuem dois valores de género sem flexão: “*o/a jornalista*”; e
- (iii) nomes substantivos com um único valor de género: “*o dente / a composição / a mensagem*”.

Tendo em conta esta proposta, a flexão de género restringe-se a pares de palavras como *menino/menina*; *gato/gata*; *mestre/mestra* porque, nestes casos, verifica-se que a uma forma nominal masculina se opõe uma outra feminina do mesmo nome, i.e., as formas representam versões diferenciadas quanto ao género do mesmo item nominal. Assim sendo, Câmara Jr. defende que a marcação do feminino se caracteriza pela presença de um

morfema de género feminino — fonologicamente /a/ — registado apenas nos nomes de entidades sexuadas que possuem um correspondente masculino, tais como *menino / menina*.

O investigador admite, assim, que o masculino constitui a forma não-marcada do nome, sendo que o feminino funciona como “uma particularização mórfico-semântica do masculino, uma forma marcada pela adunção da desinência /a/”, concluindo, então, que se trata “de uma oposição privativa, em que uma forma marcada pela desinência de feminino se afirma perante uma não-marcada, ou de desinência (Ø) para o masculino” (Câmara 1966:3). Logo, para o gramático, a categoria de género corresponde à organização dos nomes em diferentes classes, em que, no caso dos nomes com referentes animados o masculino se define como “uma forma geral, não-marcada, e o feminino indica uma especialização qualquer” (Câmara 1994:88). Câmara Jr. considera ainda que o morfema do feminino (/a/) não se encontra nos nomes comuns de dois géneros, i.e., os que possuem um diferente valor de género em função do contexto sintático em que ocorrem (*o/a motorista; o/a artista; o/a colega*), nem em substantivos de género único (*a cadeira / a mala*), porque estes não possuem uma forma masculina oposta (Câmara 1966).

Porém, se se tiver em conta a totalidade de nomes que compõem o léxico da língua portuguesa, constata-se que para além dos itens masculinos e femininos com os índices temáticos *-o* e *-a*, respetivamente, há outros que admitem diversas terminações, podendo ser de género masculino como feminino e comuns de dois (cf. Secção 1.3.1., Quadro 1.2.). Além disso, e apesar de todos os nomes possuírem um valor de género, nem todos admitem contrastes desta categoria gramatical, mesmo os terminados em *-o/-a* como, por exemplo, *o dedo / \* a deda; a mala / \* o malo*. Portanto, a variação em género não é obrigatória e afeta somente os nomes com o traço semântico [+sexuado] e, mesmo nestes, tais contrastes podem ser materializados sem o recurso à oposição de índices temáticos *-o/-a* (Villalva 2008; Mota 2016b:154).

Como vimos a propósito da descrição de género gramatical (cf. Secção 1.2.), há autores que assumem que esta categoria pode corresponder a um traço inerente ou opcional nos itens nominais. Segundo alguns autores, tais como, Augusto & Corrêa 2005, Costa *et al.* 2015 e Choupina *et al.* 2016, é possível admitir que há no português:

- (i) nomes de ‘gênero intrínseco’, cujo valor se encontra especificado no radical e é, portanto, atribuído no Léxico, como, por exemplo, *o sapato, a mesa, o coração, a canção, o elefante, a cobra, a vítima, o dia, a tribo*; e
- (ii) nomes de ‘gênero opcional’ ou ‘sintático’, cuja especificação de gênero não se encontra no radical e verifica-se apenas no plano sintático, sendo que tal especificação se pode manifestar de três formas distintas: pelo índice temático (*o menino, a menina*), pela adição de sufixos derivacionais (por ex., o sufixo *-al* origina nomes masculinos, como *o portal, o pantanal*) e pela coocorrência de palavras concordantes (especificadores e/ou modificadores) (*o estudante, a estudante*) (Costa *et al.* 2015: 332).

À luz desta proposta, os nomes de gênero intrínseco, com referentes animados e não animados, apresentam um valor de gênero único (masculino ou feminino) e não admitem a realização de contrastes de gênero. No caso dos nomes de gênero sintático, este traço atua sobre: (i) bases complexas com um gênero único e sem a possibilidade de participar em contrastes de gênero (*a modernização, a competição, o azedume*); e (iii) bases simples, nas quais o radical não se encontra especificado quanto ao valor gênero e admitem, por esse motivo, um dos dois valores disponíveis, sendo certo que a especificação dos valores de gênero das bases se manifesta, e como veremos, por mecanismos de natureza diversa (*o menino, a menina, o colega, a colega*) (Choupina, Baptista & Costa 2014; Costa *et al.* 2015:332).

Em boa verdade, no português os contrastes de gênero masculino/feminino podem manifestar-se através da oposição de constituintes temáticos, com correlato em oposições de sexo do referente de nomes com o traço semântico [+sexuado], sendo que, em grande parte dos casos, essa oposição se traduz numa oposição entre os índices *-o* e *-a* (embora também possam estar envolvidos outros índices temáticos, cf. (6a), ou pela oposição de radicais (cf. (6b)), sem se verificar, contudo, uma relação de natureza formal entre os nomes:

- (6) a. *menino* ≠ *menina*  
*professor* ≠ *professora*  
*mestre* ≠ *mestra*
- b. *homem* ≠ *mulher*  
*boi* ≠ *vaca*



Para além da marcação de classe temática, encontram-se formas de variação de género por alternância fonológica:

(7) *avô ≠ avó*

Também é possível encontrar contrastes de natureza sintática, i.e., a explicitação dos valores de género nominal associados a cada item é feita através da coocorrência de palavras concordantes (Mota 2016b:156). Os nomes comuns de dois apresentam uma única forma lexical, sendo essa forma ambígua quanto ao valor de género. Tal ambiguidade anula-se através do contexto sintático em que os nomes ocorrem:

(8) *o jornalista ≠ a jornalista*  
*este estudante ≠ esta estudante*  
*artista talentoso ≠ artista talentosa*

O modo de assinalar, linguisticamente, a oposição entre nomes que designam entidades do sexo masculino e do sexo feminino pode ainda materializar-se através de processos morfológicos como a derivação<sup>35</sup> (cf. (9a)) ou a composição (cf. (9b)).

(9) a. *duque ≠ duquesa*  
*galo ≠ galinha*  
*abade ≠ abadessa*

b. *o tigre-macho — o tigre-fêmea*  
*a baleia-macho — a baleia-fêmea*

Nos casos em que o contraste é assegurado por um processo derivacional, poder-se-á admitir que o feminino constitui uma especialização do sentido do nome masculino, sendo que, por exemplo, o derivado *duquesa* representa a mulher do *duque*.

Mattoso Câmara Jr. (1966) sugere ainda que existe variação de género em pares de palavras como *jarro/jarra*; *barco/barca*; *fruto/fruta* em que, para além da oposição masculino/feminino se processar por um contraste de vogais temáticas *-o/-a*, essa alteração do valor de género implica, consequentemente, uma alteração semântica do nome o que evidencia, assim, um comportamento de natureza derivacional. Também Bechara (1999) considera que, mesmo entre os defensores do processo de flexão em *barco/barca* se admite

---

<sup>35</sup> Segundo Villalva (2008:159) os contrastes de género estabelecidos através da derivação são pouco frequentes.

que a oposição masculino/feminino faz alusão a aspetos da realidade diferentes da diversidade de sexo, sendo o masculino a forma geral, não-marcada sintaticamente, enquanto o feminino expressa uma certa especialização<sup>36</sup>. Porém, a aplicação semântica faz com que estes pares de itens não devam ser consideradas formas geradas por um processo flexional, mas palavras diferentes geradas por um processo de derivação<sup>37</sup> (Bechara 1999).

Quanto à materialização dos contrastes por composição, constata-se que esta é exclusiva dos nomes epicenos, i.e., formas nominais que referem animais com um único valor de género gramatical independentemente do sexo do referente. Para estabelecer a oposição de género acrescenta-se à base nominal (*tigre*) a forma *-macho* (*tigre-macho*) ou *-fêmea* (*tigre-fêmea*) criando-se um composto sintático por subordinação que especifica, não o valor de género gramatical do nome, visto que esse valor se mantém inalterado (*o tigre-macho/ o tigre-fêmea*), mas o sexo da entidade referida (Villalva 2008; Mota 2016b:155).

Por fim, refira-se ainda que nem todos os nomes com referentes sexuados admitem contrastes de género. Há, de facto, um subconjunto de itens, os sobrecomuns, que possuem um único valor de género gramatical independentemente do sexo do referente<sup>38</sup> (cf. (10)):

- (10)     *o indivíduo*  
          *a pessoa*  
          *a testemunha*

---

<sup>36</sup> Efetivamente, encontramos na língua portuguesa, sobretudo em registo popular, certos vocábulos cuja distinção de género se baseia nas diferentes dimensões dos objetos designados como é o caso, por exemplo, dos pares de palavras *panelo/panela* ou *cesto/cesta* em que a forma do masculino designa, geralmente, o objeto mais pequeno (*panelo, cesto*) por oposição à forma do feminino que se refere ao mesmo tipo de objeto, mas de maiores proporções (*panela, cesta*). Além do português, há outras línguas que registam situações semelhantes relacionadas com a oposição de formas masculinas e femininas. Em italiano e em espanhol, por exemplo, o contraste masculino/feminino permite distinguir o nome da árvore do respetivo fruto. Deste modo, temos em espanhol: *aceituno*<sub>masculino</sub> ‘oliveira’ vs. *aceituna*<sub>feminino</sub> ‘azeitona’ e em italiano *melo*<sub>masculino</sub> ‘macieira’ vs. *mela*<sub>feminino</sub> ‘maçã’. Sobre estas particularidades associadas à variação de género veja-se Contini-Morava & Kilarski (2013:271-272).

<sup>37</sup> Esta posição está, assim, em linha com o que é afirmado por Rodrigues (2016). Segundo a autora (2016:37-38): “[a] morfologia flexional trata da constituição interna da mesma palavra, estudando as suas variações formais (*gato/gatos; adorou/adorou/adoraram*). A morfologia derivacional trata da constituição interna de palavras diferentes, estudando as variações formais e semânticas que permitem construir palavras a partir de outras, como *avaliação* a partir de *avaliar* ou como *contentamento* a partir de *contentar* e este verbo a partir do adjetivo *contente*.”

<sup>38</sup> Para Mota (2016b:155), nestes casos “fica claro que o género gramatical é fixo, inerente, idiossincrático, independente da classe temática e não sensível ao sexo dos referentes.”

Em suma, o sistema de atribuição de género nominal do português possui características que contrariam a perspectiva de que o género corresponde a uma propriedade da flexão nesta língua (Choupina, Baptista & Costa 2014; Mota 2016b)<sup>39</sup>. Deste modo, neste trabalho subscreve-se a posição defendida por Villalva (1994:233) que, ao constatar que a variação em género não é obrigatória e é realizável através de mecanismos linguísticos diversos, conclui que o género nominal em português:

“não é uma categoria de flexão, mas sim uma categoria morfo-sintática cuja especificação é lexicalmente determinada ou resultante de um processo morfológico não-flexional.”

### 1.3.3. Critérios de atribuição de género nominal do português

Ao descrever os diferentes sistemas de género das línguas, Corbett (1991) identificou um conjunto de critérios que atuam na atribuição dos valores de género aos nomes (veja-se, neste Capítulo, a Secção 1.2.3.) e é com base na proposta deste autor que se apresenta o sistema de atribuição de género nominal do português. Em primeiro lugar, referir-se-ão os nomes de entidades sexuadas, passando-se posteriormente a considerar os critérios relativos à atribuição do género gramatical de nomes de referentes não-sexuados.

Relativamente aos nomes com o traço semântico [+sexuado] constata-se que, em muitos dos casos, o valor de género gramatical corresponde ao género natural da entidade designada. Assim, temos formas nominais masculinas que se referem a seres do sexo masculino (cf. *homem, menino*) e formas femininas com referentes do sexo feminino (cf. *mulher, menina*). Neste tipo de itens, a atribuição de um valor de género caracteriza-se pela aplicação do critério semântico que pode, por vezes, ser coincidente com o critério formal. Por exemplo, o nome *gato* refere-se a uma entidade masculina e possui o índice temático *-o* que, tipicamente, está associado ao género masculino.

Embora se verifique a atuação do critério semântico nos nomes de entidades sexuadas, há a registar algumas restrições. Por um lado, e como já se viu, encontram-se em português alguns itens nominais, os sobrecomuns e os epicenos, que, apesar de terem referentes sexuados, possuem um valor de género gramatical único e independente do sexo do seu referente. Nestes itens, defende-se, como vimos, que o valor de género está

---

<sup>39</sup> No entanto, em documentos oficiais, nomeadamente que orientam o ensino formal, do português insiste-se na ideia de que o género é parte integrante dos processos flexionais. Veja-se a este respeito, os trabalhos de Baptista *et al.* 2013; Choupina, Baptista & Costa 2014 e de Costa *et al.* 2015.

especificado no radical nominal, sendo atribuído no Léxico (Costa *et al.* 2015; Choupina *et al.* 2016). Por outro, e apesar de muito pouco frequentes, existem determinadas formas nominais masculinas que se referem a entidades do sexo feminino (*mulherão*) e formas nominais femininas que referem seres do sexo masculino (*bicha*) (cf. Villalva 2008:100).

No que respeita aos nomes não-sexuados, a associação dos valores de género, por não se relacionar com um conteúdo semântico específico, é, tendo em conta os princípios de natureza semântica, verdadeiramente aleatória e imprevisível<sup>40</sup>. Nestes itens, e de acordo com a proposta de Corbett (1991), é possível, em alguns casos, inferir o valor de género atendendo à própria estrutura formal do nome.

Tal como se referiu anteriormente, existe uma forte tendência para se associar os índices temáticos (-o e -a) aos valores de género masculino e feminino, respetivamente. Deste modo, constata-se que temos em português:

- (i) nomes não-sexuados em que se observa uma correlação entre valores de género (masculino / feminino) e índices temáticos -o e -a, respetivamente; e
- (ii) nomes não-sexuados em que não se verifica uma correlação entre valores de género gramatical e índices temáticos -o e -a (*o programa; a moto*).

Com o objetivo de apurar a distribuição relativa dos nomes de género masculino e feminino por classe temática, Ferreira (2011) procedeu à análise dos índices de frequência dos lemas recenseados no *corpus* de frequências lexicais do português contemporâneo (Léxico Multifuncional Computorizado do Português Contemporâneo [CORLEX]). Procedeu-se, em concreto, à contagem das frequências de nomes de género masculino e feminino, formas com referentes sexuados e não sexuados, distinguindo os que possuíam os índices temáticos -o e -a, e os que pertenciam a outras classes temáticas (Ferreira 2011:33-34). No Quadro 1.3 regista-se a distribuição das frequências lexicais em função das classes temáticas e do género dos nomes, bem como os respetivos valores percentuais assim apurados:

---

<sup>40</sup> Refira-se, no entanto, que há um determinado subconjunto de formas nominais com referentes não-animados nos quais a associação dos valores de género parece estar subordinada a uma certa ‘organização semântica’ dos itens. Por exemplo, verifica-se que os dias da semana são todos de género feminino (*a segunda-feira, a terça-feira, a quarta-feira, a quinta-feira, a sexta-feira*) ao passo que os nomes relativos aos dias do fim de semana são todos de género masculino (*o sábado, o domingo*).

Classes temáticas	Valor de género gramatical	# Frequência de nomes	% Frequência de nomes
Índice Temático: <i>-o</i>	masculino	1 156 695	32
	feminino	1 613	0,04
	comum de dois (masculino e feminino)	1 127	0,03
Índice Temático: <i>-a</i>	masculino	56 362	1,6
	feminino	971 863	26,9
	comum de dois (masculino e feminino)	25 982	0,7
Outras classes temáticas	masculino	629 239	17,4
	feminino	706 851	19,6
	comum de dois	61 466	1,7
Σ		3 611 198	100

**Quadro 1.3** – Frequência e respetivo valor percentual dos lemas recensados na CORLEX em função do género gramatical e da respetiva classe temática dos nomes simples e derivados (cf. Ferreira 2011)

A partir dos resultados obtidos neste estudo, é possível verificar que a percentagem da frequência de nomes masculinos com índice temático *-o* é de, aproximadamente, 32%, sendo que apenas cerca de 1,6% das formas nominais masculinas possui índice temático *-a*. Nas restantes classes temáticas, a percentagem de nomes de género masculino apurada não ultrapassa os 17%. Quanto aos nomes de género feminino, constata-se que a percentagem da frequência de itens com índice temático *-a* atinge quase os 27%, valor muito acima do que se regista nas formas femininas de tema em *-o* (cerca de 0,04%). Nas outras classes temáticas, a percentagem de nomes femininos é de, sensivelmente, 20%. Relativamente aos nomes comuns de dois géneros, a frequência de itens com índice temático *-o* é de 0,03% e com índice temático *-a* é de 0,7% (cf. Ferreira 2011:37).

Portanto, estes dados demonstram que em cerca de 60% das ocorrências nominais do português se verifica uma correlação parcial entre o índice temático e valor de género gramatical. Nos restantes casos não há qualquer indício de natureza morfológica que permita justificar a atribuição do valor de género gramatical aos nomes. Estes dados são, assim, reveladores da fraca robustez que existe na associação entre índices temáticos e os valores de género nominal em português.

Além dos índices temáticos, existem ainda outros constituintes morfológicos que condicionam a associação dos valores de género nominal do português. Com efeito, e como atrás se referiu, os sufixos derivacionais, para além de determinarem a categoria sintática do produto final, possuem informação de género quando geram produtos nominais (Villalva 2003; Rio-Torto & Rodrigues 2016:140-150; Choupina, Baptista & Costa 2014), alterando, em certos casos, o género da base nominal a que se associam. Por exemplo, ao nome de género feminino *vasilha* pode agregar-se o sufixo (*-am(e)*) gerando-se, assim, um

produto nominal de género masculino: *o vasilhame*. Desta forma, há alguns sufixos derivacionais que dão origem a nomes masculinos enquanto outros formam nomes de género feminino. No Quadro 1.4 apresentam-se alguns sufixos que formam nomes em português e respetivos índices temáticos<sup>41</sup>, optando-se por não incluir alguns dos sufixos associados aos índices temáticos *-o* e *-a* porque, nesses casos, a correlação entre o valor de género e índice temático é perfeita, como se verifica nos sufixos *-ism(o)*: *o monarquismo*, *o romantismo* e *-ari(a)*: *a livraria*, *a pastelaria*, *a cafetaria*.

Sufixos derivacionais	Valor de género		
	Masculino	Feminino	Comum de dois
<b>-agem</b>		<i>a paisagem, a vadiagem, a contagem</i>	
<b>-am(e)</b>	<i>o vasilhame, o velame</i>		
<b>-al</b>	<i>o pantanal, o lodaçal, o lamaçal, o portal</i>		
<b>-ão</b>	<i>o rasgão, o tropeção, o pilhão</i>		
<b>-ção</b>		<i>a constipação, a elaboração, a preparação</i>	
<b>-ês<sup>42</sup></b>	<i>o economês</i>		
<b>-ez</b>		<i>a sensatez, a timidez, a estupidez</i>	
<b>-ic(e)</b>		<i>a palermice, a criancice</i>	
<b>-íci(e)</b>		<i>a imundície, a calvície</i>	
<b>-idad(e)</b>		<i>a teatralidade, a moralidade</i>	
<b>-idão</b>		<i>a escuridão, a prontidão</i>	
<b>-ist(a)</b>			<i>o/a jornalista, o/a motorista</i>
<b>-dor(a)</b>	<i>o orientador, o gerador, o aspirador</i>	<i>a orientadora, a apresentadora</i>	
<b>-il</b>	<i>o gatil, o canil</i>		
<b>-im</b>	<i>o farolim</i>		
<b>-it(e)</b>		<i>a amígdalite, a bronquite</i>	
<b>-nt(e)</b>			<i>o/a estudante, o/amante, o/a servente</i>
<b>-om(a)</b>	<i>o hematoma, o papiloma, o fibroma</i>		
<b>-or(a)</b>	<i>o leitor, o cantor</i>	<i>a leitora, a cantora</i>	
<b>-os(e)</b>		<i>a neurose, a silicose, a fibrose</i>	
<b>-tud(e)</b>		<i>a altitude, a quietude</i>	
<b>-um(e)</b>	<i>o azedume, o negrume</i>		
<b>-ugem</b>		<i>a penugem</i>	

Quadro 1.4 – Relação entre sufixos derivacionais e respetivo valor de género

<sup>41</sup> Os dados apresentados resultam da consulta da obra de Rio-Torto *et al.* 2016, Capítulo 2.

<sup>42</sup> Sobre as especificidades relativas à utilização do sufixo *-ês*, considerado sufixo formador de nomes de ‘linguagem hermética’, veja-se Rio-Torto & Rodrigues 2016:170-171.

Como se verificou, a atribuição dos valores de género aos nomes do português pode resultar da aplicação de critérios de natureza semântica e, sobretudo, formal. Nos nomes que nomeiam entidades sexuadas, e salvo algumas exceções, constata-se que existe uma correlação entre os valores de género gramatical masculino/feminino e o género natural da entidade designada, sendo que em alguns casos se observa uma correlação parcial entre o índice temático ‘-o’ e ‘-a’ e o género masculino e feminino, respetivamente.

Já em alguns nomes com o traço semântico [-sexuado] se verifica uma certa correspondência entre os índices temáticos (-o e -a) e valores de género masculino e feminino, respetivamente. Todavia, nem sempre essa correlação se observa. Na verdade, para além de haver nomes masculinos terminados em -a (*o mapa*) e femininos terminados em -o (*a tribo*), existem outras terminações que tanto podem estar associadas ao género masculino como ao género feminino.

A Figura 1.5 apresenta, tendo em conta os indícios de atribuição de valores de género nominal do português, uma proposta de distribuição dos nomes em função da aplicação de critérios semânticos e/ou formais.

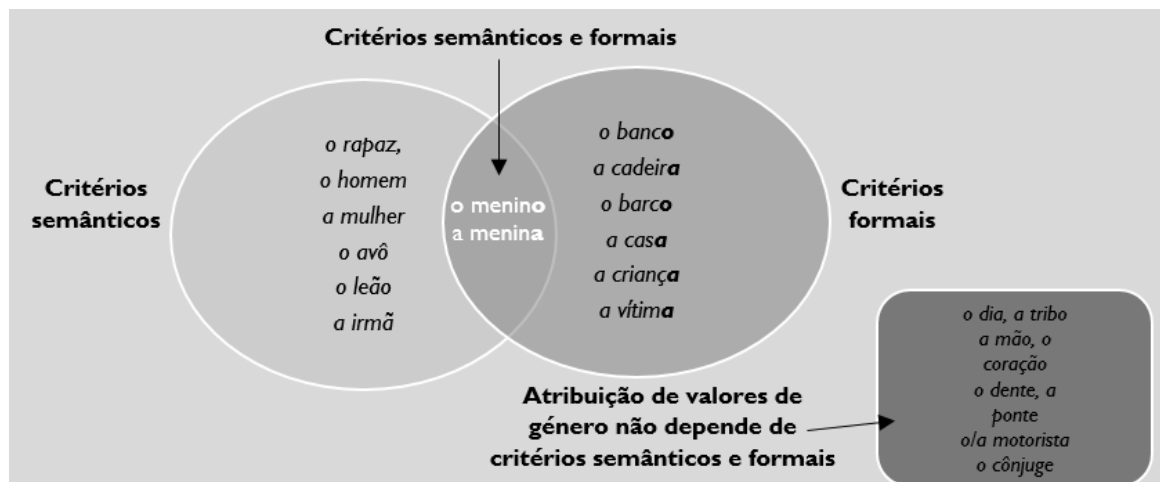


Figura 1.5 – Aplicação de critérios (semânticos e formais) de atribuição de género nominal do português

Na perspectiva de quem adquire/aprende o português como LNM, as características descritas apontam, por um lado, para a pouca robustez dos dados disponíveis no *input* no que diz respeito ao nome em si mesmo, quer numa perspectiva formal, quer numa perspectiva semântica. Por outro lado, parece evidente que a concordância sintática se constitui como um fator importante para determinar o valor de género dos nomes em português. Por conseguinte, a concordância, que é um mecanismo reconhecidamente problemático na aquisição/aprendizagem de uma LNM (cf. Franceschina 2005; Leiria 2006; Godinho 2010; Martins 2015) torna a categoria gramatical de género especialmente problemática neste contexto.



## 1.4. O género gramatical nas LM dos aprendentes de PLNM participantes no estudo

Após terem sido consideradas as principais características da atribuição de género nominal do português, procede-se em seguida à análise e descrição das estruturas de classificação nominal das diferentes línguas maternas (LM) dos aprendentes, autores dos textos escritos selecionados para o *corpus* desta investigação. Uma vez que o estudo incide sobre o (possível) papel que a LM do falante desempenha ao longo de diferentes fases do processo de aquisição/aprendizagem de português como língua não materna (PLNM), selecionaram-se textos quer de aprendentes cujas LM apresentam um sistema de atribuição de género nominal muito próximo do português, quer de aprendentes cujas LM possuem um sistema de classificação nominal diferente. Assim sendo, consideraram-se as seguintes línguas:

- (i) o **espanhol**, que, além de pertencer à mesma família linguística do português, a LA, possui o mesmo número de valores de género, sendo que o respetivo processo de atribuição morfossintática demonstra fortes semelhanças com o da língua portuguesa (cf. 1.4.1.);
- (ii) o **italiano** que, apesar de ser um idioma românico com dois valores de género gramatical, apresenta algumas diferenças em relação ao português no que concerne às características do sistema de atribuição dos valores de género (cf. 1.4.2.);
- (iii) o **alemão**, língua germânica, com três valores de género gramatical (masculino, feminino e neutro) e com distintas particularidades no que diz respeito à associação dos valores de género (cf. 1.4.3.);
- (iv) o **inglês**, língua germânica sem valores de género gramatical nos nomes, modificadores e especificadores, embora conserve vestígios de um primitivo sistema de género gramatical, sobretudo ao nível dos pronomes (cf. 1.4.4.);
- (v) o **chinês**, língua sino-tibetana que, para além de não possuir um sistema de género gramatical, não admite a realização de relações de concordância nominal (cf. 1.4.5.).

Os idiomas selecionados abrangem, assim, distintas situações no que à classificação nominal diz respeito. Por conseguinte, será possível averiguar, por exemplo, se o facto de as respetivas LM dos aprendentes serem menos distantes em relação à LA poderá

desempenhar um papel facilitador no processo de aquisição/aprendizagem da marcação de género nominal em PLNM, ao contrário do que se verifica em aprendentes cuja LM se encontra mais afastada do sistema de atribuição de género nominal do português.

Cada subcapítulo relativo à descrição dos sistemas de classificação nominal das línguas selecionadas encontra-se organizado da seguinte forma: em primeiro lugar, é feita uma breve descrição da estrutura formal dos nomes com a respetiva referência à organização do sistema de classificação dos mesmos e, posteriormente, procede-se, nos casos de idiomas com sistemas de género gramatical, à identificação dos critérios (semânticos e formais) que atuam na atribuição dos valores de género aos nomes.

## 1.4.1. O sistema de atribuição de género nominal do espanhol

### 1.4.1.1. Sobre a estrutura formal dos nomes em espanhol

Na língua espanhola, os nomes e pronomes são especificados quanto ao género gramatical e distribuem-se por dois valores: masculino e feminino. Há ainda outras classes de palavras, tais como determinantes, adjetivos e quantificadores, que recebem determinado valor de género através da concordância sintática desencadeada pelo item nominal com o qual se relacionam (Harris 1991; Ambadiang 1999:4882; Nissen 2002:254; RAE 2009:81; Dubert 2010:62), tal como se pode verificar nos exemplos seguintes:

(11)

- a. *este cuarto capítulo* ‘este quarto capítulo’
- b. *los libros eran suyos* ‘os livros eram seus’
- c. *la mesa pequeña* ‘a mesa pequena’
- d. *ella es muy simpática* ‘ela é muito simpática’

Na descrição do sistema de atribuição de género nominal do espanhol, T. Ambadiang (1994, 1999) baseia-se no conceito de “estructura flexiva de palabra”. De acordo com esta perspetiva, a configuração morfológica dos nomes inclui a possibilidade de preenchimento de informação relativa ao género e ao número (Ambadiang 1994:104).

No que respeita ao género, o mesmo autor argumenta que, em formas como *niño* ‘menino’ e *niña* ‘menina’, são observáveis padrões regulares de atribuição de género, uma vez que nestes itens atuam processos de natureza morfológica, em que, ao radical *niñ-*, poder-se-ão concatenar marcas de flexão de género que, no entender do gramático, correspondem às vogais *-o* e *-a*<sup>43</sup> (Ambadiang 1994:108). Desta forma Ambadiang defende que as desinências vocálicas átonas *-o* e *-a* em posição final, para além de serem indicadoras

---

<sup>43</sup> Segundo Ambadiang (1994) há vários indícios que justificam o tratamento das vogais *-o* e *-a* como marcas canónicas dos valores de género. Por um lado, o autor constata que no processo de aquisição linguística, quer por falantes nativos, quer por aprendentes tardios se regista uma tendência para a *sobregeneralização* do uso destas vogais, observando-se a produção de formas não dicionarizadas como *típo / típa, coche / cocha* ou *\*la fantasma* e *\*el radio*. Tais casos mostram que, inconscientemente, os falantes associam as vogais *o* e *a* aos valores de género masculino e feminino, respetivamente. Para além disso, na formação de apreciativos, em que, por exemplo de *papá* e *mamá* derivam as formas *papa(c)ito* e *mama(c)ita*, as vogais selecionadas pelo sufixo são *-o* e *-a* atendendo ao valor de género associado à forma original (1994:94). Sugere-se, então, que, e à semelhança do português, também em espanhol há uma correlação parcial entre o índice temático *-o* e género masculino e entre o índice temático *-a* e género feminino.

de classe temática a que pertencem os nomes, podem, em determinados casos, estar correlacionadas com o valor de género nominal (masculino / feminino), à semelhança do que ocorre no português, por exemplo (cf. Secção 1.3.3.). Em contrapartida, nomes em que não se verifica a variação *-o / -a*, como *mano* ‘mão’, *papel* ‘papel’, *sofá* ‘sofá’, são entendidos como “unidades léxicas irregulares (...) totalmente especificadas en cuanto a su configuración segmental y/o acentual” (1994:114). A diferença entre estes itens e casos como *niño~niña* advém do facto de, nestes últimos, o valor de género (masculino / feminino) não se encontrar especificado no radical nominal.

Ao descrever a estrutura formal do sistema de género do espanhol, Harris (1991, 1992) observa que, por não existir uma relação direta e inequívoca entre as desinências nominais *-o* e *-a* e valores de género masculino e feminino, estas vogais não podem ser tratadas como morfemas de género<sup>44</sup>. Assim sendo, o autor (1991:36) argumenta que o género corresponde a uma propriedade lexical dos itens nominais<sup>45</sup>, sendo que as vogais átonas em posição final correspondem a *word markers* (Harris 1991:30), i.e., a indicadores da classe temática do nome. Ora, pelas suas características, os *marcadores de palavra* são equivalentes aos índices temáticos apresentados para o português (Villalva 2003; veja-se ainda a Secção 1.3.1.). Portanto, e segundo este autor, os índices temáticos correspondem a constituintes morfológicos que, do ponto de vista gramatical, se caracterizam em função da sua forma fonológica — não são acentuados — e em função da sua distribuição — ocorrem sempre em posição final —, e delimitam uma palavra completa, não podendo ser seguidos de outros sufixos, quer derivacionais, quer flexionais, à exceção da marca de plural *-s* (Harris 1991:30).

Na sua proposta inicial, Harris (1991:31) distribui os nomes do espanhol por oito classes temáticas, distinguindo os itens que não apresentam qualquer índice temático, como as formas *sol*, *col* ou *amante*<sup>46</sup>, daqueles que no singular terminam em vogal seguida de consoante, como *dosis*, *lunes*. Este autor propõe ainda uma organização hierárquica dos índices temáticos e, em função do número de nomes e dos respetivos valores de género

---

<sup>44</sup> Harris admite ainda que as vogais *-o* e *-a* não são morfemas de género porque ocorrem também em palavras que não admitem serem especificadas relativamente a esta categoria gramatical como, por exemplo, os advérbios (1991:30).

<sup>45</sup> Também Dubert (2010:64) considera que “a propriedade de género no substantivo (...) forma parte da entrada léxica, (...) está lexicalmente determinada”.

<sup>46</sup> Segundo Harris (1991) as palavras atemáticas terminam, tipicamente, em consoante como *sol*, *mártir* ou em <e> como, por ex., *amante*, *padre*. Nestes casos, a presença de <e> é justificada por razões de natureza articulatória, uma vez que as sequências consonânticas <nt> ou <dr> não são permitidas em coda (1991:31).

associados a cada índice, considera que há índices temáticos *canónicos* e *não canónicos*. De acordo com esta proposta, são considerados *canónicos* os índices temáticos *-o* e *-a* quando associados aos valores de género masculino e feminino, respetivamente, como *el libro* ‘o livro’ e *la casa* ‘a casa’. Já os *não canónicos* ou *residuais* correspondem aos casos em que o índice temático *-o* se encontra em nomes de género feminino (*la mano* ‘a mão’) e o índice temático *-a* em nomes de género masculino (*el día* ‘o dia’).

Em 1992, o autor revê a sua proposta, considerando, então, que os itens lexicais do espanhol se distribuem por cinco classes temáticas<sup>47</sup>. Assim, os itens terminados em *-o* integram a Classe I (*libro* ‘livro’, *mano* ‘mão’) e os itens de tema em *-a* constituem a Classe II (*problema* ‘problema’, *cuchara* ‘colher’). A Classe III abrange três subconjuntos: os subconjuntos A, A’ e B compreendem itens terminados em *-e*<sup>48</sup>. A Classe A’ integra formas terminadas em consoante (ou em *-e* teórico (Câmara Jr. 1994), dado que esta vogal surge no plural, como em *papel~papeles*). A Classe IV engloba um número muito reduzido de itens terminados em *-s* ou em vogal seguida de *-s* (Harris 1992:66). Por fim, na Classe V, considerada ‘exótica’, encontram-se itens nominais terminados em *-i* e em *-u*, muito pouco frequentes no léxico da língua espanhola, além de alguns estrangeirismos terminados em consoante.

Veja-se, no Quadro 1.5, a distribuição dos itens nominais do espanhol, de traço semântico [+sexuado] e [-sexuado], de acordo com proposta de Harris (1992).

---

<sup>47</sup> É de ressaltar que nas classes temáticas propostas por Harris (1991, 1992) se encontram, para além dos nomes, formas adjetivais e adverbiais, sendo que estas últimas não são especificadas quanto ao género gramatical. Deste modo, o autor pretende demonstrar que mesmo palavras sem especificação de género possuem marcadores de palavra e, por conseguinte, estes constituintes não podem ser considerados como morfemas de género. No Quadro 1.5 optou-se por colocar apenas exemplos de formas nominais, atendendo aos objetivos da presente investigação.

<sup>48</sup> No entendimento do autor, nos nomes do subconjunto A, a vogal *-e* surge por questões de natureza articulatória e, no subconjunto B, encontram-se formas em que a presença da vogal não é justificada por razões fonológicas (Harris 1992:66).

Classe temática	Índice temático (marca de palavra)	Valores de género		
		Masculino	Feminino	Comum de dois (masculino e feminino)
I	-o	<i>libro</i> 'livro'; <i>banco</i> 'banco'; <i>muchacho</i> 'menino'; <i>hijo</i> 'filho'	<i>foto</i> 'fotografia'; <i>mano</i> 'mão'	<i>soprano</i> 'soprano'; <i>testigo</i> 'testemunha'
II	-a	<i>problema</i> 'problema'; <i>planeta</i> 'planeta'; <i>día</i> 'dia'	<i>cuchara</i> 'colher'; <i>mesa</i> 'mesa'; <i>muchacha</i> 'menina'; <i>hija</i> 'filha'	<i>artista</i> 'artista'; <i>camarada</i> 'camarada'; <i>colega</i> 'colega'
III (A)	-e	<i>punte</i> 'ponte'; <i>coche</i> 'carro'; <i>padre</i> 'pai';	<i>noche</i> 'noite'; <i>leche</i> 'leite'; <i>madre</i> 'mãe'	<i>amante</i> 'amante'
III (A')	-Ø, -e (teórico)	<i>sol</i> 'sol', <i>mártir</i> 'mártir';	<i>col</i> 'couve'	<i>joven</i> 'jovem'
III (B)	-e	<i>pase</i> 'passe'	<i>prole</i> 'descendência'	
IV	-Vs	<i>tórax</i> 'tórax'	<i>dosis</i> 'dose'	
V	-i	<i>táxi</i> 'táxi'	<i>metrópoli</i> 'metrópole'	
	-u	<i>espíritu</i> 'espírito'	<i>tribu</i> 'tribo'	
	-Ø	<i>chef</i> 'chefe'; <i>golf</i> 'golfe'		

**Quadro 1.5** – Distribuição dos itens nominais em espanhol em função da classe, índice temático e respectivos valores de género (de acordo com a proposta de Harris (1992:65))

A partir dos dados expostos, são assinaláveis as similaridades entre as classes temáticas do espanhol e as classes definidas para o português (cf. Quadro 1.2). Além disso, e tal como acontece na língua portuguesa, não existe uma correspondência direta e inequívoca entre determinada estrutura formal e respetivo valor de género gramatical do nome.

#### 1.4.1.2. Sobre a natureza (não) flexional do género em espanhol

Em português, e como vimos, é muito discutível a associação do género a processos de natureza flexional (cf. Secção 1.3.2.). Vejamos, para o espanhol, em que medida esta questão é também equacionada.

É no capítulo dedicado à flexão nominal, publicado na *Nueva Gramática de la Lengua Española* que Ambadiang (1999) procede à descrição detalhada e à análise dos sistemas de flexão das categorias de género e de número nominais, reconhecendo, no entanto, o primeiro sistema como mais assistemático e irregular do que o segundo (1999:4845). Tendo em conta as várias propriedades (semânticas e formais – morfológicas e fonológicas –) que estão na base da associação dos valores de género aos nomes em espanhol, quer as formas simples, quer as formas derivadas, o autor rejeita a ideia de que a especificação do género nominal seja arbitrária, uma vez que é possível encontrar uma certa base semântica e/ou formal na associação dos valores de género, tanto nos nomes com o traço semântico [+sexuado] como os que possuem o traço [-sexuado] (1999:4859).

Ambadiang (1994, 1999) admite, portanto, que há flexão de género em pares de palavras como por exemplo *niño/niña* ‘menino/menina’, *abad/ abadesa* ‘abade/abadessa’ e *gallo/gallina* ‘galo / galinha’, em que o masculino corresponde à forma não-marcada, a que se opõe uma outra forma do feminino, forma marcada pela presença de uma desinência (seja *-a*, *-esa* ou *-ina*) (1999:4861). Por sua vez, casos como *manzano/manzana* ‘macieira/maçã’, *tomatera/tomate* ‘tomateiro / tomate’ já não constituem, nesta perspetiva, formas de flexão, uma vez que configuram uma relação do tipo derivacional, em que *manzano* e *tomatera* correspondem a ‘plantas que dão origem a *manzana* ‘maçã’ e *tomate* ‘tomate’. Ou seja, em *gallo/gallina* existe flexão, porque o valor de género gramatical destas formas é, antes de mais, correspondente ao género natural dos respetivos referentes, não sendo, portanto, determinado pelo sufixo, tal como ocorre em *manzano/ manzana* ou *tomatera/tomate*.

Apesar do teor descritivo da análise de Ambadiang (1994, 1999), a verdade é que esta não é suficiente para explicitar a natureza variável da categoria de género gramatical na língua espanhola, observando-se, de certo modo, uma tendência para a simplificação dos dados reais da língua. E, como vimos, é muito discutível a associação das vogais átonas *-o* e *-a* a morfemas de género no espanhol (cf. Secção 1.4.1.1.). Além disso, ao se distinguirem, dentro da flexão, unidades lexicais ditas ‘regulares’ das unidades lexicais ‘irregulares’, i.e., tendo em conta o facto de se encontrarem especificadas ou não quanto ao valor de género (Ambadiang 1994:108), não é, em boa verdade, uma perspetiva produtiva para a descrição do sistema de género nominal do espanhol.

Assim, atendendo às características do sistema de atribuição de género do espanhol, e à semelhança do que se viu para o português (cf. Secção 1.3.), muito dificilmente se pode considerar que o género corresponde a uma categoria da flexão, já que a atribuição dos valores de género nominal não corresponde a um mecanismo sistemático nem obrigatório.

Com efeito, nem todos os nomes admitem a oposição de género gramatical. Só alguns nomes de entidades sexuadas permitem a variação em género, havendo diferentes processos linguísticos que possibilitam a materialização dos contrastes, tais como:

- (i) processos lexicais, i.e., quando duas formas lexicais distintas com referentes semelhantes e diferenciáveis pelo sexo estão associadas aos dois valores de género disponíveis no espanhol: *hombre* ‘homem’ / *mujer* ‘mulher’; *toro* ‘touro’ / *vaca* ‘vaca’ e pela distinção de constituintes temáticos: *niño* ‘menino’ / *niña* ‘menina’; *profesor* ‘professor’ / *professora* ‘professora’;

- (ii) processos morfológicos, como a derivação: *duque* ‘duque’ / *duquesa* ‘duquesa’.
- (iii) processos sintáticos: *el artista* / *la artista*.

Para além disso, nem todos os nomes com referentes sexuados admitem variação de género. É o caso dos nomes epícenos como *ballena* ‘baleia’ em que à base nominal «*ballena*» ‘baleia’ se adiciona a forma «*macho*» ‘macho’ «*ballena macho* ‘baleia-macho’» ou «*hembra*» ‘fêmea’ «*ballena hembra* ‘baleia-fêmea’», criando-se, assim, um composto sintático que especifica o sexo do referente, mantendo-se, no entanto, o mesmo valor de género gramatical (Ambadiang 1999; Dubert 2010:67). O mesmo se verifica nos nomes sobrecomuns como *víctima* (vítima) que possuem um valor de género único, independentemente do sexo da entidade à qual se referem (RAE 2009:84).

Em suma, e como se pôde constatar, os problemas que advêm na descrição do sistema de atribuição de género nominal do espanhol são, efetivamente, muito semelhantes aos assinalados em português e, ao contrário do que defende Ambadiang (1994, 1999), é muito discutível considerar que em espanhol o género corresponde a uma categoria da flexão<sup>49</sup>.

#### 1.4.1.3. Critérios de atribuição de género aos nomes em espanhol

No que diz respeito aos critérios (semânticos e formais) que atuam na associação dos valores de género dos nomes em espanhol considere-se, em primeiro lugar, os itens nominais de entidades sexuadas. Nestes casos, e de um modo geral, o valor de género gramatical corresponde ao género natural do referente. Assim, temos:

- (i) nomes com valor de género gramatical masculino que denotam seres do sexo masculino — (*el*) *padre* ‘(o) pai’; (*el*) *niño* ‘(o) menino’; e

---

<sup>49</sup> Porém, Dubert (2010:67-68) ao analisar o comportamento da categoria de género no espanhol, defende que “o xénero é unha categoría morfosintáctica flexiva simplemente porque dispara procesos de concordancia, considerados sintáticos, aínda que non se reciba por medio de ningún proceso e estea almacenada desde o principio na entrada léxica, do mesmo xeito que está almacenada desde o principio outra información sintáctica, como a categoría sintáctica Subs, ou a representación fonolóxica /mar/. De feito, o xénero é parte da información sintáctica que cargan as entradas léxicas dos substantivos. Polo tanto, aínda que en *mar* no hai un proceso de flexión (entendo por tal a introducción dunha propiedade morfosintáctica antes ausente), é unha entrada léxica flexionada, pois contén unha categoría flexiva *inherente*”.



- (ii) nomes femininos que denotam seres do sexo feminino — *(la) madre* ‘(a) mãe; *(la) niña* ‘(a) menina’.

A aplicação do critério semântico para a atribuição dos valores de género (masculino / feminino) dos nomes animados apresenta, no entanto, algumas restrições. Por um lado, e como se viu (cf. Secção 1.4.1.2.), nos nomes epicenos (*rinoceronte* ‘rinoceronte’) e sobrecomuns (*víctima* ‘vítima’), o valor de género não corresponde necessariamente ao sexo do seu referente (Ambadiang 1999:4850). Também os nomes comuns de dois géneros possuem uma única forma que tanto pode ser de género masculino como feminino. Nestes, a ambiguidade desfaz-se a partir do contexto sintático em que ocorrem: «*el artista/ la artista* ‘(o/a) artista’; *este testigo / esta testigo* ‘(esta) testemunha’» (RAE 2009).

Por fim, e apesar de pouco frequentes no léxico espanhol, há algumas formas nominais [+sexuadas] em que o valor de género gramatical é oposto ao sexo da entidade designada (Ambadiang 1999:4849). A título de exemplo refira-se o termo «*(el) marimacho* ‘(a) maria-rapaz’» que, embora masculino, designa, em tom pejorativo e coloquial, uma mulher com características masculinas<sup>50</sup>.

No que respeita aos nomes de entidades não-sexuadas, a associação dos valores de género é, atendendo ao critério semântico, verdadeiramente imprevisível e aleatória<sup>51</sup>. Ainda que se observe uma correlação parcial entre os índices temáticos *-o* e *-a* e valores de género masculino e feminino<sup>52</sup> (*el tiempo* ‘o tempo’, *la alegría* ‘a alegria’), a verdade é que nem sempre essa correlação ocorre, havendo no léxico do espanhol formas nominais masculinas de tema em *-a* «*(el) día* ‘(o) dia’» e femininas de tema em *-o* «*(la) mano* ‘(a)

---

<sup>50</sup> Este tipo de itens configura, de facto, um subconjunto muito restrito e pouco frequente em espanhol. Numa tentativa de eliminar tal ‘irregularidade’, no México e em alguns países de língua oficial espanhola, regista-se a forma «*marimacha*» ou, por vezes, opta-se pelo uso de «*marimacho*» associado ao género gramatical feminino, como se ilustra no exemplo seguinte: “*es una marimacho*” (RAE 2009:83).

<sup>51</sup> No entanto, Ambadiang (1999:4851) sugere que é possível inferir uma certa motivação semântica na atribuição dos valores de género aos nomes não-sexuados, uma vez que há determinados grupos semânticos que, em certa medida, podem condicionar o valor de género dos itens que os integram. Por exemplo, são de género masculino os nomes dos dias da semana e dos meses do ano (*el lunes* ‘segunda’, *el domingo* ‘domingo’, *el enero* ‘janeiro’) e são femininos os itens nominais que designam as letras do alfabeto (*la a* ‘a’, *la efe* ‘F’, *la ene* ‘n’). Porém, o mesmo autor reconhece que nem sempre a pertença a um determinado grupo semântico justifica o valor de género do item nominal porque, em muitos dos casos, ocorrem inúmeras exceções à ‘regra’.

<sup>52</sup> No que concerne à associação de valores de género masculino e feminino com os índices temáticos *-o* e *-a*, respetivamente, indica Nissen (2002:254) que em espanhol “99.8% of the nouns ending in *-o* are masculine, 96.6% of the nouns ending in *-a* are feminine, and 89.35% of the nouns ending in *-e* are masculine”.

mão'» (Harris 1991; Ambadiang 1999; Rae 2009). Encontram-se ainda nomes com terminações associadas aos dois valores de género como, por ex., *papel*, *limón* 'limão', *col* 'couve', etc. (cf. Quadro 1.5).

Outros constituintes morfológicos, como os sufixos derivacionais, determinam o valor de género das respetivas formas nominais, uma vez que, e tal como constata Ambadiang (1999:4876), “el afijo impone su rasgo de género (...) el género de los nombres derivados se predice a partir del de su último constituyente sufijal”. Recorde-se que esta é também uma característica do sistema de atribuição de género do português, como se verificou a partir dos dados recolhidos no Quadro 1.4, havendo, por este motivo, uma forte proximidade entre os dois sistemas de género destas duas línguas românicas.

O mesmo autor (1999:4876) observa, por fim, que há sufixos derivacionais portadores de ou que se fazem acompanhar dos índices temáticos *-o*, *-e* e *-a* e que apresentam, de um modo geral e salvo algumas exceções pontuais, a seguinte tendência:

- (i) os sufixos que terminam em *-o* e *-e* originam, tipicamente, nomes de valor de género masculino, como *-ado*, *-azgo*, *-edo*, como *arbolado* 'arvoredo' e *anclaje* 'ancoragem';
- (ii) os sufixos com índice temático *-a* originam, tipicamente, nomes femininos;
- (iii) todos os sufixos derivacionais terminados em *-d* ou *-z* formam nomes de género feminino: *la humedad* 'a humidade', *la memez* 'asneira'; e
- (iv) todos os sufixos derivacionais terminados em *-l* ou *-r* dão origem a nomes masculinos: *el berenjenal* 'plantação de beringelas'; *el pintor* 'o pintor'; *el obrador* 'trabalhador'.

## 1.4.2. O sistema de atribuição de género nominal do italiano

### 1.4.2.1. A estrutura formal dos nomes em italiano

A presente Secção é dedicada à descrição e análise do sistema de atribuição de género nominal do italiano, língua que, juntamente com o espanhol e o português, é românica. Por conseguinte, e no que ao género gramatical diz respeito, são assinaláveis algumas semelhanças entre estes três idiomas, apesar de, e como se verá, o italiano apresentar algumas particularidades distintivas.

Assim, e à semelhança do que se observa nos idiomas românicos anteriormente apresentados, também o italiano exhibe dois valores de género em oposição: o masculino e o feminino, sendo que os itens nominais e alguns pronomes possuem um valor de género intrínseco, enquanto outras classes de palavras (adjetivos, determinantes, etc.) recebem determinado valor pela concordância sintática desencadeada pelo nome com o qual estão associados (Marcato & Thüne 2002:194).

No que concerne à associação dos valores de género aos nomes, constata-se que esta resulta da aplicação de critérios de natureza semântica e formal, embora estes nem sempre atuem de modo consistente. Efetivamente, se em grande parte dos nomes de entidades sexuadas o valor de género gramatical (masculino / feminino) corresponde ao género natural dos respetivos referentes (por ex., *ragazzo* ‘rapaz’ é masculino por designar uma entidade do sexo masculino), há nomes de entidades sexuadas cujo género gramatical não corresponde ao sexo do respetivo referente, como, por ex., o sobrecomum feminino *la guardia* que designa, independentemente do sexo do referente, as pessoas encarregadas de guardar alguma coisa (Dardano & Trifone 1997; ver ainda Secção 1.4.2.2. do presente trabalho).

Em italiano, as categorias de género e de número nominais encontram-se fortemente associadas. Efetivamente, para além de serem sempre especificados quanto ao valor de género, é possível os nomes exibirem, na sua forma, informação relativa à categoria de número (*singular* vs. *plural*) e, ao contrário do que acontece no português e no espanhol, a formação do plural caracteriza-se, de um modo geral, pela alteração dos índices temáticos, sendo raros os casos da simples adição de *-s* à forma nominal do singular<sup>53</sup>. Por exemplo,

---

<sup>53</sup> A marcação do plural através da adição de *-s* ocorre, essencialmente, nos estrangeirismos ou palavras provenientes de outras línguas que ainda não integraram o sistema nominal do italiano (cf. Dressler & Thornton (1996:13-14)). Por não ser frequente a marcação do número pela adição de *-s* à forma do singular, Vincent (1988:289) considera este facto “one of the features which marks Italian (...) off from Western Romance languages such as French or Portuguese”.

e a título ilustrativo, a forma do plural do nome *libro* ‘livro’ é *libri* evidenciando-se, assim, a alteração da vogal final que, de *-o*, passou a *-i*. Já para o feminino singular *donna* ‘senhora’ a forma correspondente no plural é *donne*, com a mudança do constituinte temático final, i.e., de *-a* para *-e*. Estas alterações resultam da pertença dos nomes a diferentes classes temáticas.

Estes exemplos permitem, em parte, demonstrar uma característica da estrutura mórfica do italiano, em que, tipicamente, as classes de palavras, quer lexicais, como nomes e adjetivos, quer gramaticais, como determinantes, se caracterizam por terminar em vogal que, como se viu, pode carrear informação de género e simultaneamente de número<sup>54</sup> (Dressler & Thornton 1996; Oliphant 1997; Acquaviva 2009). Pelas características que apresentam e função que desempenham, as vogais finais dos nomes correspondem ainda a índices temáticos.

Segundo a tradição gramatical italiana, existem diferentes paradigmas nos quais as várias classes de palavras se organizam (Dardano & Trifone 1997). Mais concretamente, tanto nomes, como adjetivos e determinantes obedecem a determinados padrões, estipulados em função dos índices temáticos que apresentam no singular e no plural (Vincent 1988; Dressler & Thornton 1996; Dardano & Trifone 1997; Acquaviva 2009).

Ao descrever a morfologia nominal do italiano, relativamente à marcação de número, Dardano & Trifone (1997) distinguem as formas variáveis das não variáveis, i.e., nomes que admitem variação de número e os que não admitem essa variação. Os nomes variáveis integram diferentes classes temáticas, atendendo aos índices associados à forma do singular e à respetiva forma do plural.

Assim, entre as formas variáveis, distinguem-se:

- (i) itens terminados em *-o* no singular e em *-i* no plural (*libro-libri* ‘livro-livros’; *mano-mani* ‘mão-mãos’);
- (ii) itens terminados em *-a* no singular e em *-e* no plural (*casa-case* ‘casa-casas’; *donna-donne* ‘senhora-senhoras’);
- (iii) nomes terminados em *-e* no singular e em *-i* no plural (*monte-monti* ‘monte-montes’ / *mente-menti* ‘mente-mentes’); e

---

<sup>54</sup> A propósito desta particularidade Dressler & Thornton (1996) constatarem que: “[i]talian noun inflection has the system-defining properties of number (Sg. unmarked, Pl. marked) and gender (m. unmarked, f. marked), where both categories are simultaneously signaled on the noun. Both categories are also signaled by definite and indefinite article (...)” (1996:3).

- (iv) itens terminados em *-a* no singular e em *-i* no plural (*problema-problemi* ‘problema-problemas’; *poeta-poeti* ‘poeta-poetas’).

Por sua vez, nas formas não variáveis, existem:

- (i) alguns nomes terminados em *-a* (*vaglia* (sing.) – *vaglia* (pl.) ‘ordem de pagamento’);
- (ii) alguns itens terminados em *-o* (*radio* (sing.) – *radio* (pl.));
- (iii) nomes terminados em *-i* e em *-u* (*brindisi* (sing.) – *brindisi* (pl.) ‘brinde’; *ipotesi* (sing.) – *ipotesi* (pl.) ‘hipótese’; *gru* (sing.) – *gru* (pl.) ‘grua’);
- (iv) nomes terminados em *-(i)e*<sup>55</sup> (*specie* (sing.) – *specie* (pl.) ‘espécie’);
- (v) formas terminadas em vogal tónica (Ø) (*caffè* (sing.) – *caffè* (pl.) ‘café’; *virtú* (sing.) – *virtú* (pl.) ‘virtude’); e
- (vi) itens terminados em consoante que correspondem, geralmente, a estrangeirismos (Ø) (*film* (sing.) – *film* (pl.) ‘filme’; *reclam* (sing.) – *reclam* (pl.) ‘anúncio publicitário’);

Quanto à categoria de género, constata-se que, entre as formas variáveis, a maioria dos nomes de tema em *-o* no singular, e em *-i* no plural (*libro-libri*), e a maioria dos nomes de tema em *-a* no singular e em *-i* no plural (*problema-problemi*) é de género masculino, sendo femininos todos os nomes que, no singular, terminam em *-a* e, no plural, possuem a terminação *-e* (*casa-case*). Já os nomes terminados em *-e* no singular e em *-i* no plural (*ponte-ponti*) distribuem-se pelos valores de género masculino e feminino<sup>56</sup> (Oliphant 1997:7).

No que respeita às tendências de atribuição de valor de género às formas nominais não variáveis, verifica-se que são masculinos os itens terminados em *-a* e femininos os

---

<sup>55</sup> À exceção de *moglie*, *superficie* e *effigie*, formas variáveis com o plural em *-i* (Dardano & Trifone 1997:185-6). Sobre estes casos excepcionais, veja-se, Dressler & Thornton (1996:18, e n.33).

<sup>56</sup> Caffarra *et al.* (2015:1020) observam que nos nomes terminados em *-e* no singular se regista um equilíbrio na proporção de nomes de género masculino e feminino. Segundo os autores (2015:1020): “the gender of singular nouns ending in *-e* cannot be predicted from the word form alone since a similar proportion of masculine and feminine singular nouns end in *-e* (17% and 20%, respectively)”. Também Oliphant (1997:7) afirma que os nomes pertencentes a esta classe temática “are considered unpredictable as to gender.”

terminados em *-o*<sup>57</sup>. Os estrangeirismos terminados em consoante são, geralmente, masculinos (Acquaviva 2009:50).

Nigel Vincent (1988) identifica ainda um subconjunto de itens masculinos terminados em *-o* como, por exemplo, *uovo* ‘ovo’ e *osso* ‘osso’, em que é possível distinguir duas formas de plural, i.e., uma masculina de tema em *-i*, e outra feminina em *-a*<sup>58</sup>. Cada uma destas formas apresenta uma certa variação semântica. Ou seja, ao utilizar-se o plural em *-a*, transmite-se uma ideia de coletividade (*ossa* ‘ossos — entendidos como um conjunto de ossos que compõe o esqueleto’) ao passo que a forma de plural terminada em *-i*, não acarreta esse significado (*ossi* ‘ossos — dispersos/não entendidos como um todo, um esqueleto’). Vincent (1988:289) entende que estes casos, apesar de residuais, são vestígios do plural neutro latino (veja-se ainda a Secção 1.2.5.). Há ainda um subconjunto de itens nominais com referentes sexuados que, no singular, são comuns de dois valores de género, apesar de não o serem no plural. Com efeito, os nomes comuns de dois terminados em *-a* como *colega* ou *atleta* e os terminados em *-ista* ou *-cida*, *artista*, *omicida* ‘homicida’, quando se encontram no plural apresentam uma forma de género masculino, terminada em *-i* (*i collegi*, *gli atleti*, *gli artisti*, *gli omicidi*) e outra de género feminino, terminada em *-e* (*le college*, *le atlete*, *le artiste*, *le omicide*) (Dardano & Trifone 1997:175).

No Quadro 1.6 apresenta-se a distribuição dos itens nominais do italiano, quer as formas variáveis, quer as formas não variáveis, em função das classes temáticas a que pertencem e dos respetivos valores de género associados.

---

<sup>57</sup> Os nomes femininos terminados em *-o* correspondem, na sua maioria, a formas truncadas, como *la foto* e *l'auto*, por exemplo. Este fenómeno é, aliás, muito semelhante ao que se regista no português.

<sup>58</sup> Segundo Paolo Acquaviva (2009) a tradição gramatical italiana tem considerado como expoentes de uma classe temática casos como *osso*~*ossa*, *braccio*~*braccia* cujos valores de género são distintos no singular e no plural. No entanto, o autor defende que estes pares de palavras não configuram uma classe temática particular, dado que estamos perante dois lexemas distintos, cada um marcado com um valor de género específico (2009:51).

	Valor de género			
	Índice Temático (Singular / Plural)	Masculino (Singular / Plural)	Feminino (Singular / Plural)	Comum de dois (masculino e feminino) (Singular / Plural)
Formas variáveis	-a / -e		<i>la casa / le case</i> 'casa'	
	-o / -i	<i>il libro / i libri</i> 'livro'	<i>la mano / le mani</i> 'mão'	
	-e / -i	<i>il fiore / i fiori</i> 'flor'	<i>la classe / le classi</i> 'aula'	<i>il, la nipote / i, le nipoti</i> 'neto'
	-a / -i	<i>il problema / i problemi</i> 'problema'	<i>la ala / le ali</i> 'asa'	<i>il, la collega / i collegi, le college</i> 'colega'
Formas não variáveis	-a	<i>il / i taglia</i> 'ordem de pagamento'		
	-o		<i>la / le radio</i> 'rádio'	
	-i	<i>il / i brindisi</i> 'brinde'	<i>la / le metropoli</i> 'metrópole'	
	-u	<i>il / i guru</i> 'guru'	<i>la / le gru</i> 'grua'	
	-e	<i>il / i re</i> 'rei'	<i>la / le specie</i> 'espécie'	
	-Ø	<i>il / i caffè</i> 'café' <i>il / i film</i> 'filme'	<i>la / le virtù</i> 'virtude'	

**Quadro 1.6** – Distribuição dos nomes do italiano, formas variáveis e invariáveis, em função do índice temático e respetivo valor de género

Dressler & Thornton (1996) propõem ainda a organização das classes temáticas da língua italiana em função do seu grau de 'produtividade', ou seja, atendendo ao número de itens que as integra e respetivo valor de género que lhes está associado<sup>59</sup>. Assim sendo, por exemplo, é considerada produtiva a classe temática dos nomes terminados em *-a* no singular e em *-e* no plural, dado que todos os nomes desta classe são de género feminino (cf. Quadro 1.6); pelo contrário, não é produtiva a classe dos nomes terminados em *-e* no singular e em *-i* no plural, visto que aqui se encontram, quer itens de género masculino, quer itens de género feminino (Dressler & Thornton 1996:18). Além disso, a maior produtividade destas classes temáticas é igualmente visível pelo facto de outras classes de palavras, como adjetivos, determinantes e pronomes, seguirem, preferencialmente, este padrão (sobre este assunto, veja-se também Acquaviva 2009).

Em suma, pese embora o facto de se assinalarem algumas tendências relativamente à classificação temática dos nomes e respetiva associação aos valores de género disponíveis, a verdade é que a estrutura formal dos itens lexicais do italiano não apresenta, na sua maioria, uma relação inequívoca e uniforme entre as classes temáticas e valores de género masculino e feminino. Assim, e como observam Caffarra *et al.* (2015:1020):

<sup>59</sup> Para estes autores, há fortes evidências de que a informação relativa ao género gramatical subordina a organização do sistema de classes temáticas nominais em italiano.

“[t]he characteristics of the Italian nominal system suggest that grammatical gender information cannot be unequivocally recovered from the word form. For instance, gender-to-ending consistency is based on distributional information that can provide ambiguous (in the case of opaque nouns) or unreliable (in the case of irregular nouns) information about the gender of a noun.”

#### 1.4.2.2. Critérios de atribuição de género nominal do italiano

No que diz respeito à atribuição de valores de género é preciso distinguir os nomes de entidades sexuadas dos que têm referentes não-sexuados. No primeiro caso, e tal como acontece em português e em espanhol, observa-se uma base semântica na atribuição dos valores de género (masculino / feminino) em que, apesar de algumas exceções à aplicação deste critério, o género gramatical corresponde ao género natural do referente nominal. Já nos nomes que possuem o traço semântico [-sexuado], a determinação de um valor de género é semanticamente arbitrária (Marcato & Thüne 2002:190).

Assim sendo, em nomes referentes a entidades sexuadas aplica-se um princípio de natureza semântica, uma vez que os valores de género (masculino/feminino) correspondem, em grande parte, ao sexo do respetivo referente (Dardano & Trifone 1997:169). Portanto, e atendendo a este princípio semântico, temos:

- (i) nomes masculinos com referentes de sexo masculino: «*(il) padre* ‘(o) pai’; *(il) gatto* ‘(o) gato’»; e
- (ii) nomes femininos com referentes do sexo feminino: «*(la) madre* ‘(a) mãe’; *(l’) infermiera* ‘(a) enfermeira’».

Todavia, nem sempre o critério semântico se aplica, uma vez que há formas nominais com o traço semântico [+sexuado] cujo valor de género não corresponde, necessariamente, ao sexo do seu referente, como é o caso dos nomes sobrecomuns, masculinos (*il soprano* - ‘o soprano’, *il mezzosoprano* - ‘o meio-soprano’) e sobrecomuns femininos (*la vedetta* - ‘a vedeta’; *la sentinella* - ‘o sentinela’) (Dardano & Trifone 1997:169).

Também encontramos no italiano alguns nomes epicenos (Dardano & Trifone 1997:176), com uma só forma masculina (*corvo* ‘corvo’; *falco* ‘falcão’, *leopardo*



‘leopardo’) e uma só forma feminina (*aquila* ‘águia’, *pantera* ‘pantera’, *balena* ‘baleia’), independentemente do género natural do respetivo referente [-humano]. Nestes nomes é possível adicionar informação relativa ao sexo do referente sem, no entanto, se alterar o valor de género gramatical que lhe está associado<sup>60</sup>. Deste modo, quando se pretende distinguir, numa determinada espécie, o macho da fêmea, o nome epiceno deverá vir acompanhado das formas «*maschio* ‘macho’» e «*femmina* ‘fêmea’» respetivamente, como ilustram os seguintes exemplos:

(12)

- a. *il falco maschio* ‘o falcão macho’ - *il falco femmina* ‘o falcão fêmea’
- b. *la pantera maschio* ‘a pantera macho’ - *la pantera femmina* ‘a pantera fêmea’

Por fim, e ainda relativamente aos nomes com referentes animados, refira-se os comuns de dois géneros, com uma mesma forma associada aos dois valores de género disponíveis no italiano (Dardano & Trifone 1997:175). Nestes casos, a especificação do género gramatical é estabelecida com recurso à concordância sintática, i.e., através das palavras que coocorrem com o nome. Assim, para fazer referência a um ser do sexo masculino, recorre-se, por exemplo, ao artigo masculino *il* (*il pianista*, *l’amante*, *l’atleta*), enquanto o artigo feminino *la* serve para designar o ser do sexo feminino (*la pianista*, *l’amante*, *l’atleta*).

Convém ainda salientar que, e como atrás se viu (cf. Secção 1.4.2.1.), entre os nomes comuns de dois, há casos que apresentam duas formas distintas no plural como *pianista* que, no singular, é ambíguo quanto ao género (*il/la pianista*), mas no plural não é. Com efeito, é possível distinguir, pelo índice temático, uma forma masculina *i pianisti* e outra feminina *le pianiste* (Dardano & Trifone 1997:175). Possivelmente, este fenómeno estará correlacionado com a baixa produtividade da classe temática dos nomes masculinos de índice temático *-a*, no singular, e *-i* no plural, como *sistema-sistemi* ‘sistema’ (cf. Secção 1.4.2.1 e Quadro 1.6). Assim, face à fraca produtividade desta classe temática, no plural estes nomes comuns de dois assumem formas distintas, uma para o masculino (em *-i*) e

---

<sup>60</sup> A respeito dos nomes epicenos, Dardano & Trifone indicam alguns casos que se comportam como nomes comuns de dois géneros, em que uma mesma forma pode ser antecedida por determinante artigo masculino (*il*) ou feminino (*la*). A título de exemplo, os gramáticos referem itens como *il/la serpe* ‘a serpente’ e *il/la lepre* ‘a lepre’. Note-se, no entanto, que a variação de género gramatical não reflete a distinção de género natural. Por outras palavras, o facto de se antepor o artigo masculino *il* ao nome *lepre* não significa, necessariamente, que se esteja a referir ao sexo masculino da espécie, mas à espécie na sua generalidade (Dardano & Trifone 1997:176).

outra para o feminino (em *-e*) obedecendo, assim, aos paradigmas ditos mais produtivos, ou seja, os plurais masculinos terminados em *-i* e os plurais femininos terminados em *-e* (Acquaviva 2009).

Entre os itens de entidades sexuadas que admitem variação de género gramatical e, à semelhança de outras línguas românicas aqui descritas, a materialização dos contrastes de género é assegurada por diferentes mecanismos linguísticos (Dardano & Trifone 1997:171-174), tais como:

- (i) a oposição de constituintes temáticos, como, por ex. *il* 'alumno' 'o aluno' - *l'alumna* 'a alumna'; *il* 'signore' 'o senhor' - *la* 'signora' 'a senhora' e pelo contraste de radicais: *il* 'genere' 'o genro' - *la* 'nuora' 'a nora'; *il* 'fratello' 'o irmão' - *la* 'sorella' 'a irmã'; *il* 'toro' 'o touro' - *la* 'vacca' 'a vaca'; e
- (ii) o processo de derivação *il* 'studente' 'o estudante' - *la* 'studentessa' 'a estudante'; *il* 'difensore' 'o defensor' - *la* 'difenditrice' 'a defensora'.

Relativamente à atribuição dos valores de género aos nomes de referentes não-sexuados constata-se que, em função do critério semântico, ela é totalmente idiosincrática e imprevisível<sup>61</sup>. Considerando-se a estrutura mórfica dos nomes e, em particular, as classes temáticas a que pertencem é possível, em parte, inferir o valor de género dos nomes não-sexuados (Dressler & Thornton 1996; Dardano & Trifone 1997; Oliphant 1997). Na verdade, e como se constatou anteriormente (cf. Quadro 1.6), a maioria dos nomes terminados em *-o* no singular e em *-i* no plural é de género masculino (*libro* 'livro' / *quadro*), sendo femininos todos os nomes terminados em *-a* no singular, e em *-e* no plural: (*casa* 'casa'; *penna* 'caneta'). No entanto, e como se viu, este critério não atua de modo uniforme, dado que também se registam casos de nomes femininos terminados em *-o* (*mano* 'mão') e, sobretudo, de masculinos terminados em *-a* (*problema*) (Acquaviva 2009;

---

<sup>61</sup> Todavia, há alguns autores (Oliphant 1997; Dardano & Trifone 1997) que consideram haver algumas relações entre determinados grupos semânticos e respetivos valores de género, à semelhança do que se verificou para o português e para o espanhol (cf. Secções 1.3.3. e 1.4.1.3.). Por exemplo, em italiano a maioria dos nomes de árvores é de género masculino, enquanto os nomes de frutos são femininos (Dardano & Trifone 1997). Porém, os grupos semânticos referidos integram um número muito reduzido de itens, sendo inúmeras as exceções dentro de um mesmo grupo. Logo, e tal como em outros idiomas românicos aqui descritos, este é um critério muito pouco produtivo para explicar a associação dos valores de género dos nomes de entidades não-sexuadas.

Caffarra *et al.* 2015). No plural, os nomes que integram estas classes temáticas terminam em *-i*, preservando o mesmo valor de género<sup>62</sup>.

Para além dos índices temáticos, há também, no italiano, sufixos derivacionais associados a valores de género específicos. Por exemplo, o sufixo aumentativo *-one* dá origem a nomes de género masculino (*il cartone* ‘o papelão’) sendo femininos os nomes com os sufixos *-sione*, *-zione* (*costruzione* ‘construção; *formazione* ‘formação’) (Dardano & Trifone 1997; Oliphant 1997:7-8).

---

<sup>62</sup> Refira-se ainda o nome *eco* ‘eco’ que, apesar de no singular se usar, preferencialmente, como um nome de género feminino, como atestam os seguintes exemplos (*un'eco* ‘um eco’; *una forte eco* ‘um forte eco’), é sempre masculino quando se encontra no plural (*gli echi* ‘os ecos’) (Dardano & Trifone 1997:170).



### 1.4.3. O sistema de atribuição de género nominal do alemão

#### 1.4.3.1. Particularidades do sistema de concordância nominal do alemão (em género, em número e em caso)

O alemão possui três valores de género em oposição: masculino, feminino e neutro. Todos os itens nominais são especificados quanto ao valor de género<sup>63</sup>, embora nem sempre este seja facilmente dedutível atendendo às suas características semânticas ou fonomorfológicas. Para além do nome, a determinação do género afeta ainda outras classes de palavras, entre as quais os pronomes, os especificadores (determinantes e quantificadores) e os modificadores (Pfau 2009:106). Os nomes e restantes elementos relacionados exibem ainda concordância em número (singular vs. plural) e em caso<sup>64</sup> (nominativo, acusativo, dativo e genitivo).

Partindo dos exemplos organizados no Quadro 1.7, com sintagmas nominais constituídos por artigo definido e nome (para o masculino *der Mann* ‘o homem’; para o feminino *die Frau* ‘a mulher’ e para o neutro *das Kind* ‘a criança’), poder-se-á constatar a complexidade de que se revestem os paradigmas de concordância nesta língua germânica:

Valores de género	Número	Caso			
		Nominativo	Acusativo	Dativo	Genitivo
Masculino	Singular	der Mann	den Mann	dem Mann	des Mannes
	Plural	die Männer	die Männer	den Männern	der Männer
Feminino	Singular	die Frau	die Frau	der Frau	der Frau
	Plural	die Frauen	die Frauen	den Frauen	der Frauen
Neutro	Singular	das Kind	das Kind	dem Kind	des Kindes
	Plural	die Kinder	die Kinder	den Kindern	der Kinder

Quadro 1.7 – Paradigmas de concordância nominal do alemão (concordância em género, número e caso)

Os artigos definidos apresentam, no nominativo e acusativo singular, três formas distintas quanto ao género: *der* para o masculino, *die* para o feminino e *das* para o neutro, permitindo, deste modo, identificar o valor de género do nome com o qual estão associados. Já no plural, as formas do artigo não diferem quanto ao género, mas apenas quanto ao caso. Ou seja, nestes casos, o plural do artigo é *die* no nominativo e no acusativo,

<sup>63</sup> Tipicamente, cada nome possui um dos três valores de género disponíveis na língua alemã. Todavia, observam-se algumas exceções, visto que certos itens são associados a mais do que um valor de género, como é o caso, por exemplo de «*Bonbon* ‘bombom’» tratado ora como neutro ora como masculino e «*Butter* ‘manteiga’» considerado, por vezes, ora como nome feminino ora como nome masculino (Bußmann & Hellinger 2003:143).

<sup>64</sup> A respeito da concordância em caso, Corbett (2006a:133) considera que este é um traço formal de concordância muito particular já que “is not an inherent feature of the noun: is imposed on the noun phrase for semantic reasons or by government by some other syntactic elements (...)”

independentemente do género do nome. No dativo e genitivo plural, as formas do artigo são *den* e *der*, respetivamente, estando associadas aos três valores de género (Salmons 1993:412).

Por sua vez, os nomes apresentam formas morfológicas distintas quanto ao número e caso. Por exemplo, para o singular, a forma *Mann* no nominativo, acusativo e dativo e a forma *Mannes* no genitivo. Já no plural, *Männer* é a forma do nominativo, acusativo e genitivo, e *Männern* é a forma do dativo.

Uma outra particularidade do sistema de concordância em género do alemão consiste no facto de nem sempre os adjetivos concordarem em género com o nome. Segundo Corbett (1991), quando os adjetivos coocorrem no sintagma nominal, apresentam formas distintas de acordo com o valor de género do item a que estão associados (cf. 13a-c), mas quando assumem uma função predicativa, esses mesmos adjetivos não concordam em género com o nome, apresentando uma única forma (cf. 14a-c), conforme atestam os exemplos seguintes recolhidos de Corbett (1991:124).

- (13) a. (*masculino*) **warmer** Tee «chá quente»  
b. (*feminino*) **warme** Milch «leite quente»  
c. (*neutro*) **warmes** Wasser «água quente»

- (14) a. (*masculino*) der Tee ist **warm** «O chá está quente.»  
b. (*feminino*) die Milch ist **warm** «O leite está quente.»  
c. (*neutro*) das Wasser ist **warm** «A água está quente.»

Também os pronomes, por vezes, não respeitam o valor de género gramatical do item nominal que retomam. Por exemplo, o neutro *Mädchen* que designa “uma menina jovem” pode ser pronominalizado com o neutro ‘*es*’ respeitando-se assim a concordância gramatical ou com a forma pronominal do feminino ‘*sie*’, aludindo-se, então, ao género natural do referente nominal (Corbett 1991:227).

#### 1.4.3.2. Critérios de atribuição de género nominal do alemão

À semelhança dos sistemas de classificação nominal baseados em género das línguas românicas selecionadas para este estudo — português, espanhol e italiano — também no alemão a atribuição dos valores de género aos nomes se pauta por um certo

grau de assystematicidade e de complexidade, resultando da interação de fatores de natureza semântica e formal, nomeadamente de aspetos relacionados com a estrutura morfológica e também fonológica dos nomes<sup>65</sup> (Zubin & Köpcke 1984:43; Corbett 1991:49).

Relativamente aos critérios de natureza semântica, nos nomes de entidades sexuadas os valores de género masculino e feminino correspondem, de um modo geral, ao género natural da entidade designada. Assim sendo, e à luz deste critério, são masculinos os itens cujo referente é do sexo masculino *der Mann* ‘o homem’, e femininos os nomes com referentes femininos como, por exemplo, *die Frau* ‘a mulher’. Todavia, nem sempre este critério é suficiente para explicar a atribuição dos valores de género destes itens, uma vez que encontramos nomes com o traço semântico [+sexuado] associados ao género neutro, como é o caso de *Mädchen* ‘a menina jovem’ e *Kind* ‘a criança’. Para alguns autores (Zubin & Köpcke 1986; Aikhenvald 2004), estes neutros poder-se-ão justificar pelo facto de os respetivos referentes corresponderem a entidades que ainda não atingiram a maturidade sexual sendo, por isso, consideradas como ‘assexuadas’ (Zubin & Köpcke 1986: 147; Aikhenvald 2004:1035). Há também nomes de entidades sexuadas cujo valor de género (masculino ou feminino) não corresponde, exatamente, ao sexo do respetivo referente, como é o caso dos itens sobrecomuns, como *Person* ‘pessoa’ que possui o valor de género feminino, independentemente do género natural da entidade designada.

Para além da relação entre género gramatical e género natural do referente [+sexuado], encontram-se, na bibliografia especializada, diferentes trabalhos que têm procurado demonstrar como a própria delimitação concetual dos nomes do alemão poderá estar correlacionada com o valor de género a eles associado (Zubin & Köpcke 1984, 1986). Com base numa análise detalhada de um acervo considerável de itens nominais e respetivas relações semânticas, Zubin & Köpcke (1984, 1986) verificaram que o valor de género neutro está tipicamente associado a itens nominais que denotam uma certa ‘vagueza’, como *Ding* ‘a coisa’, *Element* ‘o elemento’ e *Object* ‘o objeto’<sup>66</sup> (Zubin & Köpcke 1986:144).

Além disso, e salvo algumas exceções, em determinadas categorias semânticas, os autores verificam que os termos hiperónimos são neutros, enquanto os respetivos hipónimos se distribuem pelos outros valores de género disponíveis no alemão, i.e.,

---

<sup>65</sup> A este respeito, Corbett (1991:49) considera que “gender can be predicted for a large proportion of German nouns, and that there is a complex interplay of overlapping semantic, morphological and phonological factors.”

<sup>66</sup> Excetuam-se os masculinos *der Gegenstand* ‘objeto’, *der Körper* ‘o corpo’ e o feminino *die Sache* ‘a coisa’ (Zubin & Köpcke 1986:143).

masculino e feminino (cf. Zubin & Köpcke 1986). Por conseguinte, Zubin & Köpcke (1984, 1986) defendem que a atribuição dos valores de género nominal em alemão não deve ser considerada semanticamente arbitrária. No entanto, e tal como o trabalho destes investigadores tem igualmente demonstrado, nem sempre as relações estabelecidas entre os grupos semânticos analisados e respetivos valores de género (masculino / feminino / neutro) justificam a atribuição de um valor em detrimento de outros (Zubin & Köpcke 1986). Por este motivo, afirma Pfau (2009:109, n.1):

“the principles that are adduced are of such complexity that one may question their validity or at least their explanatory power.”

Considere-se, em seguida, a estrutura formal dos itens nominais do alemão, nomeadamente as suas características fonológicas e morfológicas, e em que medida tal estrutura poderá contribuir para a atribuição dos valores de género aos nomes. Nas línguas românicas aqui descritas, nomeadamente no português e no espanhol, (cf. Secções 1.3. e 1.4.1.) verificou-se uma certa correspondência, ainda que parcial, entre os índices temáticos *-o* e *-a* e os valores de género masculino e feminino, respetivamente. Por sua vez, em alemão, há determinadas terminações nominais que permitem, em parte, inferir o valor de género desse item (Salmons 1993). A este respeito, constata-se que a maioria dos itens terminados em *-e* é de género feminino<sup>67</sup>, *Tasse* ‘o copo’, *Adresse* ‘a morada’ e *Flashe* ‘a garrafa’, embora sejam masculinas as formas nominais *Name* ‘o nome’ e *Käse* ‘o queijo’, bem como alguns itens referentes a entidades de sexo masculino, como *Deutsche* ‘alemão’, prevalecendo, nestes casos, o critério semântico.

Ainda que se revista de alguma complexidade, atuam no sistema de associação dos valores de género aos nomes do alemão, critérios de natureza fonológica (Zubin & Köpcke 1984, 1986; Köpcke & Zubin 1996; Corbett 1991; Salmons 1993). De acordo com o ‘consonant cluster principle’ (*Konsonantenhäufungsprinzip*), proposto em Köpcke & Zubin (1996), grande parte dos nomes monossilábicos que possuem, em posição inicial de palavra, ataque silábico ou em coda silábica, grupos consonânticos complexos, é masculina (Köpcke & Zubin 1996:476), conforme ilustram os exemplos seguintes *Axt* ‘o machado’, *Schutz* ‘proteção’. Por sua vez, observa-se uma tendência para se associar o valor de género masculino aos nomes monossilábicos que apresentam um segmento nasal seguido de

---

<sup>67</sup> Segundo Schenke & Seago (2004:44) cerca de 90% dos itens nominais do alemão terminados em *-e* são femininos.



consoante, como por ex., *Sand* ‘areia’, e os nomes monossilábicos que em coda apresentam determinadas sequências consonânticas tendem a ser femininos, como *Luft* ‘o ar’ e *Schicht* ‘a camada’ (cf. Salmons 1993; Köpcke & Zubin 1996:477-478).

Além de haver certas terminações vocálicas e consonânticas que se poderão associar aos valores de género, há alguns constituintes morfológicos (prefixos e sufixos derivacionais) que podem igualmente condicionar a atribuição do valor de género nominal. Neste idioma, grande parte dos itens nominais apresenta uma estrutura morfológica complexa e tal estrutura pode ter implicações na atribuição do género nominal. Salmons (1993:419) distingue, entre os afixos derivacionais disponíveis no alemão, os que categoricamente influenciam o valor de género do produto nominal que originam, ou seja, os casos em que se verifica uma relação direta e inequívoca entre determinado afixo e o género do nome, e os afixos que nem sempre determinam o género do item nominal. Por exemplo, os sufixos derivacionais *-chen* e *-lein* originam nomes diminutivos de género neutro. Assim, ao adicionar-se ao nome masculino *Mann* ‘homem’ o sufixo *-chen*, cria-se um nome diminutivo de género neutro *Männchen* ‘homem pequeno’ (Corbett 1991:50). Já os sufixos *-ung*, *-heit*, *-schaft*, por exemplo, originam nomes abstratos de género feminino como, por ex., *Modellierung* ‘modelagem’ (Zubin & Köpcke 1984, 1986; Salmons 1993; Köpcke & Zubin 1996).

Na associação dos valores de género aos nomes compostos, Zubin & Köpcke (1984) defendem que nestes atua o “Last Member Principle” (1984:44), princípio segundo o qual o valor de género do item nominal composto corresponde ao valor de género do constituinte nominal localizado à direita. Por exemplo, o nome *Fischauge* ‘olho de peixe’ é neutro, sendo composto por dois elementos nominais, o masculino *Fisch* ‘peixe’ e o neutro, localizado à direita, *Auge* ‘olho’ (Kilarski 2013:18). Contudo, em alguns casos particulares, este princípio não se verifica. Tal facto pode resultar da sobreposição de critérios<sup>68</sup>, nomeadamente de critérios morfológicos, porque, em certos nomes, a presença de um determinado prefixo pode determinar o respetivo valor de género, tal como observa Zubin & Köpcke (1984:45) ao referirem que:

“*der Bereich* ‘region’ and *der Verhaft* ‘arrest’ have masc-gender on the basis of gender-determining prefixes *be-* and *ver-*, although the last member of the unit has

---

<sup>68</sup> Zubin & Köpcke (1984) observam ainda que em determinadas variedades dialetais do alemão parece haver, em casos pontuais, uma prevalência do critério semântico para a associação dos valores de género nominal sobre as propriedades morfológicas do nome.

neut-gender in the first instance (*das Reich* 'kingdom') and fem-gender in the second (*die Haft* 'arrest').”.

## 1.4.4. O sistema de classificação nominal do inglês

### 1.4.4.1. Especificidades relativas ao sistema de classificação nominal do inglês

Como se verificou na parte introdutória deste capítulo, nem todas as línguas descendentes do indo-europeu preservaram um sistema de classificação dos nomes baseado em valores de género gramatical (cf. Secção 1.2. e Figura 1.1) e, no caso específico da língua inglesa, apenas se conservaram as marcas de um primitivo sistema de género ao nível dos pronomes (pessoais, possessivos e reflexos). Neste idioma, a especificação de género resulta, sobretudo, da aplicação de critérios semânticos, no sentido em que, para retomar nomes de entidades humanas masculinas ou femininas, se utilizam os pronomes *he* ‘ele’ e *she* ‘ela’, respetivamente, ao passo que entidades não humanas são referidas através do pronome neutro *it*. Há, então, uma certa distinção nos pronomes entre o traço [+animado], por um lado, e [-animado], por outro.

Alguns autores (Quirk *et al.* 1985; Corbett 1991; Siemund & Dolberg 2011; Kibort & Corbett 2008) entendem que o inglês exibe um sistema de género — ainda que com configurações muito particulares —, havendo, porém, outros investigadores (Ibrahim 1973; Aikhenvald 2000) que rejeitam essa posição. Vejamos, ainda que de forma muito breve, os argumentos que sustentam as duas posições.

Na descrição da gramática do inglês, Quirk *et al.* (1985) consideram que esta língua apresenta uma categoria semântica de género, especificada no sistema pronominal<sup>69</sup>. Da mesma forma, Corbett (1991), ao assumir que o processo de retoma anafórica pronominal está sob o escopo da definição do conceito de concordância sintática (Corbett 2006a; veja-se ainda a Secção 1.2.1.1.), defende que a língua inglesa exibe um sistema de género particular, que define como pronominal<sup>70</sup> (Corbett 1991:169).

---

<sup>69</sup> Também Foundalis (2002) considera que o inglês exibe, à semelhança do tamil, um sistema de género baseado em critérios de natureza semântica, sendo muito escassos os casos em que tais critérios não se observam.

<sup>70</sup> Também Siemund & Dolberg (2011), baseados na proposta de Dahl (2000), distinguem as noções de género lexical e de género referencial. O género referencial corresponde a uma propriedade estritamente semântica, na qual as categorias são determinadas em função das propriedades semânticas dos respetivos referentes nominais. Por sua vez, o género lexical corresponde a uma propriedade essencialmente formal, sendo a atribuição dos valores de género arbitrária e idiossincrática. Assim, segundo esta perspetiva, os autores (2011:492) defendem que os pronomes “[a]s deictic elements may reflect the properties of their referent and in this case need not agree with a controlling noun. Hence, it is not uncontroversial to view them as agreement targets (...) (2) a. *He: John, man, boy*, etc. b. *She: Mary, woman, girl*, etc.; *It: stone, table, water, grass*, etc.”

Todavia, para alguns investigadores não é defensável considerar o processo de retoma anafórica pronominal sob o escopo da concordância sintática. Segundo Aikhenvald (2000:21), no inglês, os três ‘valores’ de género visíveis ao nível dos pronomes de 3ª pessoa do singular — masculino (*he*), feminino (*she*) e neutro (*it*) — estipulados em função da animacidade e do sexo dos referentes nominais que retomam, desempenham uma função meramente anafórica. Ora, sendo certo que os sistemas de género gramatical se caracterizam por desencadearem operações sintáticas de concordância, numa manifestação reiterada de traços formais nos elementos que estabelecem entre si este tipo de relações, a autora considera inadequado assumir a anáfora como um fenómeno de concordância sintática (Aikhenvald 2000:29).

Do mesmo modo, Ibrahim (1973) considera que o inglês não é uma língua com género gramatical. Este autor defende ainda que o desaparecimento da categoria se justifica pelo facto de o inglês ter perdido, de forma gradual, as características de uma língua de morfologia rica (1973:86). Por conseguinte, importa considerar a evolução histórica da língua inglesa para compreender o modo como, progressivamente, as marcas de concordância dos valores de género gramatical foram desaparecendo do tecido gramatical do inglês (cf. Siemund & Dolberg 2011).

Remontando ao período do inglês antigo, compreendido entre o século V — com a chegada das tribos germânicas à Inglaterra — e o século XI — aquando da Conquista Normanda —, a língua possuía uma estrutura flexional forte, comparável à do latim. Com efeito, exibia diferentes paradigmas flexionais, observáveis pela especificação de três valores de género distintos (masculino, feminino e neutro)<sup>71</sup>, pela oposição de número (singular, plural e ainda com vestígios do dual oriundo do indo-europeu) e pela marcação de quatro casos nominais. Para além disso, neste período histórico da língua, diferentes classes de palavras, tais como adjetivos, determinantes e verbos, ostentavam marcas morfológicas de concordância (Hellinger 2001:106).

Foi a partir da Conquista Normanda, no século XI, que assinala o início do período medieval da língua, que se prolonga até ao século XVI, que se observa uma progressiva

---

<sup>71</sup> Apesar de os nomes serem especificados quanto ao valor de género, já se observava neste período uma certa ambiguidade, visto que alguns itens, essencialmente os que designavam entidades não-animadas não exibiam marcas formais explícitas do seu valor, podendo estar, por esse motivo, associados a mais do que um valor de género. Deste modo, e como afirma Ibrahim (1973:87), existiam formas nominais associadas, quer ao masculino, quer ao feminino, tais como, por ex., *dic* ‘canal’, *heaven* ‘céu’, enquanto outras eram simultaneamente masculinas e neutras, como *hyll* ‘monte’ e *frip* ‘paz’.

redução dos paradigmas flexionais (Siemund & Dolberg 2011)<sup>72</sup>. Com a invasão normanda, a língua francesa passa a ser dominante, sobretudo nas classes sociais mais elevadas, introduzindo-se no inglês um conjunto avultado de palavras do francês. Consequentemente, e uma vez que estes estrangeirismos possuíam estruturas morfológicas e fonológicas distintas, a especificação dos valores de género de acordo com os paradigmas já estabelecidos foi praticamente impossível (Ibrahim 1987:87).

Além disso, a Conquista Normanda contribuiu para o acelerar de um conjunto de mudanças que já se observavam no período do inglês antigo. Entre as diferentes alterações, Ibrahim (1973) destaca a que corresponde à passagem de *m* final a *n* e a posterior queda de *n* que levou, progressivamente, a uma redução de formas flexionais em *-a*, *-u*, *-e*, *-an*, *-um*, a uma única forma em *-e* (Ibrahim 1973:87). Para demonstrar o modo como tais alterações afetaram a especificação de género nominal da língua inglesa, Ibrahim refere exemplos de itens nominais com o traço semântico [+animado] e que, durante a época medieval, apresentavam duas formas, i.e., uma masculina para designar a entidade de sexo masculino (*nefa* ‘sobrinho’, *maga* ‘parente de sexo masculino’) e outra feminina para designar o ser de sexo feminino (*nefe* ‘sobrinha’, *mage* ‘parente de sexo feminino’). Com o avançar das alterações fonéticas acima referidas, tais itens nominais deixaram de se poder distinguir.

As mudanças ao nível da estrutura fonética afetaram igualmente outras classes de palavras. Os adjetivos, divididos em declinações fortes e fracas, tiveram as suas formas reduzidas a uma única não flexional. Também os artigos definidos e os demonstrativos, flexionados no inglês antigo e com concordância em género, número e em caso com o nome, surgiram no período medieval com formas idênticas para todos os valores de género (Ibrahim 1973:88). Quanto ao sistema pronominal, os dados históricos disponíveis sugerem que este sistema foi resistindo às mudanças ocorridas na língua, prevalecendo as formas que espelham uma distinção natural dos valores de género, i.e., baseadas em propriedades semânticas dos referentes que retomam, nomeadamente no que concerne ao carácter [+animado] e também [+humano]: *he* para nomes de referentes animados e humanos de sexo masculino; *she* para nomes de referentes animados e humanos de sexo feminino; e uma forma neutra, *it*, para nomes de referentes não animados e não humanos.

Ao considerar os processos de evolução dos diferentes tipos de classificação nominal, Aikhenvald (2000) constata que muitos sistemas de género gramatical podem

---

<sup>72</sup> Com efeito, e salvo raros casos, já no período do inglês medieval a marcação de género nos sintagmas nominais plurais não se verificava (Siemund & Dolberg 2011:516).

sofrer um processo de modificação, passando de uma estrutura dita mais gramatical a outra mais semântica. E se, numa dada língua, se regista uma perda gradual da marcação de concordância sintática dos valores de género, verifica-se ainda que o sistema pronominal é, em última instância, o menos afetado por estas modificações. Consequentemente, os valores de género que subsistem no sistema pronominal serão semânticos, mesmo que não o tenha sido no passado (Aikhenvald 2000:409; veja-se também Corbett 1991:248-260).

No que concerne à estrutura morfológica dos itens lexicais do inglês, constata-se que, em alguns nomes com referentes [+animados], há casos em que a designação da entidade de sexo masculino e a designação da entidade de sexo feminino é feita através de itens lexicais distintos, como *father* ‘pai’ e *mother* ‘mãe’. Há também alguns itens que possuem, na sua estrutura morfológica, uma especificação que permite distinguir o item nominal que designa a entidade de sexo masculino do item que designa a entidade correspondente de sexo feminino, como, por ex., *hero* ‘herói’ e *heroine* ‘heroína’<sup>73</sup> (Quirk *et al.* 1985:315).

Existem ainda alguns nomes de entidades [+animadas] com uma única forma lexical, quer para designar o ser de sexo masculino, quer para designar o ser de sexo feminino, como *student* ‘estudante’, *doctor* ‘doutor(a)’, tratando-se de casos com um comportamento equivalente ao verificado em outras línguas nos nomes comuns de dois<sup>74</sup>.

Em suma, e apesar de se verificar no sistema pronominal da língua inglesa a oposição de três valores de género — masculino (*he*), feminino (*she*) e neutro (*it*) — que reflete, essencialmente, uma oposição de natureza semântica, assumimos neste trabalho que tal oposição não é suficiente para se considerar que nesta língua há um sistema de género gramatical comparável ao que existe noutros idiomas, nos quais se observa o mecanismo de concordância sintática, desencadeado pelo valor de género do nome.

---

<sup>73</sup> Segundo Quirk *et al.* (1985:315), observa-se uma certa relação de natureza derivacional entre este tipo de itens, apesar de os sufixos que atuam nestes casos serem muito pouco produtivos.

<sup>74</sup> A pronominalização de certos nomes de entidades sexuadas poderá estar condicionada com um certo *status* social associado aos referentes prototípicos. A este propósito, Hellinger (2001) constata que existem alguns nomes, de profissões consideradas “de excelência” como, por ex., *lawyer* ‘advogado(a)’, *physician* ‘físico’, *scientist* ‘cientista’ que nos contextos em que o sexo do referente não é conhecido ou é irrelevante, se utiliza o pronome *he*, o mesmo não acontecendo em outros títulos de profissões tais como *secretary* ‘secretário(a)’ ou *nurse* ‘enfermeiro(a)’ em que, muitas vezes, a retoma anafórica é feita através do pronome *she*, até porque, tradicionalmente, são profissões exercidas por mulheres. Assim, e de acordo com a autora, este fenómeno ilustra uma categoria de género social. Ora, e nas palavras da investigadora, esta mesma categoria “has to do with stereotypical assumptions about what are appropriate social roles for women and men, including expectations about who will be a typical member of the class of, say, *surgeon* or *nurse*. Deviations from such assumptions will often require formal markings, for example by adjectival modification: *female surgeon* or *male nurse*.” (Hellinger 2001:108).

Consideremos, em seguida, as características do sistema pronominal do inglês, tendo em conta as especificidades associadas ao processo de retoma anafórica.

#### 1.4.4.2. O sistema pronominal do inglês

Conforme referido anteriormente, o inglês preserva marcas de um primitivo sistema de género gramatical nos pronomes — pessoais, possessivos, reflexos (3ª pessoa do singular).

Nos pronomes pessoais de 3ª pessoa do singular do inglês distinguem-se três valores — masculino (*he*), feminino (*she*) e neutro (*it*) —, sendo que no plural esta distinção deixa de estar presente. A especificação destes valores resulta, sobretudo, da aplicação de critérios semânticos, correlacionados com a natureza [+humana] e [+animada] do referente nominal. Assim, para se proceder à retoma de nomes de entidades humanas masculinas (por ex. *man* ‘homem’) ou femininas (por ex. *woman* ‘mulher’) utilizam-se os pronomes *he* ‘ele’ e *she* ‘ela’, respetivamente, ao passo que as entidades não humanas, [+animadas] (*horse* ‘cavalo’) ou [-animadas] (*house* ‘casa’) são referidas através do pronome neutro *it*.

Existe, porém, alguma variabilidade no uso destas formas pronominais. Com efeito, na referência a alguns animais domésticos, i.e., entidades [-humanas] e [+animadas], é possível distinguir o animal de sexo masculino (por exemplo, *dog* ‘cão’) do correspondente de sexo feminino (*bitch* ‘cadela’) através dos pronomes *he* ou *she*, respetivamente<sup>75</sup>. Há ainda alguns nomes com referentes não-animados que admitem a retoma anafórica pelos pronomes *he* ou *she*, como é o caso de alguns «boat nouns» como *ship* ‘barco’ que se associa ao pronome pessoal de 3ª pessoa feminino<sup>76</sup> (Corbett 1991:180).

Nos pronomes relativos (*who*, *which*) e em alguns indefinidos (*somebody*, *something*), não é possível a especificação do sexo da entidade designada, sendo que se estabelece uma oposição entre a retoma anafórica de nomes com referentes [+humanos] e a retoma de nomes que possuem o traço semântico [-humano] (cf. Quirk *et al.* 1985:341).

---

<sup>75</sup> A propósito desta particularidade, refere Corbett (1991:12) que “[d]omestic animals, particular if they are named, are masculine or feminine according to sex; and, especially in children’s stories, many animals have particular gender by convention.”

<sup>76</sup> Para Quirk *et al.* (1985:318), a utilização de um pronome pessoal tipicamente associado a entidades [+animadas] para retomar nomes de entidades [-animadas] é uma forma de o falante expressar uma certa relação afetiva para com o objeto referido.

Atendendo à aplicação de critérios semânticos e à distribuição das formas pronominais do singular, Corbett (1991:180) considera que existem, no inglês, três padrões de correferência pronominal<sup>77</sup>, nomeadamente:

- (i) *who / he* — retomam nomes que denotam entidades [+humanas] (*man*) e entidades [+animadas] (*dog*) do sexo masculino;
- (ii) *who / she* — permitem a referência a itens nominais com referentes [+humanos] (*woman*) e [+animados] (*cow*) do sexo feminino; e
- (iii) *which / it* — referem-se a nomes que possuem o traço semântico [-animado] (*house*).

Note-se, contudo, que há alguns itens nominais muito particulares cuja retoma anafórica pronominal nem sempre segue estes padrões. Por exemplo, nomes como *baby* ‘bebê’ e *child* ‘criança’ admitem as três possibilidades de correferência pronominal, i.e., através do relativo *who* e dos pronomes *he*, *she*, e através do relativo associado a nomes com referentes não-animados *which* e o pronome *it*. A retoma anafórica através dos pronomes *who* e *he/she* poder-se-á justificar atendendo à natureza [+humana] do referente nominal, sendo assim possível estabelecer uma distinção relativamente ao sexo do referente. Em contrapartida, estes nomes designam entidades que, por ainda não terem atingido maturidade sexual, serão consideradas ‘assexuadas’, o que, por sua vez, pode conduzir à preferência pelo padrão em que não é possível a especificação relativamente ao sexo da entidade designada, ou seja, através do relativo *which* e pronome pessoal *it*, tratando-se, então, de uma situação idêntica à verificada no alemão, a propósito do neutro *Mädchen* (cf. Secção 1.4.3.2.).

Como acima referido, nomes como *ship* ‘barco’, apesar de designarem entidades [-humanas], admitem, por vezes, a retoma anafórica através do pronome *she*. Todavia, estes itens nunca admitem a retoma pelo pronome relativo *who*, como atestam os exemplos (15)

---

<sup>77</sup> Já Quirk *et al.* (1985:314), a propósito da descrição da gramática da língua inglesa, admitem que este idioma possui nove categorias de género semântico, estabelecidas em função do carácter [ $\pm$ humano] dos referentes nominais e em função das múltiplas possibilidades de correferência pronominal no singular admitidas pelos nomes. Todavia, Corbett (1991:161) considera que a proposta dos gramáticos é insatisfatória, visto que se baseia em padrões de correferência pronominal muito particulares, não permitindo dar conta das generalizações que efetivamente se observam na língua inglesa. Para além disso, o autor considera que, ao se adotar esta perspectiva tão particularizada do fenómeno, se transmite erradamente a ideia de que muitos dos sistemas de classificação nominal baseados em género, tipicamente considerados similares, são divergentes.



e (16) recolhidos em Corbett (1991:181). Ou seja, o traço [ $\pm$ humano] parece ser mais forte e determinante para a seleção do pronome relativo a usar.

(15) I sailed on the Canberra. She's a fine ship.

(16) \*The Canberra, who is a fine ship...

Assim sendo, apesar de se evidenciar no inglês uma certa correlação entre algumas formas pronominais e as propriedades semânticas dos itens nominais que retomam, com a distinção, por exemplo, de três formas de pronomes pessoais para o singular — masculino, feminino e neutro —, e apesar de haver alguns itens nominais com formas lexicais distintas que refletem, precisamente, uma oposição de género natural das entidades designadas, como *boy* ‘rapaz’ e *girl* ‘rapariga’, a verdade é que estas características verificadas na estrutura gramatical do inglês em muito se distanciam das configurações de outras línguas previamente analisadas, como o português (Secção 1.3.), o espanhol (Secção 1.4.1.), o italiano (Secção 1.4.2.) e o alemão (Secção 1.4.3.).

Além disso, e como se pôde verificar, ao longo da sua evolução histórica, a língua inglesa foi perdendo traços característicos dos idiomas de morfologia rica, sendo notória a simplificação dos paradigmas flexionais e a conseqüente perda do esquema de concordância. Este fenómeno contribuiu, assim, para o desaparecimento gradual de um sistema de género.

Por fim, nesta língua germânica o género corresponde, essencialmente, a uma categoria de natureza lexical. Poder-se-á assim postular que os falantes deste idioma conceberão o género como uma categoria fortemente interligada com a noção de sexo dos referentes nominais, porque, por exemplo, os possessivos *his* e *hers* são quase exclusivamente utilizados para a referência do possuidor, entidade de sexo masculino ou feminino, respetivamente, enquanto o possessivo *its* serve, sobretudo, para designar o possuidor, entidade não animada e, conseqüentemente, assexuada.



## 1.4.5. O sistema de classificação nominal do chinês (mandarim)

### 1.4.5.1. Características da estrutura formal do chinês

O chinês<sup>78</sup>, idioma da família das línguas sino-tibetanas, é falado por um elevado número de falantes nativos, evidenciando uma multiplicidade dialetal única. Face à variedade de dialetos existentes no país, foi necessário instituir uma norma-padrão, o que ocorreu no ano de 1955, altura em que o Governo da República Popular da China decretou o mandarim como língua oficial. O mandarim corresponde a um dos idiomas mais falados no Norte da China, mais concretamente na capital, Beijing (Pequim), uma cidade cuja importância política e cultural se observou durante séculos (Li & Thompson 1981:1).

A língua chinesa apresenta substanciais diferenças em relação às restantes línguas que foram descritas nas secções anteriores deste capítulo. Em primeiro lugar, o chinês é uma língua do tipo isolante, em que a maioria das palavras é constituída apenas por um só morfema, sem ser possível a sua segmentação em unidades morfológicas de nível inferior (Li & Thompson 1981:11). Além disso, se línguas como o português, o espanhol e o alemão exibem marcas morfológicas no léxico que permitem distinguir, em grande parte dos casos, as várias classes de palavras entre si, ou seja, se se trata de um nome, verbo, adjetivo, advérbio, etc., tal não é possível em chinês, visto que uma mesma forma de um item pode, dependendo do contexto, ter a função de nome ou de verbo, por exemplo (Norman 1988:87).

De acordo com a tradição gramatical chinesa<sup>79</sup>, e atendendo ao seu conteúdo semântico, as palavras classificam-se como “abstratas” (*xū cí*) quando não exibem um significado específico, e “concretas” (*shí cí*). As primeiras formam uma classe fechada,

---

<sup>78</sup> Na República Popular da China coexistem inúmeros dialetos, destacando-se o mandarim, oriundo do Norte do país que é, além disso, a língua-oficial, e o cantonês, utilizado, sobretudo, nas províncias de Guanghong, Guangxi e Hainan, além das ilhas de Hong Kong e de Macau. Apesar de haver alguns pontos divergentes entre si, não existem substanciais diferenças no que respeita ao sistema de classificação nominal. Por conseguinte, e atendendo ao facto de a bibliografia existente tratar, sobretudo, o mandarim, neste trabalho ao referirmos “chinês” e “língua chinesa” dever-se-á ter em conta que a descrição incide essencialmente na estrutura do mandarim. É também de ressaltar que, no *corpus* desta investigação, encontra-se um conjunto de informantes cuja LM é o chinês, sendo que este conjunto integra falantes nativos de mandarim e falantes nativos de cantonês, uma vez que nem sempre os aprendentes especificaram, ao preencherem a ficha de identificação, se a respetiva LM correspondia ao mandarim ou ao cantonês. Note-se ainda que, na China, todos os estudantes universitários dominam também o mandarim.

<sup>79</sup> Saliente-se que apesar de o chinês ser uma língua com uma longa e vasta história, a verdade é que o estudo da sua gramática é ainda muito recente, datando dos inícios do século XIX, e fortemente influenciado pela tradição gramatical ocidental (Norman 1988:152).

cuja principal função é a de estabelecer várias relações gramaticais, e da qual fazem parte, por exemplo, as preposições e os advérbios (Norman 1988:158), enquanto as segundas integram uma classe aberta, constituída por nomes, pronomes, verbos, adjetivos, etc.<sup>80</sup> (Norman 1988:157).

Neste idioma não se verifica o fenómeno da concordância sintática, quer verbal, quer nominal, característico de línguas de morfologia rica, como é o caso do português e do alemão, por exemplo, havendo outro tipo de soluções para expressar determinadas relações gramaticais<sup>81</sup> (Norman 1988). Assim, e como seria expectável dado o seu perfil, o chinês não exhibe valores de género gramatical morfologicamente marcados, i.e., os nomes e pronomes não apresentam informação relativa ao género, nem outras classes de palavras estabelecem relações de concordância sintática com o item nominal (Li & Thompson 1981).

Há, no entanto, algumas formas lexicais com referentes que possuem o traço semântico [+animado], em que se diferencia o nome que designa o ser de sexo masculino do nome que designa a entidade de sexo feminino (Etnner 2002:37). Efetivamente, existem determinados itens nominais de entidades humanas em que, na escrita, é possível especificar o sexo do respetivo referente através da anteposição de um carácter particular. Assim, no caso de nomes com referente de sexo masculino antepõe-se o carácter 男 *nán* ‘macho’ e, quando se trata de um item nominal cujo referente é do sexo feminino, utiliza-se 女 *nǚ* ‘fêmea’, obtendo-se, desta forma, um nome composto. Por exemplo, ao item nominal *rén* ‘pessoa’, representado ortograficamente por 人, antecede-se 男 *nán*, criando-se, desta forma, o composto 男人 *nán-rén*, que designa o ‘homem, pessoa do sexo masculino’, e, com o carácter 女 *nǚ*, obtém-se a forma composta 女人 *nǚ-rén*, para designar a ‘mulher, pessoa do sexo feminino’. Quando não é possível saber o sexo do referente nominal com o traço semântico [+humano], utiliza-se, simplesmente, o carácter 人 *rén*. No caso de nomes

---

<sup>80</sup> Influenciado pela tradição gramatical ocidental, Mǎ Jiànzhōng foi o primeiro gramático que procurou, de modo sistemático, distinguir as classes de palavras na estrutura do chinês. Desde então, este tem sido um dos tópicos mais discutidos entre os gramáticos e linguistas chineses, havendo uma multiplicidade de propostas. Veja-se, a este respeito, (Norman 1988:157).

<sup>81</sup> Por exemplo, não sendo possível marcar no verbo informação relativa ao tempo, este valor é apenas depreendido através do contexto da frase ou da presença de expressões adverbiais, tais como *jīntiān* ‘hoje’, *qùnián* ‘no ano passado’, *xiànzài* ‘agora’, etc. (Li & Thompson 1981:321). Todavia, há alguns morfemas que, combinados com o verbo, permitem marcar informação de natureza aspetual como, por exemplo, o auxiliar *le* que permite marcar uma ação completa ou o aspeto perfeitivo da ação realizada (Li & Thompson 1981:185).

de entidades [-humanas], para designar a espécie de sexo masculino, emprega-se o carácter 公 *gōng* ‘macho’ e, para a espécie de sexo feminino, recorre-se ao carácter 母 *mǔ* ‘fêmea’ (Ettner 2003:36-37).

Veja-se, no Quadro 1.8, alguns exemplos de itens nominais em que é possível especificar, na escrita, o valor semântico de sexo (macho ou fêmea) do respetivo referente nominal:

Nome referente [+ humano]	Caracteres usados na especificação do sexo do referente nominal	
	男 <i>nán</i> (macho)	女 <i>nǚ</i> (fêmea)
<i>xué-shēng</i> ‘estudante’	<i>nán-shēng</i> ‘estudante do sexo masculino’	<i>nǚ-shēng</i> ‘estudante do sexo feminino’
<i>báizǐ</i> ‘criança’	<i>nán-báizǐ</i> ‘rapaz / criança do sexo masculino’	<i>nǚ-báizǐ</i> ‘criança do sexo feminino’

**Quadro 1.8** – Nomes antecidos de caracteres que permitem a especificação do sexo do referente [+humano] (exemplos recolhidos de Ettner (2002:37))

Este sistema particular aproxima-se do que se verifica, no caso dos nomes sobrecomuns, em algumas línguas românicas que, como vimos, recorrem às formas ‘macho’ ou ‘fêmea’ para criar itens nominais compostos para a especificação do sexo do referente do item nominal sem, contudo, se alterar o respetivo valor de género gramatical.

Relativamente ao sistema pronominal, constata-se que os pronomes de segunda e terceira pessoas do singular podem exibir informação relativa ao sexo do referente, embora esta ocorra só no registo escrito<sup>82</sup>, não havendo, na oralidade, a possibilidade de se estabelecer esta distinção (Norman 1988:157). Mais especificamente, na segunda pessoa do singular, para referir uma entidade do sexo masculino, utiliza-se o carácter 你 ‘*nǐ*’ ‘tu’ e, para referir uma entidade de sexo feminino, utiliza-se 妳 ‘*nǐ*’<sup>83</sup>; já na terceira pessoa do singular, o carácter 他 ‘*tā*’ é usado para designar as entidades de sexo masculino ou em contextos onde não é possível saber-se o sexo do referente<sup>84</sup>, e o carácter 她 ‘*tā*’ é usado

<sup>82</sup> A este respeito, Ramsey (1987:50-51) observa que: “[a]round the late 1910s writers began to use three separate graphs to distinguish ‘he’, ‘she’, and ‘it’. All three of these characters were pronounced exactly the same way, *tā*, since there was (and is) only one third-person pronoun in Mandarin. But the use of separate graphs gave the illusion that Chinese had the same kind of gender distinction that many Western languages did.”

<sup>83</sup> É preciso, no entanto, referir que a utilização destes caracteres é diferente nos dois padrões de escrita vigentes na China, ou seja, no chinês simplificado e no chinês tradicional. Segundo foi possível apurar, por indicação dada por uma falante nativa chinesa, professora de português como língua não materna numa universidade chinesa, só no chinês tradicional, utilizado, sobretudo, em Hong Kong, Macau e Taiwan, é que se regista o emprego de dois caracteres distintos, o 你 para se referir a entidades do sexo masculino e 妳 para designar uma pessoa do sexo feminino. No chinês simplificado utiliza-se apenas o carácter 你 para referência, quer a pessoas de sexo masculino, quer a pessoas de sexo feminino.

<sup>84</sup> Nas situações em que os falantes não conseguem saber, ao certo, o sexo do referente assumem, por defeito, que os pronomes de 2ª e 3ª pessoas, 你 ‘*nǐ*’ e 他 ‘*tā*’, respetivamente, designam pessoas do sexo

para referir seres de sexo feminino, utilizando-se também o carácter 它 ‘*tā*’ para referentes não animados ou [-humanos].

Para além da ausência da categoria de género gramatical morfológicamente codificada, o chinês também não exhibe a categoria morfológica de número. De facto, a marcação explícita do número nos pronomes só é possível através do uso da forma *men*. Assim, por exemplo, o pronome de 3ª pessoa do plural ‘eles’ corresponde a ‘*tāmen*’ (Norman 1988: 159), sendo esta uma forma que não admite a especificação do sexo do seu referente (Ettner 2002:37). Em certos casos, *men* pode ainda ser agregado a nomes com o traço semântico [+humano], obtendo-se formas que, além de não serem obrigatórias em todos os contextos e de não poderem surgir associadas a numerais, denotam um sentido de coletividade, como *hāizimen* ‘(um certo grupo) de crianças’ (cf. Norman 1988:159). O conceito de pluralidade é, portanto, transmitido através de expressões de quantificação que exigem, e como se verá em seguida, a presença de classificadores (Norman 1988:157).

#### 1.4.5.2. O sistema de classificadores do chinês: características gerais

Em chinês, o nome pode ser modificado por *demonstrative-measure compounds*, i.e., por expressões compostas por demonstrativos e classificadores (Norman 1988:157). Em contextos de quantificação, alguns itens nominais<sup>85</sup> admitem a coocorrência de constituintes designados como classificadores numerais<sup>86</sup> (Aikhenvald 2000; Grinevald 2000) (veja-se ainda Secção 1.2.2. do presente trabalho). Tal como a própria designação sugere, estes classificadores ocorrem tipicamente associados a numerais, uma vez que nem

---

masculino. O mesmo acontece com o nome de pessoa 人 ‘*rén*’ que é utilizado quando não se sabe o sexo de determinada entidade, entendendo o falante que o referente é do sexo masculino (Ettner 2002:36). Assim, e nas palavras de Ettner (2002:36), “*nǐ, tā* and *rén* have a supposed generic quality in oral contexts (...) but each connote the male gender by default”.

<sup>85</sup> De facto, nem todos os nomes admitem a presença de um classificador como é o caso de *tiān* ‘dia’ visto que a significação deste item já tem implícita a ideia de uma certa noção de ‘extensão’ e de ‘quantidade’ (cf. Li & Thompson 1981:105).

<sup>86</sup> Na bibliografia disponível, observa-se uma certa imprecisão relativamente à delimitação concetual do termo ‘classificadores’. Ou seja, para alguns autores, o elemento localizado à esquerda do nome e antecedido por numeral e/ou demonstrativo é referido como *classifier*, enquanto outros preferem o termo *measure word*. J. Norman (1988) considera que os classificadores correspondem a *measures words* dado que “measures are bound morphemes that follow either numerals or demonstratives” (1988:157). O mesmo acontece em Chao (1968) que prefere o termo *measures* a *classifiers*. Por sua vez, Li & Thompson (1981) tratam o conceito de *measure* sob o escopo dos classificadores, referindo que “any measure words can be a classifier” (1981:106). No presente trabalho, preferimos, por questões de natureza terminológica, o termo *classificador*.

sempre o numeral pode surgir imediatamente antes do nome<sup>87</sup>. Por exemplo, a seguinte sequência de numeral e de nome é considerada agramatical \**liang xueshēng* ‘dois estudantes’, sendo que, para se tornar gramatical, é obrigatória a inserção do classificador ‘ge’ *liang ge xueshēng*<sup>88</sup>. Há ainda outros determinantes (como os demonstrativos *nèi* ‘aquele’, *zhèi* ‘este’) e quantificadores (*zhěng* ‘todo’, *mǒu yī* ‘um certo’<sup>89</sup>) que exigem, geralmente, a presença de um classificador numeral<sup>90</sup> (Li & Thompson 1981:104).

Do ponto de vista semântico, os classificadores numerais podem estar associados, quer a itens nominais contáveis, quer a itens massivos, possibilitando-se, assim, a expressão da sua contabilização e medição (Li & Thompson 1981:105-113). Com efeito, os classificadores possuem diferentes especificidades em função das características semânticas dos respetivos itens nominais que modificam.

Atendendo à origem e ao significado veiculado, Chao propõe, na sua gramática, uma tipologia dos classificadores da língua chinesa<sup>91</sup>, com uma descrição exaustiva de cada tipo proposto (1968:619). Ainda de acordo com esta classificação, constata-se que existem alguns nomes que requerem classificadores particulares ou individuais<sup>92</sup> (Chao 1968:619),

---

<sup>87</sup> Há, contudo, determinados casos em que é possível um numeral anteceder o nome, sem qualquer elemento entre eles. Estes casos registam-se, sobretudo, na oralidade e em alguns ditos e expressões populares, como, por ex.: “九牛二虎之力” (*jiǔ niú èr hǔ zhī lì*) que se traduz por “A força de nove bois e dois tigres” e é utilizada quando se pretende transmitir a ideia de que para a concretização de determinado objetivo é necessário ter muita força física e/ou psicológica. Nesta frase temos os numerais *jiǔ* ‘nove’ e *èr* ‘dois’ que antecedem os nomes *niú* ‘boi’ e *hǔ* ‘tigre’, respetivamente.

<sup>88</sup> Exemplos recolhidos de Li (2013:13).

<sup>89</sup> A expressão *mǒu yī* ‘um certo’ exige a presença de um classificador uma vez que *yī* corresponde ao numeral ‘um’. Assim, quando *mǒu* ‘certo’ é utilizado sem o numeral *yī*, não exige a presença de um classificador, como se pode verificar nas seguintes expressões: *mǒu dì* ‘certo lugar’, *mǒu nián* ‘certo ano’.

<sup>90</sup> Todavia, nem sempre os demonstrativos *nèi* e *zhèi* exigem a presença de um classificador, como atestam os seguintes exemplos de formas coloquiais frequentemente utilizadas no chinês: *zhèi shū* ‘este livro’, *zhèi yǐzi* ‘esta cadeira’ e *nèi rén* ‘aquela pessoa’.

<sup>91</sup> Segundo a proposta de Chao (1968:619), existem, no chinês, os seguintes tipos de classificadores: (i) individuais, requeridos por itens nominais específicos; (ii) coletivos, que modificam o nome adicionando-lhe o sentido de ‘reunião / conjunto de’; (iii) partitivos, que denotam uma parte do todo referido pelo item nominal; (iv) de conteúdo, que correspondem a formas nominais de conteúdo usadas, por vezes, como classificadores; (v) provisórios, que fornecem, de modo pouco preciso, a informação relativa à quantidade de um determinado objeto; e (vi) de medida, que correspondem a unidades de medida como *metro*, *quilo*. Entre os classificadores numerais, Chao (1968:593) refere-se ainda a um subconjunto muito particular, utilizado com verbos transitivos. Há ainda um subconjunto de classificadores verbais associados a verbos de ação e cuja função é a de indicar a frequência de ocorrências da ação denotada pela forma verbal (1968:615). Por fim, o autor identifica um outro tipo de classificadores numerais, designados como *quasi-measures* ou *autonomous measures* (1968:608) que, ao contrário dos outros classificadores descritos, possuem uma maior autonomia no que respeita às possibilidades de ocorrência.

<sup>92</sup> De facto, há alguns nomes que exigem um classificador individual, pese embora o facto de ser cada vez mais frequente na língua a generalização de uso do classificador *ge*, mesmo em contextos em que um classificador individual seria exigido (Li & Thompson 1981).

nem sempre sendo fácil identificar que propriedades semânticas dos itens conduzem à preferência de determinado classificador em detrimento de outros<sup>93</sup>. Por exemplo, o classificador *tiáo* modifica nomes referentes a objetos ou seres alongados, tais como *lù* ‘estrada’, *shé* ‘cobra’, *yú* ‘peixe’, modificando igualmente alguns nomes de animais mamíferos quadrúpedes. Todavia, certos nomes de objetos alongados, como *máobǐ* ‘escova do cabelo’ e *jiàn* ‘seta’ não admitem o classificador *tiáo*, associando-se, por sua vez ao classificador *zhī*. Já itens como *xīnwén* ‘notícia’ ou *fǎlǐ* ‘lei’ admitem, ao contrário do que seria de esperar, e dadas as características semânticas do referente, o classificador *tiáo* (Li & Thompson 1981:112).

Um dos maiores problemas relativos à descrição dos classificadores do chinês prende-se com o facto de nem sempre haver um entendimento claro sobre a natureza gramatical destes constituintes. A este respeito, Wu & Bodomo (2009:488) consideram que os classificadores correspondem a morfemas com conteúdo semântico, na medida em que “they often carry information beyond that carried by their associated noun”. Assim, um mesmo item nominal pode ser antecedido de diferentes classificadores aos quais se associam sentidos diferentes<sup>94</sup>. Por exemplo, o nome *mi* ‘arroz’ pode, em função do contexto, estar associado a classificadores diversos, conforme se verifica em *yi lì*<sub>classificador</sub> *mi* ‘um grão de arroz’; *yi wǎn*<sub>classificador</sub> *mi* ‘uma taça de arroz’; *yi tǒng*<sub>classificador</sub> *mi* ‘um barril de arroz’; *yi guō*<sub>classificador</sub> *mi* ‘um wok de arroz’ (cf. Wu & Bodomo 2009).

Em contrapartida, Li (2013) advoga que o facto de alguns classificadores numerais em chinês veicularem um significado lexical específico e de terem origem nominal, verbal e adjetival, não é suficiente para considerar estes itens como palavras lexicais, i.e., com conteúdo semântico. O autor argumenta, por exemplo, que, apesar de as preposições transmitirem determinadas informações semânticas, não deixam de ser consideradas, nas línguas em que se manifestam, como palavras gramaticais. Assim sendo, e apesar de haver, no chinês, itens lexicais que funcionam, simultaneamente, como nomes e classificadores, quando tais formas surgem como classificadores observa-se que o sentido original associado ao nome não se preserva<sup>95</sup>. Por conseguinte, o autor entende que em chinês é

---

<sup>93</sup> A seleção de um classificador é de tal modo idiossincrática que, para Li & Thompson, o aprendente de chinês deve memorizar que nomes admitem determinados classificadores (1981:112).

<sup>94</sup> Também Contini-Morava & Kilarski (2013:273) referem esta especificidade dos classificadores do mandarim.

<sup>95</sup> Li (2013) ilustra esta sua afirmação com o seguinte exemplo: “[w]hen the lexical item *bao* ‘bag’ is used as a noun, it refers to a certain bag, say, a school bag or a plastic bag. However, when it is used as a classifier (...) the main function of *bao* ‘bag’ is to express a containment relation.” (Li 2013:29).



possível uma mesma forma lexical ser utilizada em dois contextos sintáticos diferentes, com dois sentidos semânticos igualmente distintos.

Em suma, e nas palavras de XuPing Li (2013:30) os classificadores numerais em chinês:

“are functional (...) they do not have «descriptive content» (...) and they express the grammatical function of counting or measuring entities.”

#### **1.4.6. Síntese**

As Secções anteriores deste capítulo foram dedicadas à descrição e análise dos sistemas de classificação nominal, quer do português, quer das LM dos aprendentes, autores dos textos escritos que compõem a base empírica da presente investigação. Atendendo às questões colocadas no âmbito deste trabalho (cf. Introdução e Capítulo 3, Secção 3.1.1.), no qual se pretende averiguar o papel que os conhecimentos linguísticos prévios desempenham no processo de assimilação do sistema de atribuição de género nominal em PLNM, foram selecionadas também línguas que não exibem sistemas de classificação dos nomes baseados em género morfológicamente marcado, como é o caso do chinês. Para além disso, mesmo entre as línguas com esta categoria, são assinaláveis as distintas particularidades de atribuição dos valores de género, que confirmam a natureza idiossincrática da categoria gramatical em análise. Com efeito, e como se viu, na maioria dos sistemas de atribuição de género nominal aqui descritos, apesar de o género constituir uma propriedade gramatical inerente aos itens nominais, nem sempre determinado valor de género (masculino, feminino e neutro) é facilmente dedutível a partir das características semânticas e/ou formais dos nomes (Corbett 1991; Aikhenvald 2000). Assim sendo, é através do mecanismo da concordância sintática que a categoria de género se manifesta.



## **1.5. O género gramatical no contexto da aquisição linguística: LM vs. LNM**

A bibliografia especializada sobre a aquisição nativa de uma língua com um sistema de género nominal demonstra que, e independentemente da língua em análise, esta categoria, bem como a concordância dela decorrente, se encontra estabilizada logo nos primeiros anos de vida do falante, assinalando-se poucos desvios em tarefas de concordância nas produções infantis (Corbett 1991; Oliphant 1997; Clark 1998; Figueira 2001; Augusto & Corrêa 2005; Costa *et al.* 2015; Corrêa & Augusto 2017).

Em contrapartida, dados sobre a aquisição/aprendizagem de uma LNM revelam que, quanto mais tardia for a exposição do falante aprendente à LA, maiores dificuldades se registam na assimilação do sistema de atribuição de género aos nomes dessa língua, e os desvios são persistentes mesmo em produções de aprendentes com níveis de proficiência e de competência linguísticas mais avançados (veja-se, para espanhol como LNM, Franceschina 2003, 2005; para o italiano, Oliphant 1997; para o francês, Dewaele & Veronique 2000, 2001; para o português, Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Martins 2015; Pinto 2015).

Ainda que de forma breve, consideram-se, em seguida, os contributos da investigação sobre a aquisição de género na LM e na LNM, com especial enfoque para os estudos sobre a assimilação do género gramatical no português.

### **1.5.1. Aquisição de género gramatical por falantes nativos**

A investigação sobre a aquisição da categoria de género gramatical por falantes nativos de línguas de matriz indo-europeia revela que esta estrutura gramatical se encontra estabilizada logo nos primeiros anos de vida do falante (Corbett 1991; Clark 1998; Figueira 2001; Augusto & Corrêa 2005; Costa *et al.* 2015; Corrêa & Augusto 2017). Em função das características dos sistemas de atribuição de género da língua em aquisição, são reconhecidos marcos etários diferentes para o desenvolvimento do processo de identificação das propriedades relativas à associação dos valores de género aos nomes nas produções infantis<sup>96</sup> (Corbett 1991:82). No caso específico do português, estudos vários

---

<sup>96</sup> Segundo Aikhenvald (2000), a existência de marcos etários distintos para a aquisição dos valores de género por falantes nativos está intimamente associada às propriedades dos sistemas de atribuição de género nominal dos idiomas. Por conseguinte, a autora (2000:415) observa que “[m]orphological

têm vindo a demonstrar que a identificação da informação relativa ao género se manifesta por volta dos dois anos de idade (Augusto & Corrêa 2005) e que aos três anos se estabiliza o conhecimento relativo às propriedades do sistema de concordância de género (Costa *et al.* 2015; Corrêa & Augusto 2017).

Uma parte substancial da investigação procura identificar, a partir de dados empíricos, que fatores contribuem para a capacidade precoce da criança de atribuição dos valores de género aos nomes. A este respeito, Clark (1998:381) observa que “factors related to the forms of gender marking appear to be an important determinant of how early and how easily children acquire gender marking.” Ou seja, são sobretudo os fatores de natureza linguística (indícios de natureza morfofonológica e sintática) que contribuem para o processamento e aquisição da informação de género (Oliphant 1997:2; Aikhenvald 2000:414; Dewaele & Veronique 2001:277).

No português, estudos vários contribuem para uma maior compreensão não só do desenvolvimento da aquisição nativa de género, mas também das estratégias mobilizadas pelas crianças na atribuição dos valores de género aos nomes (Figueira 2001; Corrêa *et al.* 2004; Augusto & Corrêa 2005; Costa *et al.* 2015; Choupina *et al.* 2016; Corrêa & Augusto 2017). Nestes trabalhos, é geralmente aceite a ideia de que o reconhecimento das relações de concordância sintática entre os elementos que compõem o grupo nominal detém um papel essencial para a aquisição nativa de género, sendo que o processamento sintático contribui igualmente para a identificação da categoria de género gramatical bem como das propriedades associadas à atribuição dos valores de género aos nomes.

Em estudos sobre a aquisição do português do Brasil (Corrêa *et al.* 2004; Augusto & Corrêa 2005; Corrêa & Augusto 2017) verifica-se que, por volta dos dois anos de idade, a criança demonstra um conhecimento implícito relativamente às propriedades das unidades do léxico que contribuem para a combinação dos itens nas unidades sintáticas. No léxico, os itens distribuem-se por duas categorias essenciais: lexicais e funcionais/gramaticais. As categorias lexicais configuram uma classe aberta de palavras com conteúdo semântico e concetual e dela fazem parte os nomes, os verbos, os adjetivos e os advérbios. As categorias funcionais compreendem uma classe fechada constituída por

---

complexity may slow down the process of gender acquisition. Children have more difficulties in mastering gender systems of languages with three genders where gender markers interact with other categories, such as number and case (...). In contrast, children learning languages with clearly differentiated and regular marking of just two genders (...) master these systems more easily and more quickly”.

especificadores, verbos auxiliares, entre outros constituintes de natureza estritamente gramatical<sup>97</sup> (Corrêa & Augusto 2017). Postula-se que o reconhecimento das categorias funcionais, anterior à sua mobilização para atividades de produção, serve como um instrumento ao serviço da aquisição do léxico nos primeiros anos de vida da criança (Corrêa & Augusto 2017). Dados empíricos analisados por Corrêa *et al.* (2004) são reveladores de que logo “no segundo ano de vida a criança é sensível à forma fônica de determinante” (2004:135) sendo que, já no final desse período, apresenta um grau elevado de sensibilidade não só relativamente à forma e posição do determinante, mas também à presença de marcas incongruentes de género entre determinante e nome.

No estudo conduzido por Choupina *et al.* (2016) sustenta-se igualmente a importância da concordância para a aquisição dos valores de género. Neste trabalho, verificou-se que crianças, falantes nativas do português europeu, a frequentar o 1º e 2º ciclos escolares, são mais proficientes em tarefas de produção de concordância do nome com especificadores e adjetivos do que na produção de cadeias de correferência. Neste estudo, registam-se maiores dificuldades nas tarefas de identificação explícita dos valores de género de nomes como *tribo* ou *panda* justificáveis, para os autores, pelo facto de no ensino explícito de género se incluir, muitas vezes, a ‘pseudoregra’ de associação entre valores de género masculino e feminino e índices temáticos *-o* e *-a*, respetivamente (Choupina *et al.* 2016:226).

A partir da análise de produções espontâneas (e com um certo efeito humorístico) proferidas por crianças entre os 2 e 4 anos de idade, Figueira (2001) considera que, no processamento da informação relativa ao género gramatical, as crianças se baseiam, primeiramente, na identificação do género natural (macho vs. fêmea) das entidades animadas para distinguir os valores de género gramatical (masculino vs. feminino). Como exemplos ilustrativos, refira-se a ocorrência de casos de sobregeneralização da regra ‘se item termina em *-o* é de género masculino’ (*bom dia*) que são logo associados, pela criança, à informação de sexo (*Bom dia é para homem. Bom dia é para mulher*) (Figueira 2001:315).

---

<sup>97</sup> Na proposta de Raposo (2013), os advérbios não figuram como elementos pertencentes às classes lexicais, uma vez que estes não têm referente extralinguístico. Assim, para este autor (2013:333) “[a]s classes lexicais são aquelas cujos membros possuem um significado descritivo, independente de qualquer morfema flexional a que possam estar ligados, significado esse que denota entidades ou situações exteriores à linguagem (coisas, pessoas, animais, qualidades, lugares, ideias, ações, etc.). Os nomes, os adjetivos e os verbos constituem as classes lexicais. De facto, a generalidade destes itens pode ser definida em função de um significado descritivo extralinguístico, que remete para o mundo *fora da linguagem*, para as entidades, estados, ações e processos que nele existem, e para as propriedades destes. Em contrapartida, os membros das classes gramaticais têm essencialmente uma função estruturadora, quer ao nível da frase quer ao nível do discurso.”

Com estes dados, a investigadora considera que, a partir dos 3 anos de idade, já é possível “contemplar na criança uma atitude reflexiva em relação à linguagem” (Figueira 2001:319), ou seja, defende que a criança mobiliza, desde muito cedo, conhecimentos de natureza metalinguística relativamente às marcas de atribuição de género nominal.

Augusto & Corrêa (2005) chamam ainda a atenção para o facto de, logo aos dois anos de idade, as crianças revelarem sensibilidade para a distinção de nomes de “género intrínseco” e de nomes de “género opcional” (2005:228). De acordo com as autoras, nos nomes de género intrínseco, o género encontra-se especificado no radical, incluindo-se, nestes casos, os itens com referentes não animados (*a mesa, o papel, a ponte*) e animados humanos e não humanos (*a vítima, o elefante*). Nos nomes de género opcional, este não se encontra especificado no radical e verifica-se apenas na sintaxe com a consequente atribuição do valor de género pelos mecanismos de concordância sintática (*o/a estudante, a menina, o menino*) (Costa *et al.* 2015:332). Assim, para as investigadoras, o reconhecimento da opcionalidade do traço formal do género contribui para que a criança reconheça o valor genérico associado ao masculino e que, por ser usado na nomeação dos elementos de uma classe, não possui, exclusivamente, uma relação com o sexo das entidades designadas. Por conseguinte, Augusto & Corrêa (2005:229) concluem que:

“aos dois anos de idade (...) a criança já está inserida no sistema gramatical do português e parece identificar a marcação morfofonológica relativa a género opcional como informação relevante não só para o processamento sintático, mas também para a constituição de categorias semânticas com implicações para a compreensão e para o estabelecimento da referência genérica.”

Em suma, os dados da aquisição no português como LM são reveladores de que a assimilação dos valores de género nominal ocorre em fases precoces do desenvolvimento do falante. Destaca-se ainda o papel das estratégias de natureza formal<sup>98</sup>, nomeadamente o reconhecimento das relações sintáticas entre as unidades do léxico, i.e., entre o nome, determinantes e adjetivos, para o processo de associação dos valores de género por parte

---

<sup>98</sup> Também Oliphant (1997:3), ao analisar um conjunto de estudos sobre a aquisição linguística nativa do género gramatical, conclui que “[t]he L1 research makes clear that young children learning their first language do not use a semantic strategy in acquiring grammatical gender, while the relative weight they give to morphophonological and syntactic cues varies from study to study, seemingly a function of age level. It is apparent that children develop an increasing ability to process and integrate competing cues which allows them to cope with the complexities of gender agreement.”

das crianças, já que “a relação entre o gênero gramatical de uma palavra e a forma fonológica não é direta” (Augusto & Corrêa 2005:215).

### 1.5.2. Aquisição do gênero gramatical por falantes não nativos

A investigação sobre o processo de aquisição/aprendizagem de uma LNM, sobretudo em fases tardias do desenvolvimento do falante aprendente, tem demonstrado que, mesmo nos níveis de proficiência linguística avançados, se registam desvios de atribuição de gênero nominal<sup>99</sup> que nunca chegam a ser erradicados por completo das produções dos aprendentes.

Dados empíricos relativos ao processo de aquisição/aprendizagem da atribuição de gênero nominal por aprendentes tardios de português como LNM, seja como língua segunda ou língua estrangeira (Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Mariotto 2014) seja como língua terceira ou língua adicional (Lacsán 2015; Pinto 2015) demonstram que, apesar de se registar uma melhoria global no desempenho em função do nível de proficiência linguística dos aprendentes, os desvios de atribuição dos valores de gênero nunca chegam a ser totalmente eliminados. Para além disso, trabalhos que analisam dados relativos ao processo de aquisição da concordância nominal por aprendentes de PLNM (Leiria 2006; Godinho 2010; Martins 2015) e de falantes de variedades não nativas do português (Gonçalves 1997, 2015; Lucchesi 2009; Inverno 2009) identificam o gênero como uma área crítica.

A maioria dos estudos produzidos baseia-se, essencialmente, na análise de desvios de atribuição de valores de gênero aos nomes e de concordância nominal em gênero registados em produções escritas por aprendentes tardios de PLNM em diferentes estágios de aprendizagem. Nestes trabalhos, são colocadas hipóteses explicativas para a persistência dos comportamentos desviantes ao longo do seu desenvolvimento interlinguístico, tendo-se em conta não só os (possíveis) constrangimentos associados ao perfil do aprendente – nomeadamente a idade e período de exposição do aprendente à LA, o nível de proficiência e de competência linguísticas do aprendente (nível QECRL), a LM ou outras línguas previamente conhecidas –, mas também as condicionantes associadas às características do sistema de atribuição de gênero nominal do português. Também os dados analisados sobre

---

<sup>99</sup> Veja-se, por ex., para o espanhol como LNM, os estudos de Alarcón 2004, Franceschina 2005, Isabelli-Garcia 2010; para o italiano como LNM, Oliphant 1997; para o francês, Dewaele & Veronique 2000, 2001.

o estabelecimento da concordância de género por falantes de variedades não nativas do português (Gonçalves 1997:60; Cavele 1999; Inverno 2009:165) identificam padrões de elevada variação relativos a esta estrutura, observando-se uma tendência constante para a redução do paradigma de marcação morfossintática desta categoria gramatical, sobretudo por falantes com baixos índices de escolaridade (Figueiredo 2009:51; Gonçalves 2015:12).

Ferreira (2011) analisou desvios de atribuição de género nominal detetados em textos escritos por aprendentes tardios de PLNM com experiência instrucional. Na análise dos dados, foi tido em conta o nível de proficiência linguística da turma frequentada pelos sujeitos informantes, assumindo-se, a partir desse indicador, que é possível aferir diferentes momentos do desenvolvimento das interlínguas dos aprendentes. A partir dos resultados obtidos, verificou-se que, embora se registre uma diminuição considerável do número de desvios nos níveis de proficiência mais avançados, os desvios nunca chegam a desaparecer totalmente. Logo, sugere-se que, ao longo do desenvolvimento interlinguístico, os aprendentes vão reconhecendo alguns dos indícios que lhes permitem identificar os valores de género dos nomes. Ainda assim, o número de desvios registado sugere que, no processo de aquisição/aprendizagem, a assimilação do valor de género de um nome é progressiva e relativamente lenta, processando-se item a item (Ferreira 2011:95).

Para além de Ferreira (2011), outros trabalhos revelam que os aprendentes de níveis avançados registam um melhor desempenho na atribuição dos valores de género nominal em relação aos falantes aprendentes dos níveis elementares (Godinho 2010; Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Lacsán 2015; Martins 2015; Pinto 2015).

No que concerne às estratégias adotadas pelos aprendentes ao longo do processo de aquisição/aprendizagem da atribuição de género nominal em PLNM, os dados apurados em diferentes estudos demonstram que a maioria dos desvios analisados resulta da adoção de uma estratégia de “marcação *default* da forma masculina” (Mariotto & Lourenço-Gomes 2013:1282), i.e., no momento de atribuir um valor de género aos nomes, observa-se uma preferência pela seleção do masculino, a forma não-marcada. O recurso a esta estratégia de simplificação do material linguístico verifica-se não só a partir dos dados de desempenho de aprendentes tardios de PLNM (Mariotto & Lourenço-Gomes 2013; Martins 2015) e de falantes de variedades não nativas do português (Figueiredo 2009), mas também em dados recolhidos de outras línguas no contexto de aquisição/aprendizagem tardia (Bruhn de Garavito & White 2002; Alarcón 2004; Isabelli-Garcia 2010). Em contrapartida, Pinto (2015), baseado em dados recolhidos de produções escritas por aprendentes marroquinos, falantes bilíngues, a frequentar cursos de português numa universidade em Marrocos,



conclui que os desvios resultam de um uso generalizado ora da forma não-marcada, o masculino, ora da forma marcada, o feminino, não havendo, portanto, e na opinião do autor, uma clara preferência de um valor em detrimento de outro. Assim, a partir da análise dos dados, o autor conclui que estes alunos “adquirem o género isoladamente para cada entrada lexical” (Pinto 2015:109).

Relativamente às características semânticas e formais dos nomes afetados, uma parte substancial dos estudos revela que nomes cujo valor de género gramatical corresponde ao género natural da entidade designada apresentam índices de desvios de atribuição de género muito baixos em todos os níveis de proficiência<sup>100</sup> (Ferreira 2011; Martins 2015; Pinto 2015).

Os nomes cujo valor de género não depende de critérios de natureza semântica nem de critérios de natureza formal apresentam maiores índices de desvio, independentemente do perfil do aprendente e do nível de proficiência linguística em português (Ferreira 2011; Martins 2015; Pinto 2015). Tendo em conta os dados analisados por Ferreira (2011) e por Martins (2015), extraídos da mesma base de dados, o *Corpus* PEAPL2, verifica-se que os nomes atemáticos registam um maior número de desvios de atribuição de género nominal. Porém, curiosamente, também os nomes cujo valor de género, masculino e feminino, seria corretamente dedutível a partir da aplicação da ‘pseudorregra’: “se o nome termina em *-o* é de género masculino e se o nome termina em *-a* é de género feminino”, registam igualmente índices de desvio consideravelmente elevados (Ferreira 2011:63; Martins 2015:41). Como vimos, em português, cerca de 60% das ocorrências nominais evidenciam uma relação, ainda que parcial, entre os índices temáticos *-o* e *-a* e o valor de género do nome (cf. Secção 1.3.3.), sendo que, nos restantes casos, a relação entre valores de género e marcadores morfológicos não se verifica. Atendendo a este facto, poder-se-á assumir que a fraca robustez do *input* linguístico, no que à atribuição dos valores de género diz respeito, contribui para uma certa ‘desconfiança’ dos aprendentes relativamente ao poder preditivo dos indícios morfológicos.

---

<sup>100</sup> Dados empíricos recolhidos de variedades não nativas do português revelam que os falantes seguem os padrões de concordância de género gramatical quando o núcleo do sintagma nominal é um item cujo valor de género gramatical corresponde ao sexo do respetivo referente. Lucchesi (2009:317) observa que em crioulos de base lexical portuguesa de África: “a estrutura linguística que tende a oferecer maior resistência à erosão gramatical ocorrida durante o processo de transmissão linguística irregular é a marcação de género nos núcleos nominais que designam os seres animados, nos casos em que a indicação referencial do sexo dos indivíduos é relevante.”

Por sua vez, Pinto (2015) regista um maior número de desvios em itens com índice temático (-o, -a, -e) por oposição aos que não possuem este constituinte morfológico. Para o investigador, a ausência de índice temático conduz os aprendentes a focalizarem a sua atenção no valor de género do nome e, para além disso, a tentarem “adquirir o género da palavra independentemente da sua forma” (Pinto 2015:100), mas tal assunção não se aplica aos nomes com índice temático -o, -a e -e, uma vez que os dados de Pinto (2015) revelaram grande percentagem de desvios nos mesmos. Nos atemáticos terminados em -ão o autor verificou que os aprendentes tendem a atribuir o género masculino, podendo este comportamento justificar-se pelo facto de os alunos interpretarem -o final como um índice temático (Pinto 2015:100). Tais dados parecem indiciar que estes aprendentes assumem que a informação de género se concentra nas vogais finais dos nomes, ao contrário do que se verifica nos dados de Ferreira (2011) e de Martins (2015). Tal hipótese, avançada pelo investigador, é suportada pelo facto de neste grupo de informantes a aprendizagem se basear, essencialmente, na leitura e em atividades de produção escrita, havendo pouco acesso a *input* oral da LA. Por conseguinte, é a forma escrita das palavras que parece exercer uma maior influência na sua aquisição (Pinto 2015:100).

Quanto aos constituintes sintáticos nos quais se explicitam as marcas de atribuição desviante do valor de género, a maior parte dos estudos revela que é no especificador, nomeadamente nos determinantes, que se evidencia um maior número de ocorrências desviantes de atribuição de género nominal, por oposição aos modificadores (Godinho 2010; Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Martins 2015; Pinto 2015). Ferreira (2011:64-65) postula que os determinantes “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”. No confronto das produções desviantes de dois grupos de aprendentes, falantes nativos de espanhol e de inglês, Mariotto & Lourenço-Gomes (2013:1283) verificaram igualmente uma maior frequência de desvios nos especificadores. Para além disso, os dados revelaram que os aprendentes, falantes nativos de inglês, apresentam uma maior tendência para a alteração da forma morfológica do nome e conseqüente alteração do seu valor de género por oposição ao que se verifica no grupo de falantes nativos de espanhol. Da análise global do desempenho dos aprendentes e dos desvios produzidos por cada segmento da amostra, as autoras (2013:1283) concluem que:

“adultos que aprendem uma língua com estrutura morfológica semelhante a de sua língua materna tendem a ser mais consistentes na aplicação da concordância de gênero na língua que está sendo aprendida do que adultos aprendendo uma língua com estrutura morfológica muito distinta da sua língua materna.”

Em contrapartida, Lacsán (2015:70-71), ao analisar dados relativos à assimilação da categoria de gênero por aprendentes tardios, falantes nativos de húngaro, observa que há uma maior dificuldade na atribuição dos valores de gênero nos adjetivos, sobretudo quando estes se encontram mais afastados do núcleo nominal. Lacsán considera que o facto de os aprendentes serem bem-sucedidos na marcação de gênero no determinante demonstra que estes são capazes de reconhecer o valor de gênero do nome, por um lado, e, por outro, o pior desempenho nas tarefas de concordância do nome com o adjetivo é revelador da incapacidade do aprendente em estabelecer relações de concordância nominal. Por conseguinte, Lacsán (2015:120) conclui que:

“L2 Portuguese speakers seem to have been more successful at acquiring gender as the lexical property of nouns, but did not seem to succeed so well at the acquisition of gender agreement which is a syntactic process.”

No estudo de Lacsán, avaliou-se também o peso dos conhecimentos linguísticos prévios para a aquisição/aprendizagem da atribuição dos valores de gênero aos nomes em PLNM e, do confronto do desempenho de falantes nativos de línguas românicas e falantes nativos de línguas não românicas, a autora constata que é sobretudo nos níveis iniciais que se regista um melhor desempenho global do primeiro grupo de aprendentes, não havendo, em contrapartida, diferenças substanciais entre os dois grupos nos níveis de proficiência linguística mais avançados<sup>101</sup>.

Num estudo de natureza psicolinguística, Mariotto (2014) procurou averiguar, tendo em conta dados extraídos de tarefas de leitura, se aprendentes falantes nativos de inglês, no nível B1 de proficiência linguística em português, seriam sensíveis à violação da concordância de gênero entre nomes [+animados] e [-animados] e adjetivos com função predicativa. Neste trabalho, foram criadas 24 frases manipuladas em função do traço

---

<sup>101</sup> A autora observa ainda que “level B2 Non-Romance L2 Subgroup participants had similar or even higher scores than advanced level (level C1) Romance L2 speakers.” (Lacsán 2015:122).

[±animado] dos itens nominais e da congruência<sup>102</sup> na concordância de género entre nome e adjetivo. A hipótese avançada neste estudo foi a de que falantes nativos de inglês, por não terem na sua LM um sistema de atribuição de género nominal, seriam menos sensíveis às variações de concordância entre nome e adjetivo, sendo essa dificuldade acrescida em nomes com referentes não animados (Mariotto 2014:54). Face aos resultados, constatou-se que, efetivamente, os falantes não nativos apresentam tempos de leitura maiores em relação ao grupo de controlo, composto por falantes nativos, sobretudo quando confrontados com frases malformadas. Além disso, os dados extraídos da tarefa de aceitabilidade das frases demonstram que o traço semântico [+animado] do referente não exerce qualquer influência no comportamento dos aprendentes, ou seja, não há um melhor desempenho nos tempos de leitura de frases com itens com referentes sexuados. Mariotto (2014:58) entende que estes resultados podem estar condicionados pelo nível de proficiência linguística dos informantes em análise, já que, “[a] pesar de os participantes apresentarem um nível limiar de proficiência, (...) talvez ainda estejam numa fase de reconhecimento deste traço na língua-alvo, e (...) o processamento do género ainda não seja uma competência alcançada”.

Na investigação de Leiria (2006) relativa ao processo de aquisição do léxico do português, baseada em dados recolhidos de produções escritas por aprendentes tardios de PLNM, a autora observa que o facto de a LM do informante possuir um sistema de atribuição de género nominal não auxilia o informante na associação dos valores de género aos nomes do português. Como assinala a investigadora, entre os falantes nativos de chinês regista-se um reduzido número de desvios de atribuição de género relativamente aos assinalados nos falantes nativos de línguas com sistemas de classificação nominal baseados em género, como o alemão e o sueco (Leiria 2006:241). Assim, poder-se-á supor que, para o desenvolvimento das interlínguas, a presença do género gramatical na LM não constitui, *por si só*, um requisito favorável para a assimilação dos valores de género dos nomes em português. Nota ainda a autora que, do conjunto total de desvios detetados nos diferentes grupos de aprendentes, e independentemente da LM do falante aprendente, cerca de 24% dos desvios incidem sobre os mesmos nomes. As ocorrências desviantes em nomes como *pessoa* e *gente* levam Leiria (2006:242) a assumir que estes itens são representados no léxico mental dos aprendentes como se fossem nomes comuns de dois. Por conseguinte, segundo a autora (2006:242-243):

---

<sup>102</sup> A congruência vs. incongruência resultou da alteração do valor de género do adjetivo (Mariotto 2014:51).

“sendo o masculino (...) o valor não marcado, é ele que é selecionado sempre que se pretende obter uma referência genérica. Daí *os pessoas, pessoas portuguesas, pessoas primitivos* e o pronome pessoal masculino *eles*.”

Do mesmo modo, os dados apurados por Martins (2015) revelam que o facto de a LM do falante aprendente exibir um sistema de género não auxilia o aprendente no processo de assimilação dos valores de género nominal em português. Da análise dos resultados, a autora observa que os falantes nativos de chinês registam um melhor desempenho global na atribuição dos valores de género nominal face ao registado em aprendentes falantes nativos de espanhol e de alemão. Infere-se, assim, um efeito de transferência negativa “não tanto direta de valores de género da LM para a LA em vocábulos semanticamente equivalentes (...), mas mais propriamente sob a forma de uma atitude de insegurança por parte do aprendente nestas circunstâncias” (Martins 2015:44).

### 1.5.3. Síntese

A partir da análise dos dados relativos à aquisição do género por falantes nativos de português e por aprendentes tardios de português como língua não materna, é possível traçar as principais particularidades associadas a um e a outro processo (cf. Quadro 1.9).

	Aquisição do sistema de atribuição de género nominal do português	
	Falantes nativos	Falantes não nativos
<b>Cronologia dos processos</b>	Aos 3 anos de idade a criança demonstra um conhecimento implícito das propriedades do sistema de atribuição de género nominal.	Aprendentes tardios apresentam dificuldades em associar os valores de género aos nomes em português.
<b>Ocorrências de desvios</b>	Registam-se poucos desvios de concordância nominal nas produções dos sujeitos infantis e maiores dificuldades em tarefas de identificação explícita e em tarefas de correferência.	Dificuldade na erradicação de desvios de atribuição de género e de concordância de género, mesmo em de níveis de proficiência mais avançados.
<b>Estratégias mobilizadas</b>	Mobilizam-se estratégias sobretudo formais: o reconhecimento dos padrões morfológicos associados às relações sintáticas entre as unidades do léxico é fundamental para a aquisição dos valores de género.	Aprendentes tardios revelam dificuldades em reconhecer e aplicar, com sucesso, os padrões de concordância sintática e de atribuição dos valores de género nominal.

**Quadro 1.9** – Síntese das particularidades associadas à aquisição do sistema de atribuição de género nominal do português por falantes nativos e por falantes não nativos

Descritas as características associadas à aquisição de género por falantes nativos e não nativos do português, o próximo Capítulo é dedicado à discussão das questões que se prendem com a aquisição/aprendizagem tardia de uma LNM. Mais especificamente, e uma vez que neste trabalho se procura testar, através da análise de uma base empírica robusta descrita no Capítulo 3, em que medida a LM produz, ou não, efeitos no comportamento linguístico dos falantes aprendentes de PLN, ter-se-ão em conta vários contributos teóricos para a descrição do papel dos conhecimentos linguísticos prévios para a assimilação dos diferentes planos estruturados de uma LNM.

# Capítulo 2

## Para uma avaliação do papel dos conhecimentos linguísticos prévios na aquisição/aprendizagem de uma LNM

---

### 2.1. Introdução

No presente capítulo, propomo-nos analisar o papel dos conhecimentos linguísticos prévios para a aquisição/aprendizagem do género gramatical numa LNM tal como estes têm sido enquadrados pelas principais correntes teóricas que atuam no âmbito da investigação em aquisição de línguas não maternas.

Partindo do pressuposto de que o processo de aquisição por falantes nativos é distinto do processo de aquisição/aprendizagem tardia por falantes não nativos, nomeadamente no que à assimilação dos sistemas de género nominal diz respeito – e como se viu na Secção 1.5.2.–, o principal objetivo deste capítulo é o de compreender, à luz de diferentes contributos teóricos, o eventual papel dos conhecimentos linguísticos prévios para justificar as maiores dificuldades que parece exhibir o aprendente tardio neste domínio. Por conseguinte, consideram-se as questões teóricas relacionadas com o lugar dos conhecimentos linguísticos prévios, sobretudo os provenientes da LM, para a construção das interlínguas dos aprendentes e de como estes interagem, ao longo dos vários estágios da aquisição/aprendizagem dos diferentes planos estruturados da língua à qual os falantes estão expostos.

Refira-se, agora, a estrutura deste capítulo. A primeira parte é dedicada à reflexão em torno dos principais aspetos relacionados com a aquisição/aprendizagem tardia de uma LNM (Secção 2.2.), equacionando-se, a este respeito, os condicionamentos associados à hipótese do ‘período crítico’ (ou ‘sensível’) para a aquisição linguística (Secção 2.2.1.) e as particularidades relativas à caracterização do conhecimento gramatical dos aprendentes tardios, i.e., a interlíngua (Secção 2.2.2.). A partir do conceito de interlíngua, consideram-se

os contributos teóricos para a configuração do conhecimento gramatical por parte dos aprendentes tardios, nomeadamente no que diz respeito ao grau de acessibilidade a diferentes componentes que configuram o conhecimento linguístico inato, a Gramática Universal (GU) (Secção 2.2.2.1.), bem como as questões associadas ao processamento da informação linguística, atendendo à relevância que, a este nível, tem sido dada às características do *input* linguístico (Secção 2.2.2.2.) e ao papel diferenciado dos subsistemas cognitivos envolvidos para o processamento da informação linguística (Secção 2.2.2.3.).

A ponderação destes aspetos será útil para, posteriormente, na Secção 2.3., se poderem enquadrar devidamente as hipóteses teóricas relativas ao papel dos conhecimentos linguísticos prévios dos aprendentes para a construção dos sistemas interlinguísticos e, mais especificamente, para o desenvolvimento da aquisição/aprendizagem tardia do sistema de atribuição de género em PLNM. Assim, e após a caracterização do conceito de ‘transferência linguística’ (Secção 2.3.1.) e da referência às principais manifestações a ele associadas (Secção 2.3.2.), analisam-se as hipóteses evocadas na bibliografia especializada para compreender o grau de transferibilidade do género gramatical (Secção 2.3.3.).



## 2.2. Aquisição da LM vs. aquisição/aprendizagem de uma LNM

Como área multidisciplinar, a investigação produzida no campo da Aquisição de Línguas Não Maternas<sup>103</sup> (ALNM) reúne contributos de diferentes modelos teóricos, tanto na área da Linguística como da Psicolinguística que, em conjunto, têm procurado identificar fatores de diversa natureza (linguística, cognitiva, cultural e social) para uma maior compreensão das especificidades relativas do processo de aquisição de uma LNM, sobretudo quando este ocorre em fases tardias do desenvolvimento do aprendente (Ellis 2000; Gass & Selinker 2008).

É geralmente aceite a ideia de que aquisição da língua materna (LM) se distingue, em alguns aspetos essenciais, da aquisição/aprendizagem tardia de uma LNM (Bley-Vroman 1990; Ritchie & Bhatia 1996; Ellis 2000; Gass & Selinker 2008) na medida em que:

- (i) quando inicia o processo de assimilação das estruturas de uma LNM, o aprendente já possui, pelo menos, um sistema linguístico prévio, i.e., uma LM; e
- (ii) muito dificilmente o aprendente tardio atinge, satisfatoriamente, níveis de competência e de proficiência linguísticas consideradas nativas em todos os planos estruturados da língua-alvo (LA).

No que ao género gramatical diz respeito, vimos que esta estrutura gramatical constitui uma área crítica no contexto da aquisição/aprendizagem tardia<sup>104</sup> do PLNM

---

<sup>103</sup> Neste trabalho, preferiu-se utilizar a expressão “língua não materna” (LNM) para designar qualquer sistema linguístico não nativo a cujo *input* linguístico o aprendente é exposto depois de adquirida a língua materna (LM), já em fases mais tardias do seu desenvolvimento linguístico (Leiria 2006:7; Flores 2013:35; Madeira 2017:306). Trata-se, portanto, de um termo tomado numa aceção mais lata e cuja preferência se justifica pelas dificuldades associadas ao uso de outras expressões utilizadas no contexto da investigação na área, tais como “língua segunda” (LS) e “língua estrangeira” (LE) que, pese embora o facto de designarem o sistema linguístico não nativo cuja exposição a *input* ocorre tardiamente, apresentam divergências no que concerne à realidade sociolinguística a que se circunscrevem (Leiria 2004; Flores 2013). Em boa verdade, a LS corresponde a uma LNM que numa determinada comunidade possui um estatuto sociopolítico definido (Leiria 2004:1) e pode ser, então, uma das línguas oficiais do país, que é ensinada nas escolas e possui um papel crucial dentro das fronteiras territoriais onde é utilizada (Leiria 2004). Em contrapartida, a LE corresponde a um sistema linguístico que é aprendido, sobretudo, em contexto formal, e é tipicamente utilizado numa comunidade onde essa língua não possui um estatuto sociopolítico definido (Leiria 2004:4).

<sup>104</sup> Optou-se, no presente trabalho, pela expressão “aquisição/aprendizagem de uma LNM” para designar o processo de assimilação das estruturas linguísticas em fases tardias do desenvolvimento do falante aprendente. Esta opção justifica-se pelo facto de, no contexto da investigação em ALNM, não

(Godinho 2010; Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Martins 2015; Pinto 2015), já que os desvios de atribuição de género ocorrem mesmo em níveis de proficiência e de competência linguísticas avançados, não o sendo, porém, na aquisição linguística por falantes nativos (Corrêa *et al.* 2004; Augusto e Corrêa 2005; Costa *et al.* 2015; Choupina *et al.* 2016; Corrêa & Augusto 2017).

Dados empíricos disponíveis sobre a assimilação tardia do género gramatical no português<sup>105</sup> permitem aferir algumas das especificidades que envolvem a assimilação tardia da associação dos valores género aos nomes, bem como a concordância sintática decorrente dela por aprendentes com diferentes perfis sociolinguísticos (Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Lacsán 2015; Martins 2015; Pinto 2015).

Ao contrário do que se observa na aquisição/aprendizagem tardia do sistema de atribuição de género nominal e da concordância nominal em género, estudos sobre o processo de aquisição nativa revelam, e como se viu, que a assimilação desta estrutura gramatical não se afigura particularmente problemática, começando a estabilizar-se logo nos primeiros anos de vida do falante (Corbett 1991:82; Clark 1998; Costa *et al.* 2015; Corrêa & Augusto 2017).

O confronto dos dados da aquisição do sistema de atribuição de género nominal do português por falantes nativos e por aprendentes tardios de PLNM conduz, necessariamente, à reflexão em torno dos constrangimentos associados à assimilação tardia desta estrutura gramatical e, por este motivo, consideram-se em seguida as questões teóricas relacionadas com a possibilidade de haver uma idade ‘crítica’ ou ‘sensível’ para a aquisição/aprendizagem dos diferentes planos estruturados de uma LNM (Cf. Secção 2.2.1.). Para além disso, o facto de o processo aquisitivo da categoria de género ser

---

existir consenso sobre os fatores que contribuem para a delimitação concetual dos termos “aquisição”, por um lado, e “aprendizagem” de uma LNM, por outro (Ellis 2000; Gass & Selinker 2008). Em linhas gerais, o termo “aquisição” é tipicamente utilizado para designar o processamento das estruturas de uma língua em fases precoces do desenvolvimento e que decorre, prototipicamente, em contextos naturais, i.e., em situações em que os falantes estão expostos ao *input* linguístico e, naturalmente, sem qualquer atenção consciente, adquirem as estruturas linguísticas que lhe permitem comunicar nessa língua. Em contrapartida, o conceito de “aprendizagem” compreende o processo de assimilação tardia das estruturas linguísticas, circunscrevendo-se, prototipicamente, a um contexto formal. Há, para além disso, uma clara consciência, por parte do aprendente, das diferentes fases do processo de aprendizagem do novo idioma. O facto de não haver consenso relativamente aos critérios que permitem distinguir claramente o domínio da “aquisição” e da “aprendizagem” de uma LNM, tem levado muitos investigadores à utilização dos termos de forma indiferenciada (Ellis 2000:6; Gass & Selinker 2008). Sobre este assunto, veja-se ainda a proposta de Krashen (1981).

<sup>105</sup> Cf. Capítulo 1, Secção 1.5.2.

manifestamente divergente em falantes nativos e em aprendentes tardios leva a que se analisem as particularidades associadas à descrição dos sistemas interlinguísticos característicos dos falantes aprendentes tardios, sendo certo que, no presente trabalho, se dará especial atenção ao papel dos conhecimentos linguísticos prévios para a assimilação do sistema de atribuição de género nominal em PLNM, com o intuito de compreender, à luz de diferentes contributos teóricos, se esta categoria gramatical é potencialmente transferível de uma LM para a LA de aprendizagem e, no caso de não ser transferível, que outros fatores atuam ao longo das diferentes fases de assimilação dos valores de género nominal por aprendentes tardios de PLNM com diferentes perfis linguísticos.

### 2.2.1. Hipótese do período ‘crítico’ (ou ‘sensível’) para a aquisição linguística

A noção tradicional de ‘período crítico’ para a aquisição linguística remonta, essencialmente, aos trabalhos de investigação sobre a aquisição monolíngue de Lenneberg (1967 [1985]). Este investigador, apoiado por argumentos de natureza neurofisiológica, entende que existe um período de tempo no desenvolvimento da criança no qual a aquisição de um idioma deve ocorrer, já que, em fases tardias do seu desenvolvimento ontogénico, se observam constrangimentos biológicos que impedem, necessariamente, a assimilação natural das estruturas linguísticas (Lenneberg 1967 [1985]; Birdsong 1999:1).

Lenneberg (1967 [1985]) advoga que a aquisição linguística ocorre até à puberdade, sendo este limite cronológico coincidente com o final do processo de lateralização das funções linguísticas no cérebro<sup>106</sup> (Lenneberg 1967 [1985:187]). Deste modo, o investigador considera que, se durante o desenvolvimento da criança não houver suficiente exposição a *input* linguístico, o processo de lateralização das funções linguísticas não se consolida e fica, após a puberdade<sup>107</sup>, inevitavelmente comprometido.

---

<sup>106</sup> A lateralização define-se como um processo de natureza fisiológica, considerado um dos mais relevantes para o desenvolvimento e amadurecimento do cérebro e do qual está dependente a especialização de diferentes funções cognitivas, entre as quais as que dizem respeito à capacidade para a linguagem (Ritchie & Bhatia 1996:967). A lateralização das funções linguísticas no cérebro consiste, precisamente, na definição e no amadurecimento de duas áreas corticais no hemisfério esquerdo e que são tipicamente referenciadas para o processamento linguístico: as áreas de Broca (responsável pela produção) e de Wernicke (responsável pela compreensão). Com o culminar do processo de lateralização, o cérebro deixa de revelar o grau de plasticidade que antes o caracterizava. A plasticidade cerebral corresponde a “the ability of neurons to make new connections and varied connections depending on stimulus.” (Eubank & Gregg 1999:69).

<sup>107</sup> Lenneberg (1967 [1985]) considera que é na puberdade, sensivelmente aos treze anos de idade, que o processo de lateralização se encontra totalmente estabelecido. A determinação deste marco

Na bibliografia especializada, encontram-se, no entanto, opiniões divergentes relativamente à pertinência da lateralização das funções linguísticas no cérebro como fator de determinação de um limite cronológico de um período ‘crítico’ (McLaughlin 1984; Singleton 2005:271). Partindo dos contributos de diferentes estudos empíricos, Krashen (1981:73), por exemplo, observa que “lateralization is "firmly established" much earlier, at least by age 5, and that the preconditions for lateralization may be present even at birth.”. Assim sendo, para este investigador, são as modificações que operam no plano cognitivo e afetivo, registadas na puberdade, i.e., sensivelmente aos 12 anos de idade, que condicionam necessariamente o processo de assimilação das estruturas linguísticas. Mais especificamente, o autor considera que as mudanças ocorridas neste período do desenvolvimento do falante “may have the effect of boosting language *learning* potential, while limiting or weakening the language *acquisition* potential” (1981:76), estabelecendo, então, uma relação entre a idade do falante aprendente e os processos associados à aquisição linguística, automática e inconsciente, por um lado e, por outro, à aprendizagem, na qual é assumido que o aprendente tem consciência das diferentes etapas por que passa para assimilar a informação linguística.

A ausência de dados empíricos robustos que permitam comprovar a existência de um limite cronológico específico para a determinação de um período crítico para a aquisição linguística levou a que muitos autores colocassem em causa os fundamentos propostos pela perspetiva original da ‘Hipótese do Período Crítico’ (Hamers & Blanc 1990:223). Na verdade, os dados apurados em trabalhos de investigação sobre crianças privadas de *input* linguístico após a puberdade demonstraram que, mesmo nestas circunstâncias, há indícios de que é possível assimilar, mais tardiamente, determinadas áreas da estrutura linguística, nomeadamente itens vocabulares (McLaughlin 1984; Hamers & Blanc 1990), ainda que outras, relativas à estrutura gramatical, se revelem comprometidas.

---

cronológico resultou da análise de um conjunto de dados recolhidos em pacientes com afasias, tendo o investigador constatado que as possibilidades de recuperação depois da puberdade diminuem de forma considerável. Assim, e como afirma o autor (1967 [1985:179]): “[e]n la época de la pubertad se alcanza un punto de giro. Las afasias que se desarrollan a partir de esta edad o que no han tenido tiempo de desaparecer completamente hasta esta etapa, dejan normalmente rastro que el paciente no puede superar.”

No contexto da aquisição/aprendizagem de uma LNM, é geralmente aceite a ideia de que a idade do aprendente desempenha um papel fundamental para o processo de assimilação das estruturas linguísticas (Johnson & Newport 1989; Birdsong & Molis 2001). Em boa verdade, enquanto aprendentes precoces conseguem atingir, com relativa facilidade, níveis de proficiência e de competência linguísticas consideradas nativas em todos os domínios estruturados da LA de aprendizagem, o mesmo não se verifica habitualmente em aprendentes tardios (Larsen-Freeman & Long 1991 [1994]; Bley-Vroman 1990; Ritchie & Bhatia 1996; Selinker 1972 [1992]; Gass & Selinker 2008).

Assiste-se, porém, a um longo debate entre os membros da comunidade científica relativamente à pertinência dos fundamentos propostos na hipótese do período ‘crítico’ para o contexto da aquisição/aprendizagem de uma LNM (Ioup 2005; Herschensohn 2013). Por um lado, alguns investigadores assumem que o facto de um aprendente tardio dificilmente conseguir atingir níveis de proficiência e de competência linguísticas consideradas nativas em todos os domínios estruturados da LA é representativo da possibilidade de se estabelecer um limite cronológico para a assimilação das estruturas de uma LNM (veja-se, por exemplo, o trabalho seminal de Johnson & Newport 1989 e a replicação desse estudo por Birdsong & Molis 2001).

Por outro, há autores que rejeitam a aplicação de barreiras cronológicas para este efeito, uma vez que, se existisse, de facto, um limite ‘crítico’ para a aprendizagem de uma LNM, um falante adulto não seria capaz de aprender um ou mais idiomas depois de ter adquirido a sua LM<sup>108</sup> (McLaughlin 1984; Hamers & Blanc 1990; Bialystok & Miller 1999). Procurando evidências empíricas que suportem a hipótese do período crítico para a aquisição/aprendizagem de uma LNM, à luz do que defendem Johnson & Newport (1989), Bialystok & Miller (1999) analisaram dados recolhidos de várias tarefas realizadas por aprendentes precoces e aprendentes tardios de inglês como LNM, que são falantes de LM chinesa e de LM espanhola. Neste estudo, os investigadores replicaram a metodologia adotada por Johnson & Newport (1989), a fim de verificar o desempenho dos diferentes grupos de aprendentes, distinguidos em função da idade de início de exposição à LA e da

---

<sup>108</sup> Também Lenneberg (1967 [1985:206]) considera não ser possível limitar a aprendizagem de um idioma não nativo a um período cronológico, ao afirmar que “[l]a mayoría de los individuos de inteligencia media son capaces de aprender un segundo lenguaje después de comenzada la segunda década”. O autor entende ainda que, embora o falante possa iniciar tardiamente a aprendizagem de uma língua e comunicar nessa mesma língua, evidenciar-se-ão problemas característicos de uma aprendizagem tardia, ou seja, serão evidentes “bloqueos para el aprendizaje del lenguaje” (Lenneberg 1967 [1985:206]).

respetiva LM, em tarefas de produção e de compreensão (Bialystok & Miller 1999:131). Os resultados obtidos revelaram que, independentemente da idade de início de exposição à LA, os falantes nativos de espanhol apresentam um melhor desempenho global do que os falantes nativos de chinês. Por conseguinte, os autores argumentam que, se houvesse um período ‘crítico’ para o desenvolvimento da aquisição/aprendizagem de uma LNM, poder-se-ia esperar que tal barreira cronológica influenciasse, do mesmo modo, o desempenho dos diferentes grupos de aprendentes nas várias tarefas propostas (Bialystok & Miller 1999:140). Sugere-se, então, que para além da idade, outros fatores contribuem para o desempenho dos falantes aprendentes tardios, nomeadamente a configuração do conhecimento linguístico prévio.

Porém, mesmo entre os autores que rejeitam a existência de uma idade ‘crítica’ para o contexto da aquisição/aprendizagem de uma LNM, há os que consideram que a assimilação de determinados planos da estrutura de um idioma não nativo parece estar condicionada pelo período cronológico em que o falante aprendente inicia a exposição à LA (Herschensohn 2013:321). Entre os planos estruturados da língua, o plano fonológico é, talvez, um dos mais suscetíveis a uma idade ‘sensível’ para a aquisição linguística de um idioma não nativo (Lenneberg 1967 [1985:205]); Johnson & Newport 1989; Martins 2008). A este respeito, observa Lenneberg (1967 [1985:205]) que “[l]os acentos extranjeros no pueden superarse fácilmente después de la pubertad.”, i.e., findo o período ‘crítico’ da aquisição linguística. Do mesmo modo, Johnson & Newport (1989) constataram que falantes aprendentes precoces apresentam um melhor desempenho na aquisição fonológica do que os aprendentes tardios<sup>109</sup>.

Por sua vez, Slabakova (2006) defende que o domínio da semântica, sobretudo o domínio da semântica frásica, não é afetado pela idade em que o falante inicia a exposição à LA. Para a investigadora, não existem evidências empíricas robustas que comprovem um pior desempenho em tarefas de compreensão e de interpretação dos enunciados por aprendentes tardios comparativamente a aprendentes precoces (Slabakova 2006:30).

---

<sup>109</sup> Saliente-se que na postulação de um período cronológico ‘crítico’ ou ‘sensível’ para a aquisição fonológica é preciso distinguir dois planos: o da produção e o da receção linguísticas. A este respeito, afirma Martins (2008:165) que “na apresentação da problemática relativa à existência de uma idade crítica para a aquisição do sistema fonológico de L2 fica sempre implícita a ideia de que é plausível um bilingue tardio possuir um índice de percepção e discriminação fonológica em relação às unidades de L2 substancialmente superior à sua capacidade de realização fonética nativa dessas mesmas unidades.” O que parece, então, estar em causa não é a delimitação de um período cronológico específico para todos os planos da aquisição fonológica, mas antes o da identificação de uma idade ‘crítica’ para a aquisição, por parte do aprendente, dos movimentos articulatórios de que necessita para a produção fonética das unidades fonológicas da LNM.

Quanto à aquisição de propriedades morfossintáticas e, em particular, do género gramatical, vimos anteriormente (cf. Capítulo 1, Secção 1.5.) que a assimilação tardia desta categoria gramatical parece estar sujeita a uma idade ‘crítica’. Em função destas evidências, surgem, na literatura especializada, várias hipóteses teóricas que procuram compreender o facto de os aprendentes tardios evidenciarem um conhecimento incompleto e deficitário do domínio morfossintático da LA (cf. Hawkins & Chan 1997; Franceschina 2003, 2005; Tsimpli 2003; Tsimpli & Dimitrakopoulou 2007; Tsimpli & Mastropavlou 2008; veja-se ainda a Secção 2.3.3.2. do presente trabalho).

A *Interpretability Hypothesis* (Tsimpli 2003; Tsimpli & Dimitrakopoulou 2007; Tsimpli & Mastropavlou 2008) baseia-se no pressuposto teórico minimalista de Chomsky (1995 [1999:320]) de que as categorias morfossintáticas, como género e número nominais, correspondem a traços- $\phi$  (traços-*phi*) que, na interface semântica, são representados ora como traços interpretáveis, dotados de informação semântica, ora como não-interpretáveis, sem qualquer informação de natureza concetual. Segundo Lopes (2006:161), os traços interpretáveis fazem “parte de um léxico universal” da língua. Em contrapartida, os traços não-interpretáveis são apenas instrumentais para a computação sintática, i.e., para a realização de operações de concordância (Corrêa 2002:121). De acordo com os fundamentos postulados pela *Interpretability Hypothesis*, na distinção entre a aquisição linguística nativa e não nativa, avalia-se o grau de acessibilidade dos aprendentes aos traços interpretáveis da língua, por um lado, e, por outro, a inacessibilidade dos traços não-interpretáveis, por estes estarem sujeitos a um período crítico (Lopes 2006:163). Deste modo, Tsimpli & Dimitrakopoulou (2007:211) postulam que:

“interpretable features are accessible to the L2 learner whereas uninterpretable features are difficult to identify and analyze in the L2 input due to persistent, maturationally-based, L1 effects on adult L2 grammars.”

Portanto, é assumido que aprendentes tardios terão, inevitavelmente, maiores dificuldades no processamento e codificação da informação de natureza formal, não interpretável, em detrimento da informação semântica. Este pressuposto está também presente, e como veremos na Secção 2.2.2.2., nos modelos teóricos que analisam o processamento tardio do *input* linguístico à luz das respetivas propriedades semânticas e formais. Segundo VanPatten (2008:120), atendendo aos pressupostos postulados no âmbito da teoria do Processamento de *Input*, os aspetos semânticos de uma língua são os primeiros

a serem assimilados pelos falantes aprendentes tardios, sendo que os aspetos formais dos itens são secundários no processo de aquisição/aprendizagem de uma LNM.

O género é tipicamente referenciado como um traço interpretável no nome, não o sendo, todavia, nos elementos que com ele concordam, tais como nos determinantes e nos adjetivos (Chomsky 1995 [1999]; Carstens 2000). Assim, à luz dos fundamentos postulados no âmbito da hipótese da interpretabilidade, haverá necessariamente uma maior dificuldade na aquisição/aprendizagem tardia do estabelecimento das relações de concordância do que na assimilação do valor de género dos nomes (White *et al.* 2004:109; Sabourin *et al.* 2006:26). Pese embora o facto de Chomsky (1995 [1999]) e de outros investigadores, tais como Carstens (2000), assumirem o género como um traço interpretável no nome, há algumas reservas a ter em conta em relação a esta proposta. Com efeito, e como observa Carstens (2000:328, n.12), o género também pode ser perspectivado como um traço não-interpretável, já que se trata de uma categoria arbitrária que não afeta, necessariamente, a interpretação semântica do item nominal.

Por fim, as dificuldades inerentes à determinação de uma idade ‘crítica’ para a aquisição/aprendizagem de uma LNM levaram alguns estudiosos a considerar como mais plausível a noção de período ‘sensível’. Esta expressão denota, assim, a natureza gradual do desenvolvimento linguístico do falante aprendente tardio e coloca em evidência o facto de o processo de assimilação das estruturas linguísticas não depender só do momento cronológico em que se dá a exposição ao *input*, mas de outros fatores, tais como, por exemplo, os conhecimentos linguísticos prévios dos falantes aprendentes, fatores de natureza afetiva (personalidade, motivação do aprendente para a aprendizagem), bem como as especificidades do *input* linguístico ao qual o aprendente está exposto (Bialystok & Hakuta 1999; Bialystok & Miller 1999; Ioup 2005; Herschensohn 2013).

Pese embora o facto de não haver, entre a comunidade científica, um consenso relativamente ao papel da idade para a aquisição/aprendizagem de uma LNM, a verdade é que a assimilação plena dos diferentes planos estruturados parece estar sujeita ao momento cronológico em que o falante aprendente inicia a exposição à LA. Para além disso, o próprio sistema linguístico atualizado pelo aprendente tardio de uma LNM apresenta distintas particularidades, já que tipicamente não corresponde aos enunciados produzidos por um falante nativo dessa língua. Estas evidências deram origem ao conceito de ‘interlíngua’ (Selinker 1972 [1992]), que se define em seguida.



### 2.2.2. Interlândia

Termo central para a investigaão da ALNM, a ‘interlândia’<sup>110</sup> designa, em linhas gerais, o sistema lingüístico que o aprendiz tardio evidencia ao longo da assimilaão das estruturas de uma LNM (Selinker 1972 [1992]). Segundo Selinker, a interlândia corresponde a um sistema lingüístico único, permeável e sujeito a variaão, que os aprendentes tardios de uma LNM evidenciam<sup>111</sup> (Gass & Selinker 2008:14).

Do ponto de vista estrutural, as interlândias não correspondem nem às estruturas da língua de origem, i.e., da LM do aprendiz, nem às estruturas da LA de aprendizagem. As interlândias são consideradas sistemas lingüísticos uma vez que, durante o processo da sua construão, os fenômenos desviantes em relaão à gramática da LA de aprendizagem são, de certo modo, previsíveis. Na verdade, e à semelhana do que é postulado para a aquisião do idioma nativo, os desvios sistemáticos são representativos das diferentes etapas de desenvolvimento do processo de assimilaão das estruturas de uma LNM. Por outras palavras, os desvios sistemáticos evidenciam um processo gradual e transicional, podendo ser entendidos como um reflexo do que Corder (1967 [1992:36]) designa como o ‘programa interno’ (*built-in-syllabus*) do aprendiz. Assumindo, então, que o processo de aquisião/aprendizagem de uma LNM corresponde a um processo criativo, Corder (1967 [1992:36]) entende que os erros cometidos pelos aprendentes constituem evidências de que esses mesmos aprendentes vão colocando hipóteses sobre o *input* da LA. Deste modo, para o autor, é partindo da análise dos desvios sistemáticos que se pode comprovar a existêcia de um sistema lingüístico intermédio, a interlândia, condicionado por regras próprias.

O reconhecimento da interlândia pressupõe que os aprendentes terão interiorizado uma gramática mental, um sistema lingüístico abstrato que, regendo-se por um conjunto de princípios e de regras, vai condicionar, quer o domínio da compreensã, quer o domínio da produão de enunciados da LNM (Ellis 2003:33). Trata-se, assim, de um sistema transitório e em permanente reestruturaão, a que Ellis (2003:33) se refere como um *interlanguage*

---

<sup>110</sup> A formulaão deste conceito é atribuída a L. Selinker aquando da publicaão, em 1972, do artigo seminal intitulado *Interlanguage* e, desde então, constitui-se como um elemento essencial para a investigaão no campo da aquisião/aprendizagem de uma LNM (Larsen-Freeman & Long 1991 [1994]; Ellis 2000, 2003; Gass & Selinker 2008).

<sup>111</sup> Na literatura encontram-se ainda outras propostas terminológicas para designar o sistema lingüístico intermédio que o aprendiz tardio atualiza ao longo da aquisião/aprendizagem de uma LNM, tais como ‘dialecto idiossincrático’, expressã proposta por Corder (1971 [1992]) e ‘sistema aproximado’, proposto por Nemser (1971 [1992:52]). No entendimento de Larsen-Freeman & Long (1991 [1994:63]) a preferêcia pelo conceito de ‘interlândia’ na investigaão em ALNM deve-se, em parte, “a la actitud neutral que representa, ya que los otros términos connotaban un enfoque hacia la LM”.

*continuum*. Segundo Selinker (1972 [1992:82]), para a assimilação das estruturas de uma LNM, é mobilizada uma ‘estrutura psicológica latente’, um conceito que se baseia na noção de ‘estrutura latente da linguagem’ proposta por Lenneberg (1967 [1985]). A estrutura latente da linguagem conduz a uma competência e proficiência nativas do idioma que a atualiza e, conseqüentemente, não pode ser recrutada durante todas as fases do desenvolvimento do indivíduo (cf. Secção 2.2.1.). Assim, no contexto da aquisição tardia de uma LNM, e face à indisponibilidade da ‘estrutura latente da linguagem’ Selinker (1972 [1992:83]) considera que:

“los que aprenden una L2 activan una estructura diferente, aunque también genéticamente determinada, cada vez que intentan expresar significados que poseen previamente en la lengua que están aprendiendo”.

Segundo Selinker (1972 [1992]), na construção das interlínguas estão envolvidas distintas estratégias e processos de natureza psicolinguística, que se caracterizam sumariamente em seguida.

Ao longo da assimilação das estruturas da LA, o aprendente recorre a diferentes estratégias de aprendizagem. Definir o que constitui uma ‘estratégia’ é, contudo, tarefa difícil, já que “sabemos muy poco acerca de las estrategias que utilizan los hablantes” (Selinker 1972 [1992:88]). Como exemplo de estratégia de aprendizagem, Selinker refere-se ao facto de, mesmo inconscientemente, o aprendente tender à simplificação do material linguístico da LA, omitindo, por exemplo, o uso de determinadas estruturas gramaticais, tais como, por ex., os artigos e as marcas de flexão verbal e nominal (Selinker 1972 [1992:88]). Neste contexto, observa-se igualmente que a transferência de estratégias mobilizadas na instrução formal atuam também como estratégia de simplificação que são, portanto, induzidas pelo modo como o *input* é apresentado ao aprendente tardio em contexto instrucional<sup>112</sup>.

---

<sup>112</sup> Efetivamente, alguns recursos didáticos podem condicionar a percepção dos aprendentes e, por conseguinte, conduzi-los a produções não convergentes com a língua-alvo. Selinker (1972 [1992:87-88]) exemplifica este processo com um curioso caso verificado num conjunto de falantes nativos de servo-croata, aprendentes tardios de inglês como LNM. Estes aprendentes manifestavam dificuldades em distinguir os pronomes pessoais *he/she*, situação que, segundo o autor, decorre do facto de, quer os manuais didáticos, quer os professores formularem exercícios apenas com o recurso ao pronome masculino *he*. Evocando este mecanismo como participante na construção das interlínguas Selinker (1972 [1992]) atribui, concomitantemente, um papel relevante à experiência instrucional para a configuração do conhecimento do aprendente tardio relativamente à LNM.

Outro processo atuante na construção das interlínguas diz respeito à sobregeneralização das regras da LA. Este pode também ser entendido como uma estratégia de simplificação do material linguístico da LA por parte dos aprendentes (Selinker 1972 [1992]). A tendência recorrente para simplificar o material linguístico pode, por sua vez, conduzir à ‘estabilização’ de determinadas estruturas linguísticas, levando, em última instância, ao cessar do desenvolvimento da interlíngua.

A par das estratégias de aprendizagem, as estratégias de comunicação contribuem igualmente para a simplificação do material linguístico da LA (Gass & Selinker 2008:439). Na verdade, numa situação comunicativa, o aprendente pode optar por evitar o uso de determinadas estruturas, por considerar que estas não são essenciais para os seus objetivos comunicativos. Na assimilação do sistema de atribuição de género nominal em português como LNM, dados empíricos recolhidos em diferentes trabalhos (Godinho 2010; Ferreira 2011; Martins 2015; Pinto 2015) revelam que os aprendentes tendem, de um modo geral, a ‘ignorar’ as marcas de atribuição de género nos especificadores, sobretudo nos determinantes. Ora, esta tendência, observada em aprendentes tardios de PLNM com diferentes perfis linguísticos, poderá estar correlacionada com o facto de o aprendente considerar que estas classes de palavras têm, do ponto de vista semântico, pouco valor comunicativo, já que, “mesmo “ignorando” estas marcas, consegue atingir os seus objetivos comunicativos” (Ferreira 2011:65).

Por fim, na construção da interlíngua atua ainda o processo de transferência linguística que, segundo esta perspetiva, corresponde à presença de desvios e/ou de outras manifestações na produção dos aprendentes tardios relacionada com a influência da configuração do conhecimento linguístico prévio que provém, nomeadamente da LM (Selinker 1972 [1992]; Ellis 2003:51). Na Secção 2.3., e face aos objetivos do presente estudo, será feita uma análise detalhada da noção de transferência linguística, bem como das suas implicações para o desenvolvimento das interlínguas dos falantes aprendentes e, mais especificamente, para a aquisição/aprendizagem da categoria de género gramatical numa LNM.

Convém ainda assinalar que a interlíngua não corresponde a um sistema linguístico que apenas integra manifestações não convergentes com a LA de aprendizagem. Em boa verdade, e como já referido anteriormente, este sistema é o resultado das hipóteses que, em diferentes estádios de desenvolvimento do processo aquisitivo, o aprendente tardio coloca sobre o *input* linguístico (Corder 1967 [1992]; Selinker 1972 [1992]; Ellis 2003; Gass &

Selinker 2008). Sendo certo que apenas uma ínfima percentagem de aprendentes atinge um grau de competência e de proficiência linguísticas consideradas nativas<sup>113</sup>, observa-se que uma parte substancial das manifestações linguísticas da interlíngua são, na verdade, convergentes com a LA, evidências, aliás, assinaladas em muitos estudos (Gass & Selinker 2008). Os fenómenos não convergentes com a LA podem, por seu turno, reaparecer após um período de ‘remissão’ apresentando um movimento de *backsliding*, i.e., o reaparecimento de desvios que já se pensavam erradicados das interlínguas dos aprendentes (Selinker 1972 [1992]; Gass & Selinker 1993, 2008; Ellis 2003). De facto, e apesar de uma contínua exposição a *input* linguístico, de uma forte motivação para aprender o novo idioma e de suficientes oportunidades para comunicar nessa língua, os aprendentes tardios demonstram muitas dificuldades na progressão da aquisição/aprendizagem de uma LNM (Selinker 1972 [1992]; Bley-Vroman 1990; Larsen-Freeman & Long 1991 [1994]; Gass & Selinker 1993, 2008; Ellis 2003). Tendo em conta estas evidências, o conceito de ‘interlíngua’ surge associado à noção de ‘fossilização’ (Selinker 1972 [1992]).

Em linhas gerais, os fenómenos ‘fossilizados’ caracterizam-se pela persistência de comportamentos não convergentes aos da LA no sistema interlinguístico dos aprendentes e que são impeditivos do desenvolvimento do processo aquisitivo<sup>114</sup>. Em última instância, a fossilização é considerada inevitável e permanente<sup>115</sup> (Selinker 1972 [1992:85]; Bley-Vroman 1990:9).

Apesar de, na prática, os fenómenos fossilizados serem difíceis de identificar, é possível recorrer a alguns indícios. Por exemplo, Selinker & Lakshmanan (1993:197) argumentam que a presença constante de estruturas próprias da interlíngua na produção de enunciados da LA constitui evidência empírica de estruturas fossilizadas. Contudo, e na perspetiva de Ellis (1986), a fossilização resulta não só em formas desviantes com a estrutura da LA, mas também em formas convergentes<sup>116</sup>. Este autor assume, então, que a

---

<sup>113</sup> Segundo Selinker (1972 [1992:83]), apenas 5% dos aprendentes tardios consegue atingir níveis de proficiência superiores em todos os planos estruturados da LA.

<sup>114</sup> Também Weinreich (1968:48) já havia assinalado a presença de fenómenos de “permanent grammatical interference” sem, contudo, se referir especificamente ao conceito de ‘fossilização’, termo usado pela primeira vez por L. Selinker.

<sup>115</sup> Para Selinker (1972 [1992:85]), a fossilização é um mecanismo próprio da estrutura psicológica latente e os fenómenos fossilizados correspondem a “ítems, reglas y subsistemas lingüísticos que los hablantes de una LM particular tienden a conservar en su IL (...) sin importar cuál sea la edad del alumno o cuánto entrenamiento haya recibido (...). Es importante observar que las estructuras fosilizables tienden a permanecer (...), reemergiendo en la producción de una IL incluso cuando parecían ya erradicadas.”

<sup>116</sup> A este respeito, afirma Ellis (1986:48) que “[i]f when fossilization occurs, the learner has reached a stage of development in which feature X in his interlanguage has assumed the same form as in target-

fossilização é um processo característico do desenvolvimento interlinguístico, uma vez que os fenómenos fossilizados, desviantes ou não, ocorrem inevitavelmente a determinado ponto do processo<sup>117</sup> e, por conseguinte, estende a fossilização a todos os casos em que as regras gramaticais – compreendidas no seu sentido mais lato – se fixam de um modo permanente na gramática da interlíngua do aprendente.

A extensão da fossilização a formas convergentes não é, porém, inteiramente aceite. Bley-Vroman (1990), por exemplo, assume que a fossilização corresponde à presença constante de estruturas não convergentes com a LA nos enunciados, constituindo, para além disso, um importante fator para distinguir o desempenho linguístico de aprendentes precoces dos aprendentes tardios, na medida em que, na aquisição da LM por crianças, “there is no fossilization (short of success). Stages are inevitably passed through, the system remains plastic until success is achieved.” (1990:10).

Em alternativa à noção de ‘fossilização’, Long (2003:521) considera como mais operativo o conceito de ‘estabilização’, estado que, em todo o caso, antecederá o de uma eventual fossilização. Tal posição decorre do facto de o investigador ter verificado que os fenómenos fossilizados são, em rigor, difíceis de detetar empiricamente.

Os dados empíricos relativos à aquisição/aprendizagem tardia de género gramatical em PLNM (Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Mariotto 2014; Lacsán 2015; Martins 2015; Pinto 2015), e também em outras LNM (Oliphant 1997; White *et al.* 2004; Franceschina 2005; Sabourin *et al.* 2006; Foucart 2008), revelam a persistência dos desvios relativos a esta estrutura mesmo em aprendentes de níveis de proficiência linguística mais avançados. Por conseguinte, poder-se-á assumir que esta categoria gramatical é suscetível a um estágio de estabilização, e até, porventura, de fossilização, sendo necessário compreender os fatores que conduzem a categoria de género a esta situação.

Como se depreende, a noção de interlíngua integra um conjunto de pressupostos relativos ao dinamismo da aquisição linguística tardia e, na linha do que afirma Ellis

---

language, then fossilization of correct form will occur. If, however, the learner has reached a stage of development in which feature Y still does not have the same form as the target-language, the fossilization will manifest itself as a error”.

<sup>117</sup> Saliente-se, no entanto, que a perspetiva adotada por Ellis (1986) espelha a definição de fossilização proposta por Vigil & Oller em 1976. Segundo estes autores “[w]e will extend the notion of fossilization to any case of grammatical rules, construed in the broadest sense become relatively permanently incorporated into a psychological grammar. (...) It is not only the fossilization of so called ‘errors’ that must be explained, but also the fossilization of correct forms that conform to the target language norms.” (1976:282 *apud* Han 2004:17).

(2003:35), este conceito “can be viewed has a metaphor of how L2 acquisition takes place.”. Percorrendo a bibliografia especializada, é possível encontrar um conjunto vasto de contributos para a compreensão dos processos envolvidos na construção das interlínguas que, em função do posicionamento teórico de que emanam, atribuem papéis distintos a diversos fatores para justificar a relativa ‘incapacidade’ do aprendente tardio para a aquisição plena dos diferentes planos estruturados da LA (Gass & Selinker 2008). Assim, na caracterização da interlíngua, é necessário ter em linha de conta as implicações teóricas associadas à configuração do conhecimento linguístico inato, às características do *input* e aos mecanismos cognitivos envolvidos no processamento da informação linguística. A consideração destes aspetos é essencial para a posterior análise do fenómeno da influência do conhecimento linguístico prévio para a construção das interlínguas, bem como para compreender as hipóteses teóricas que, a este respeito, são colocadas.

#### 2.2.2.1. O papel da Gramática Universal

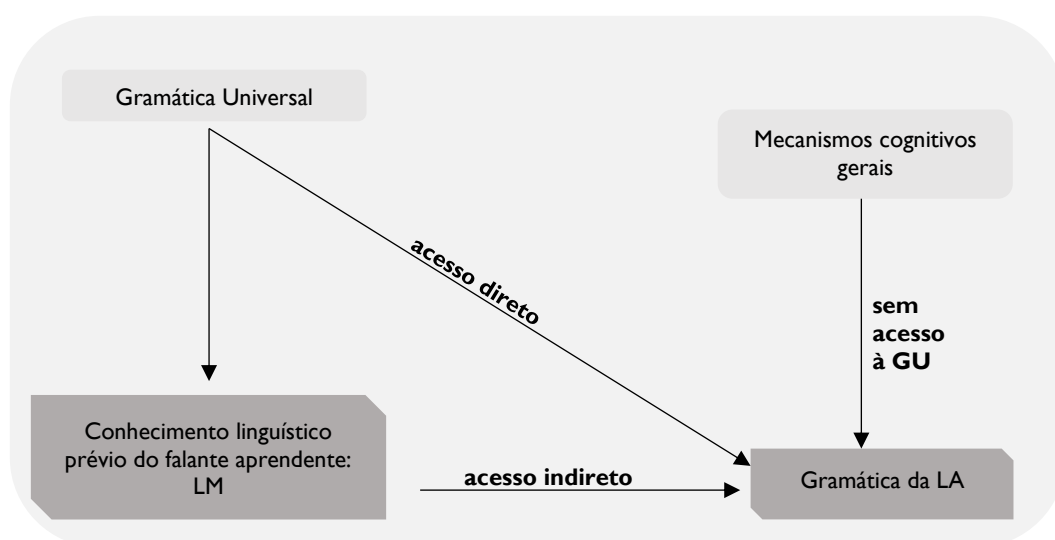
Os investigadores que perfilham o quadro teórico da gramática generativa assumem que a construção das interlínguas e a assimilação das estruturas da LA está dependente do grau de acessibilidade (direto/indireto e parcial/total) a componentes específicas que configuram a Gramática Universal (GU) (Ellis 1986; White 1993, 2003; Cook 1994; Schachter 1996; Gass 1996; Gass & Selinker 2008).

A GU faz parte de um modelo teórico relativo à natureza do conhecimento linguístico e à aquisição linguística que foi postulado por N. Chomsky (1965 [1975]), e no quadro do qual se reconhece a existência de um dispositivo cognitivo inato e especificamente programado para a aquisição linguística, denominado *Language Acquisition Device* (LAD). Na GU estão integradas as informações relativas à forma possível que qualquer gramática pode assumir.

Em linhas gerais, a GU corresponde a um conjunto de princípios e de parâmetros que configuram uma estrutura mental inata para a linguagem (Chomsky 1965 [1975]; Cook 1985, 1994; White 1993, 2003, 2008) e define-se como “a set of general principles that apply to all grammars and that leave certain parameters open” (Cook 1985:3). De acordo com esta perspetiva, os princípios são responsáveis pelos aspetos comuns a todos os idiomas do mundo e os parâmetros justificam a variação registada (Chomsky 1965 [1975]; Cook 1994:25; White 2008:42).

Desde muito cedo que a teoria chomskyana sobre a aquisição linguística acolheu um especial interesse para o campo da investigação sobre a aquisição de línguas não-maternas (White 1990:50; Gass 1996:330; Ritchie & Bhatia 1996:7). De facto, investigadores que perfilham este quadro teórico (White 1990, 1993, 1995, 2003, 2008; Cook 1994; Vainikka & Young-Scholten 1994; Flynn 1996; Schachter 1996; Schwartz & Sprouse 1996; Hawkins & Chan 1997; Franceschina 2003, 2005; White *et al.* 2004) procuram avaliar se os princípios e parâmetros da GU continuam inteiramente disponíveis durante o processo de assimilação tardia das estruturas de uma LNM, fazendo previsões acerca da natureza das representações gramaticais dos falantes aprendentes tardios, com assunções específicas sobre o desenvolvimento aquisitivo das diferentes categorias gramaticais.

Neste quadro teórico são colocadas diferentes hipóteses relativamente ao ‘estádio inicial’ da aquisição/aprendizagem de uma LNM, não havendo igualmente consenso no que respeita ao grau de acessibilidade (direto/indireto; total/parcial) por parte dos aprendentes aos princípios e parâmetros da GU (Cook 1994:34; Gass 1996; White 2008:39). Observam-se, a este respeito, hipóteses nas quais se postula: (i) o acesso pleno e direto à GU (White 1990, 2008; Flynn 1996); (ii) o não acesso à GU (Bley-Vroman 1990; Schachter 1996); e (iii) o acesso parcial (direto ou indireto) à GU, através da configuração da gramática do idioma nativo (Hawkins & Chan 1997; Tsimpli 2003; Franceschina 2003, 2005) (cf. Figura 2.1).



**Figura 2.1** – Grau de acessibilidade da GU no contexto de aquisição/aprendizagem de uma LNM (figura baseada na proposta de Cook (1994:33))

Destas hipóteses decorrem, conseqüentemente, perspectivas diversas acerca dos padrões de desenvolvimento interlingüístico. Com efeito, os defensores do pleno acesso à GU advogam que o processo de aquisição/aprendizagem de uma LNM é, de certo modo, semelhante ao da aquisição do idioma nativo, na medida em que a GU atua na configuração da representação da gramática mental do falante aprendente tardio<sup>118</sup>. Todavia, o acesso pleno à GU não implica, necessariamente, que o aprendente tardio atinja um grau de competência e de proficiência linguísticas consideradas nativas em todos os planos estruturados da LA. Assim, e nas palavras de White (2008:46):

“the claim that interlanguage grammars are UG-constrained is a claim that the linguistic representations of the L2[learners] are subject to principles of UG, like other natural languages. It is not a claim that L2ers will necessary achieve the same grammar as a native speaker would. (...) Many factors come into play in L2 that simply not arise in L1 acquisition – including prior knowledge of another language and possible deficiencies in the input – which may prevent native-like attainment.”

Em contrapartida, alguns investigadores assumem que o processo de assimilação linguística por falantes nativos e não nativos é fundamentalmente distinto<sup>119</sup> (Bley-Vroman 1990:23; Schachter 1996:187). Neste contexto, assume-se que o aprendente tardio, apesar de prejudicado pela impossibilidade de aceder à GU ou a mecanismos que atuam no âmbito do dispositivo cognitivo para a aquisição linguística, o LAD, pode, em alternativa, compensar esses constrangimentos através dos conhecimentos linguísticos prévios e de outros mecanismos cognitivos gerais que tem à sua disposição. Assume-se, portanto, que a assimilação das estruturas de uma LNM depende de variáveis relacionadas, quer com o próprio indivíduo, associadas à personalidade, à predisposição para aprender idiomas e à

---

<sup>118</sup> Num estudo conduzido por Flynn (1996), a autora verificou que, em contexto de imersão linguística, um conjunto de aprendentes de LM japonesa conseguiu adquirir, com sucesso, certas propriedades gramaticais do inglês como LNM, apesar de tais propriedades não se encontrarem representadas na língua japonesa (Flynn 1996:150). A partir destes dados, a investigadora assume, assim, que a faculdade inata para a aquisição linguística continua ativa durante a aquisição/aprendizagem tardia de uma LNM.

<sup>119</sup> Bley-Vroman (1990) constata ainda que, na assimilação tardia das estruturas de uma LNM, os aprendentes tardios atingem um grau de desenvolvimento inferior ao que é alcançado por um falante nativo, não conseguindo, então, progredir na aprendizagem. Por outras palavras, este autor assume que a ‘estabilização’ ou ‘fossilização’ da progressão da aprendizagem das estruturas da LA é inevitável. Neste contexto, o autor (1990:10) admite que a presença constante de fenómenos fossilizados na aprendizagem tardia de uma LNM “constitutes a serious obstacle to the assertion that adult and child language acquisition are fundamentally the same.”



motivação, quer com o próprio processo de aprendizagem, caracterizado pela instrução formal e pela ocorrência de *feedback* corretivo (Bley-Vroman 1990).

Para além destas posições diametralmente opostas, uma parte substancial da investigação produzida à luz dos pressupostos do modelo gerativo avança com hipóteses teóricas sobre a possibilidade de haver acesso parcial à GU, seja direto ou indireto, pela configuração da gramática do idioma nativo do falante aprendente (Hawkins & Chan 1997; Ellis 2000, 2003; Tsimpli 2003; Franceschina 2003, 2005). Tais hipóteses acarretam implicações várias para o modo como é feita a assimilação dos diferentes planos estruturados da LA de aprendizagem e, em particular, para a assimilação da categoria de género gramatical. Uma vez que as considerações teóricas a este respeito integram posições muito específicas acerca da influência dos conhecimentos linguísticos prévios para a construção das interlínguas e das condições promotoras e inibidoras da transferibilidade das propriedades gramaticais, retomaremos esta discussão na Secção 2.3.3.2.

#### 2.2.2.2. O papel do *input* linguístico

Nas diferentes teorias, é tipicamente reconhecida a necessidade de exposição a *input* linguístico para a aquisição/aprendizagem de uma LNM, pese embora o facto de não haver consenso no que concerne à sua relevância para o desenvolvimento do processo aquisitivo (Ellis 2000:243; Carroll 2001:8; Gass & Selinker 2008:304). Em linhas gerais, o *input* define-se como o conjunto de estruturas linguísticas a que o falante aprendente está exposto e a partir do qual são formuladas hipóteses (Crossley *et al.* 2014:303).

Corder (1967 [1992]) foi um dos primeiros investigadores a chamar a atenção para os processos subjacentes à transformação do *input* a *intake*. Assim, e nas palavras do investigador (1967 [1992:36]):

“[u]na forma lingüística no adquire necesariamente el estatuto de dato de entrada por el mero hecho de que le presente al alumno en el aula, porque dichos datos son los que “entran”, no los que están disponibles para entrar, y parece razonable suponer que es el alumno quien controla dichos datos, o, para ser más exacto, su toma (“intake”).”

Assumindo a existência de um dispositivo inato para a aquisição linguística, Krashen (1982) defende que a ativação desse mesmo dispositivo depende do grau de exposição do aprendente a *input* considerado compreensível e é com base neste pressuposto que o autor formula a Hipótese do *Input* (*Input Hypothesis*). Segundo esta hipótese, a

aquisição das estruturas linguísticas ocorre sempre que haja exposição a *input* com formas e estruturas que vão mais além do estado atual de compreensão da língua por parte do aprendiz (Krashen 1982:20). Para Krashen, em sala de aula, o professor deve garantir o fornecimento do *input* compreensível. Todavia, o autor não esclarece que condições estão subjacentes, quer à configuração deste tipo particular de *input*, quer à sua interação com o respetivo estágio de desenvolvimento do aprendiz, tornando, assim, difícil a sua aplicabilidade empírica (para uma revisão crítica desta proposta, veja-se, por exemplo, McLaughlin (1978 [1992])).

Modelos teóricos como a Teoria de Processamento do *Input* (*Input Processing Theory*) de VanPatten (1996, 2008) e a Teoria da Processabilidade (*Processability Theory*) de Pienemann (2008a, 2008b) refletem especificamente sobre a natureza do processamento do *input* linguístico e sobre o modo como decorre a transformação do *input* a *intake* (Corder 1967 [1992]). VanPatten (2008) apresenta uma série de princípios subjacentes ao processamento da informação linguística, definidos de acordo com as características do *input*. Para este investigador, o foco inicial do aprendiz incidirá na função do item linguístico e só depois na sua forma, sendo que o valor comunicativo da forma é essencial para que seja processada e integrada na gramática da interlíngua (VanPatten 2008:120). Já a Teoria da Processabilidade de Pienemann (2008b) baseia-se no pressuposto de que há diferentes estágios na construção da representação gramatical da LA que envolvem a integração gradual dos itens linguísticos pelo processador. A passagem de um estágio de desenvolvimento para o seguinte é sequencial, não sendo possível ‘saltar’ etapas nesse processo (2008a:141). Portanto, a lógica subjacente a este modelo teórico é a de que, em qualquer estágio do desenvolvimento, o aprendiz produz e compreende apenas as formas e estruturas linguísticas da LA de aprendizagem que o estágio atual do processador linguístico é capaz de assimilar.

Assim sendo, as tarefas de compreensão e de produção dos enunciados da LNM estão intrinsecamente dependentes não só das propriedades do *input* que favorecem a sua integração na gramática dos aprendizes, relacionadas, por exemplo, com a frequência dos itens lexicais e com a sistematicidade das estruturas linguísticas (Ellis 2002), mas também com as próprias limitações e capacidades do processador linguístico do aprendiz (VanPatten 2008; Pienemann 2008a, 2008b; Barcroft & Wong 2013:637).

Segundo Martins (2015:28) o processador:

“consiste num dispositivo complexo que convoca o contributo da atenção e da memória de curto prazo, que são determinantes para a extração da informação relevante do input ao qual os aprendentes estão expostos, mas também as interações entre estes sistemas cognitivos e os subsistemas de memória de longo prazo, que são cruciais para a codificação, o armazenamento e posterior recuperação das representações linguísticas (...).”

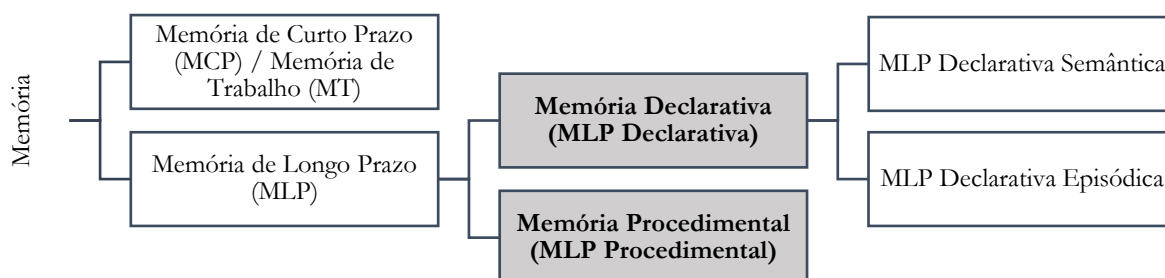
Estes modelos delimitam, portanto, um quadro geral para a compreensão das diferentes fases do desenvolvimento interlinguístico, concedendo especial atenção à influência das características dos dados linguísticos à disposição dos aprendentes para a sua aquisição (Carroll 2001). Para além disso, permitem uma análise mais fina relativamente às características associadas à assimilação dos diferentes planos estruturados dos idiomas e às condições que justificam uma maior ‘resistência’ de determinadas estruturas linguísticas ao processo aquisitivo. No contexto da análise da aquisição/aprendizagem da categoria de género numa LNM, constata-se que as propriedades semânticas e formais dos itens acarretam implicações para a atribuição de um valor de género por parte dos aprendentes tardios. A este respeito, dados apurados em diferentes estudos, previamente considerados neste trabalho, têm dado conta, precisamente, dos diferentes padrões associados à atribuição dos valores de género em função das especificidades dos respetivos itens lexicais e gramaticais.

Para além do papel das propriedades do *input* para a assimilação linguística, importa ter em linha de conta o papel de mecanismos cognitivos, nomeadamente, o dos subsistemas de memória envolvidos no processamento e codificação da informação linguística.

#### **2.2.2.3. O papel diferenciado dos subsistemas de memória: o modelo declarativo/procedimental**

Embora a memória se apresente como um sistema com uma organização verdadeiramente complexa (cf. Figura 2.2) e existam estudos sobre o papel, quer da Memória de Curto Prazo (MCP), quer da Memória de Longo Prazo (MLP) para a assimilação linguística, na presente Secção deste trabalho dar-se-á especial atenção ao modelo declarativo/procedimental (Ullman 2001a, 2004). Este modelo teórico assenta,

justamente, no pressuposto de que dois subsistemas da MLP, a MLP Declarativa e a MLP Procedimental, assumem papéis distintos na aquisição das estruturas linguísticas da LM e na aquisição/aprendizagem tardia da LNM, contribuindo, assim, para compreender as especificidades inerentes a um e a outro processo.



**Figura 2.2** – Subsistemas de memória (esquema baseado em Baddeley (1999:1))

Cada um dos subsistemas de memória é responsável pela realização de tarefas distintas, sendo que a MLP Procedimental e a MLP Declarativa desempenham um papel fundamental para a assimilação e codificação linguísticas (Ullman 2001a, 2001b, 2001c, 2004; Paradis 2004). A MLP Declarativa abrange o conhecimento explícito sobre factos e eventos relacionados com o mundo e experienciados pelo próprio indivíduo (Ullman 2001a, 2004; Paradis 2004:8). As conceptualizações armazenadas neste subsistema de memória são flexíveis, explícitas e, em grande medida, conscientemente manipuláveis (Ullman 2001b:106; Martins 2016:106). Por sua vez, a MLP Procedimental abrange representações envolvidas na produção de tarefas motoras e cognitivas de natureza implícita e automática (Ullman 2001a, 2001b, 2001c, 2004; Paradis 2004:7), constituindo-se como um conjunto de ‘saberes como’ que não são acessíveis ao nível do consciente<sup>120</sup>.

No modelo declarativo/procedimental (Ullman 2001a, 2004), pressupõe-se que a linguagem depende de duas capacidades mentais distintas: (i) de memorização, com

<sup>120</sup> Atendendo às características das representações associadas a cada um dos subsistemas de memória de longo prazo, a bibliografia especializada tende à associação, nem sempre consensual, dos termos dicotómicos ‘declarativo’ e ‘explícito’, por um lado, e ‘procedimental’ e ‘implícito’, por outro. A este respeito, observa Martins (2008:63) que a rejeição, por parte de alguns investigadores, da equivalência concetual dos termos “abre caminho para que se quebre a associação tradicional e exclusiva entre saberes explícitos e declarativos, (...) por via da abertura da possibilidade de estes últimos poderem assumir, tal como os procedimentais, um carácter implícito.” Pese embora este dado, neste trabalho optou-se por não estabelecer uma distinção concetual entre os termos (veja-se, a este respeito, Martins 2008).

incidência no léxico mental; e (ii) de computação, que diz respeito à gramática mental. O léxico mental define-se como o acervo memorizado das características idiossincráticas das palavras de uma língua<sup>121</sup> (Ullman 2004). Já a gramática mental envolve a capacidade de computação de regras necessárias à manipulação das palavras e das representações de estruturas linguísticas mais complexas<sup>122</sup> e compreende, assim, “o conhecimento linguístico implícito da fonologia, morfologia, sintaxe e certos aspetos (composicionais) da semântica” (Martins 2016:107). Por conseguinte, Ullman (2001c:717) associa o processamento da informação da gramática mental a módulos ‘encapsulados’, constituídos por mecanismos cognitivos inacessíveis à consciência.

Na aquisição linguística nativa, é assumido que a MLP Declarativa é responsável pela aquisição, representação e uso do léxico mental e, mais especificamente, de todas as propriedades idiossincráticas do léxico (como a associação forma – significado). Este subsistema de memória abrange ainda o conhecimento metalinguístico explícito, que se afigura essencial para o uso de estruturas e de representações que configuram a competência linguística implícita (Paradis 2004:11), na medida em que as representações são mobilizadas aquando da monitorização dos usos linguísticos (Martins 2008). A MLP Procedimental abrange, por sua vez, a computação da gramática mental (Ullman 2001c), sendo, portanto, determinante para a configuração da competência linguística implícita (Paradis 2004:9).

No contexto da aquisição/aprendizagem tardia de uma LNM, Ullman (2001b) assume que a MLP Procedimental deixa de estar inteiramente disponível para a computação da gramática mental. Assim, e nas palavras deste autor (2001b:109):

“linguistic forms that are computed grammatically in procedural memory in L1 may depend largely on declarative/lexical memory in L2”.

---

<sup>121</sup> Segundo Ullman (2004:233), o léxico mental corresponde a: “all idiosyncratic word-specific information. Thus, it includes all words whose phonological forms and meanings cannot be derived from each other (i.e. their sound meaning pairings are arbitrary) (...). It also contains other irregular— i.e. not entirely derivable — word-specific information, such as the particular arguments that must accompany a given verb (...) and any unpredictable forms that a word takes (...). The mental lexicon may comprise other distinctive information as well, smaller or larger than words: bound morphemes (...), and representations of complex linguistic structures whose meanings cannot be transparently derived from their parts (...)”.

<sup>122</sup> A gramática mental envolve um conjunto de regras que, de acordo com Ullman (2001c:717): “constrain how lexical forms can combine to make complex representations, and allow us to interpret the meanings of complex forms even if we have no heard or seen them before”.

Portanto, tanto o léxico como a gramática mental passarão, pelo menos em fases iniciais do desenvolvimento interlinguístico dos aprendentes, a ser suportados pela MLP Declarativa. Contudo, Ullman (2001b:110) não anula a hipótese de a MLP Procedimental servir a computação de certos aspetos da gramática mental da LNM. Ou seja, é assumido que é possível ao aprendente tardio atingir uma certa ‘automatização’ do uso de estruturas da LNM, ainda que, e de acordo com Morgan-Short *et al.* (2010:157), “such “proceduralization” will depend on a number of factors, including the type and the amount of L2 experience and training, as well as individual differences, such as procedural learning abilities”.

No que à aquisição do género diz respeito, os dados empíricos disponíveis são reveladores de que o processamento dos sistemas de atribuição de género e da concordância é diferente nos falantes nativos e nos aprendentes tardios, sobretudo quando estes se encontram em fases iniciais do seu desenvolvimento interlinguístico (cf. Capítulo 1, Secção 1.5.). Estudos sobre os mecanismos cognitivos envolvidos no processamento dos sistemas de atribuição de género gramatical por falantes nativos demonstram que esta categoria é processada por via da MLP Procedimental (Foucart & Frenck-Mestre 2011) e, por contraste, na aquisição/aprendizagem do género numa LNM, o processamento desta categoria é feito por via da MLP Declarativa (Morgan-Short *et al.* 2010), justificando-se, então, a partir destes indícios, as diferenças substanciais detetadas entre a aquisição desta estrutura por falantes nativos e por falantes não nativos.

## 2.3. O papel dos conhecimentos linguísticos prévios para a construção das interlínguas

### 2.3.1. Para uma definição de transferência linguística

A investigação levada a cabo na área da ALNM procurou, desde muito cedo, compreender em que medida o conhecimento linguístico prévio dos aprendentes, sobretudo da sua LM, afeta as distintas fases do processo de assimilação das estruturas de uma LNM<sup>123</sup> (Odlin 1989, 2013; Ellis 1986, 2003; Gass & Selinker 1993, 2008; Gass 1996; Alonso 2002). É essencialmente na década de 50 que a influência interlinguística passou a ser referida como ‘transferência linguística’ (*linguistic transfer*) e, a partir desse momento, este conceito foi sendo sucessivamente reconfigurado pelas diferentes correntes teóricas em que é utilizado. Por conseguinte, um dos problemas colocados ao estudo da transferência linguística prende-se com o facto de não existir, na literatura, consenso no que respeita ao recorte concetual deste termo (Odlin 1989; Gass 1996). Para além disso, também proliferam termos diferentes para designar fenómenos que, genericamente, se podem albergar sob o termo ‘transferência’.

Com efeito, nem sempre as diferentes correntes teóricas adotaram o termo ‘transferência’ para se referirem à influência dos conhecimentos linguísticos prévios para a assimilação das estruturas de uma LNM. Este conceito, importado da Psicologia, e introduzido na investigação em ALNM pelos behavioristas, designava, nesse quadro teórico, a presença de informação linguística da LM nos enunciados da LA de aprendizagem (Lado 1957). Esta visão essencialmente negativa de transferência foi, desde cedo, colocada em causa (Ellis 1986) e, por este motivo, alguns investigadores propuseram, em alternativa, expressões consideradas ‘neutras’, i.e., menos marcadas por uma tradição teórica específica (Gass 1996:318), preferindo referir-se ao ‘papel da língua materna’<sup>124</sup>

---

<sup>123</sup> As primeiras referências à influência interlinguística remontam aos trabalhos no âmbito do bilinguismo e do estudo das línguas em contacto (Martins 1997). A este respeito, refira-se o trabalho pioneiro de Weinreich (1968) que, em 1953, na publicação da 1ª edição da obra *Languages in contact: findings and problems*, utiliza a expressão ‘interferência linguística’ para se referir a (1968:1) “instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language”.

<sup>124</sup> Segundo Corder (1993), o uso da expressão ‘papel da língua materna’ para caracterizar a influência linguística na aquisição/aprendizagem de uma LNM tem algumas vantagens. Por um lado, esta expressão permite integrar no estudo da influência linguística outros fenómenos igualmente condicionantes na produção dos aprendentes. Entre esses fenómenos destaca-se, por exemplo, o de evitamento de uso de determinadas estruturas que, como Schachter (1974 [1992]) defende, poderá estar condicionado com a informação da LM do aprendente. Por outro lado, retira à influência linguística

para designar a influência do conhecimento linguístico prévio ao longo do desenvolvimento interlinguístico (Krashen 1981:64; Corder 1993:19). Já autores como Sharwood Smith & Kellerman (1986:1 *apud* Ellis 2000:301) sugerem como mais adequada a expressão ‘influência interlinguística’ (*crosslinguistic influence*), uma vez que esta coloca em evidência o facto de a aprendizagem de uma LNM poder ser condicionada, quer pelos conhecimentos provenientes da LM, quer por outras línguas previamente conhecidas pelos aprendentes<sup>125</sup> (Pinto 2012:175). Todavia, Odlin (2013:151) sublinha que:

“crosslinguistic influence (CLI) means roughly the same thing as certain other terms, including language transfer and interference: each signifies the metaphoric notion of one language influencing another”.

Do mesmo modo, Ringbom (2013:396) admite que, embora o termo ‘transferência’ seja, de certo modo, mais limitativo do que a expressão ‘influência interlinguística’, a verdade é que os dois são usados indiferenciadamente na bibliografia especializada.

A expressão ‘transferência linguística’ surge muitas vezes associada ao conceito de ‘empréstimo’ e, para Corder (1993), é preciso estabelecer uma distinção entre os dois termos<sup>126</sup>. Observa o autor que o ‘empréstimo’ é uma estratégia comunicativa, característica da *performance* dos aprendentes, i.e., do desempenho, verificável quando o falante recorre a itens provenientes da sua LM ou de outras línguas previamente conhecidas

---

uma certa conotação negativa que se foi disseminando nas primeiras abordagens teóricas à transferência. Deste modo, o sistema linguístico nativo deixa de ser encarado como um obstáculo à aprendizagem, configurando-se, antes, como “a heuristic tool in the discovery of the formal properties of the new language, facilitating the learning of those features which resemble features of the mother tongue” (Corder 1993:29).

<sup>125</sup> Também Ellis (2003:54) considera como mais adequada a expressão ‘influência interlinguística’, pois, para este investigador, o conceito de transferência “is yet another metaphor for explaining L2 acquisition. In some ways it is an inappropriate one. When we transfer money we move it out of one account and into another, so one account gains and the other loses. However, when language transfer takes place there is usually no loss of L1 knowledge. This obvious fact has led to the suggestion that a better term for referring to the effects of L1 might be ‘cross-linguistic influence’”.

<sup>126</sup> Já Weinreich (1968) alertara para a necessidade de se distinguirem as noções de ‘interferência linguística’ e de ‘empréstimo’. Para o investigador, o empréstimo apresenta um carácter mais acidental e superficial relativamente à interferência. Porém, nem sempre a literatura sobre o bilinguismo encarou o fenómeno de empréstimo tal como este é perspectivado por Weinreich, registando-se, neste domínio, uma certa “fluidez conceptual” (Martins 1997:88) entre os termos. Tal fluidez justifica-se, por um lado, pelo facto de a delimitação concetual de interferência de Weinreich estar muito próxima da definição de empréstimo proposta por Einar Haugen que, num artigo publicado em 1950, refere “borrowing is the attempted reproduction in one Language of patterns previously found in another” (1950:212). Face a esta delimitação geral da noção de empréstimo, facilmente se compreende a sua estreita aproximação à definição de interferência. Por outro lado, o facto de Weinreich ter incidido a sua análise em fenómenos de interferência ao nível do léxico aproxima este conceito do de empréstimo, já que também Haugen estuda os empréstimos do tipo lexical.



com o intuito de preencher lacunas que tem ao nível da representação gramatical da LA de aprendizagem<sup>127</sup> (Corder 1993:27), ao passo que a influência interlinguística se encontra ao serviço do processo de aprendizagem. O ‘empréstimo’ é, portanto, um fenómeno de natureza accidental, sem implicações para o desenvolvimento da aprendizagem, ao contrário do que se observa na transferência linguística.

As diferentes propostas terminológicas acima referidas resultam de conceitualizações distintas deste fenómeno. Com efeito, numa perspetiva tradicional, baseada nos pressupostos teóricos behavioristas e da Análise Contrastiva (Lado 1957), a transferência linguística correspondia unicamente à presença de desvios na produção dos enunciados da LNM. Neste contexto, o conhecimento linguístico prévio constituía o único fator de bloqueio à aprendizagem, sendo que a ‘dificuldade’ em aprender um novo idioma se perspetivava em função da relação tipológica entre as línguas, postulando-se, assim, que, quanto mais diferenças estruturais existissem entre os dois idiomas, LM e LA, maiores dificuldades teria o aprendiz (Larsen-Freeman & Long 1991 [1994:57]). Por conseguinte, recorreu-se a uma análise contrastiva das estruturas dos sistemas envolvidos no processo de aquisição, já que “in the comparasion between native and foreign language lies the key to ease or difficulty in foreign language” (Lado 1957:1).

Gradualmente, a noção de transferência linguística foi sendo reconfigurada e ajustada aos novos paradigmas teóricos da ALNM. Com efeito, com o advento do pensamento cognitivista, sobretudo a partir dos anos 70 e 80, a investigação, implícita e explicitamente, redefiniu a noção de transferência linguística, procurando integrá-la no conjunto mais amplo de processos de natureza cognitiva que interagem em conjunto na construção das interlínguas dos aprendentes tardios, i.e., com outros fatores característicos do próprio desenvolvimento interlinguístico (Acosta & Leiria 1997; Alonso 2002; Gass & Selinker 2008). Assim, e segundo Gass (1996:321):

“[a]n important finding was that L1 influences occur not only as direct linguistic reflexes, but they also indirectly underlying organizational principles of language”.

---

<sup>127</sup> O autor sublinha ainda o facto de, na literatura especializada, o processo de preenchimento de lacunas através do uso de itens importados da LM ser muitas vezes tratado como uma característica típica do fenómeno de transferência linguística. Todavia, tal associação é, para Corder (1993:26), inadequada, dado que nestas situações “nothing is being transferred from anywhere to anywhere”.

Para além do conhecimento linguístico prévio proveniente da LM, Schachter (1993) integra sob o escopo da transferência linguística o conhecimento linguístico ‘imperfeito’ que o aprendiz tem da própria LA de aprendizagem. Nesta perspetiva, também o estágio atual do aprendiz relativamente ao conhecimento das estruturas da LA desempenha um papel relevante para a formulação de hipóteses sobre o *input* linguístico (1993:36).

Na literatura, observam-se ainda posições distintas relativamente à natureza da transferência linguística. Com efeito, alguns investigadores assumem que a transferência corresponde a um processo atuante ao longo do desenvolvimento interlinguístico dos aprendizes (Selinker 1972 [1992]; Corder 1967 [1992]). Entende-se, portanto, que a transferência é uma estratégia ao serviço do aprendiz, mobilizada para a resolução de problemas com os quais o falante se vê confrontado. É esta a posição que adota Kellerman (1977:93) ao afirmar que “[t]ransfer (...) is a strategy which the learner applies to the problem of how to express himself in the T[arget] L[anguage].”

Para Krashen (1981:67), a transferência é uma estratégia na medida em que resulta do facto de o falante se vir confrontado com situações comunicativas nas quais precisa de colocar em prática algo que ainda não aprendeu. A transferência linguística, assumida como ferramenta ao serviço do aprendiz, permite, assim, preencher lacunas do conhecimento do aprendiz relativamente às estruturas da LA, sendo certo que este ‘recuo’ (*fall back*) à LM se vai atenuando à medida que o falante aprendiz for progredindo no processo de aprendizagem da LA (Kellerman 1977; Krashen 1981). Assume-se, portanto, que a influência do conhecimento linguístico prévio será mais proeminente em estádios iniciais do processo de desenvolvimento linguístico do que nos estádios mais avançados (veja-se ainda a Secção 2.3.3.3.).

Embora considere que existem na literatura diferentes propostas terminológicas alternativas ao conceito de ‘transferência linguística’, associadas, conseqüentemente, a conceptualizações distintas do fenómeno, Odlin (1989) opta pela utilização deste termo e propõe, em alternativa, uma definição mais operativa. De acordo com a proposta deste autor (1989:27), a transferência linguística corresponde a:

“the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired”

Poder-se-á admitir que a descrição de Odlin é um pouco vaga, pelo facto de caracterizar a transferência como uma simples ‘influência’ linguística. Com efeito, pouco ainda se sabe sobre as motivações que determinam as ocorrências de tal influência, já que esta pode, entre outros factores, resultar das perceções várias que os aprendentes possuem acerca do potencial de transferibilidade das estruturas linguísticas (Kellerman 1977, 1995; Krashen 1981). Pese embora este facto, consideramos, no entanto, que a concetualização da transferência proposta por Odlin (1989) constitui, desde já, um bom ponto de partida para a sua análise, e em linha com esta proposta, neste estudo utilizaremos a expressão ‘transferência linguística’.

A transferência linguística tem implicações nos vários domínios estruturados dos sistemas linguísticos, podendo incidir sobre os planos fonético-fonológico, morfossintático, léxico-semântico e discursivo-pragmático (Odlin 1989; Ellis 1986, 2000; Ringbom 2013), sendo certo que, entre os diferentes planos estruturados, há, contudo, uns que são mais permeáveis à transferência do que outros, i.e., há determinadas estruturas da língua mais potencialmente transferíveis do que outras (Alonso 1999:142; Ellis 2000:316).

O plano fonético-fonológico é, talvez, um dos que maior permeabilidade apresenta (Ellis 1986:40; Ringbom 2013:397). A este respeito, é globalmente aceite a ideia de que quanto mais tardia for a exposição à LA de aprendizagem, maiores constrangimentos correlacionados com a influência do idioma nativo se observarão (veja-se, a este respeito, o trabalho de Johnson & Newport 1989; e as implicações associadas à postulação da hipótese do período ‘crítico’ consideradas na Secção 2.2.1. do presente trabalho).

A transferência linguística também acarreta implicações para o desenvolvimento da aquisição/aprendizagem do plano discursivo-pragmático de uma LNM, já que comportamentos não convergentes com a LA podem estar relacionados com a influência dos padrões culturais e discursivos da LM do falante aprendente (Ellis 2000:316; Ringbom 2013:397). Neste contexto, alguns trabalhos identificaram falhas de ajustamento às normas da LA para cumprimentar, responder a agradecimentos e negociar pedidos, causadas pela percepção social que os aprendentes possuem a partir da respetiva LM<sup>128</sup> (Kasper & Blum-Kulka 1993:10).

---

<sup>128</sup> Segundo Odlin (1989:37): “the use of apologies appears to be more frequent in American English than in Hebrew, and English speakers learning Hebrew appear to follow the norms of their native language in making apologies”.

O conhecimento linguístico prévio afeta também o plano léxico-semântico na medida em que pode influenciar o modo como os aprendentes concetualizam e utilizam as palavras da LA. Ringbom (2013:398) associa a transferência lexical à transferência concetual<sup>129</sup>, ao assumir que o conhecimento linguístico prévio produz efeitos, quer para o domínio da compreensão, quer para o domínio da produção das palavras de uma LNM. Segundo Pinto (2012:174), ao longo do desenvolvimento da aquisição/aprendizagem das palavras, observa-se que:

“[o]s alunos estabelecem relações de equivalência entre as palavras das línguas previamente adquiridas e as da língua-alvo e, com base nesta identificação interlinguística, transferem itens lexicais que consideram comuns nas línguas em comparação”.

Em relação ao que se regista nos outros planos estruturados da língua, o plano morfossintático é considerado um dos menos suscetíveis à transferência linguística (Ellis 2000; Ringbom 2013). Assumindo a dificuldade em determinar o grau de propensão para as ocorrências de transferência linguística sobre os diferentes domínios da estrutura de uma língua, Ellis (2000:317) observa, no entanto, que o nível sintático é talvez um dos menos afetados, justificando esta sua posição ao afirmar que:

“most learners have a much more highly developed metalingual awareness of grammatical properties than of phonological or discourse/pragmatic properties. This awareness may enable learners to control their choice of linguistic form at the level of grammar to a greater extent than at the other language levels and this may inhibit transfer.”

Portanto, Ellis (2000) associa uma menor influência linguística no domínio da morfossintaxe à mobilização do conhecimento metalinguístico, i.e., à reflexão, por parte dos aprendentes, sobre as características das estruturas em aquisição. Infere-se, então, que um maior foco na forma linguística conduzirá, em princípio, a uma menor incidência dos efeitos da transferência linguística.

---

<sup>129</sup> A transferência concetual corresponde a um fenómeno que opera ao nível da conceptualização da língua por parte dos aprendentes e não propriamente ao nível do uso e da produção de formas da LA. Assim, e no entender de Jarvis (2013:115): “conceptual transfer is assumed to originate from language-specific characteristics of a person’s conceptual system rather than from his or her linguistic (including semantic) knowledge per se.”

No que concerne ao género gramatical, vários trabalhos têm vindo a debruçar-se sobre o papel que a configuração do conhecimento linguístico prévio do aprendente poderá desempenhar para a assimilação desta estrutura gramatical e, mais especificamente, sobre o potencial de transferibilidade do género gramatical (White *et al.* 2004; Franceschina 2005; Sabourin *et al.* 2006; Foucart 2008). Estes trabalhos têm como principal objetivo compreender em que medida a presença na LM de sistemas de atribuição de género nominal com características muito próximas das da LA de aprendizagem afeta o desempenho linguístico dos aprendentes e se, em última instância, este é um fator determinante para a aquisição plena desta estrutura. Os resultados obtidos nestes trabalhos não são, contudo, inteiramente convergentes, já que, por um lado, se admite a possibilidade de os aprendentes adquirirem satisfatoriamente a estrutura de género gramatical, mesmo quando esta não está representada na sua LM, defendendo-se que a configuração do conhecimento linguístico prévio não impede o pleno acesso aos mecanismos cognitivos inatos para a aquisição da linguagem (White *et al.* 2004). Em contrapartida, outros estudos apontam para a impossibilidade da aquisição plena desta estrutura gramatical por aprendentes tardios cuja LM não possui esta categoria (Franceschina 2005; Sabourin *et al.* 2006; Foucart 2008; veja-se ainda a Secção 2.3.3.2.).

Ao analisar as implicações da transferência linguística para a assimilação do sistema de atribuição de género nominal do holandês por aprendentes tardios, que são falantes nativos de alemão, de inglês e de línguas românicas, Sabourin *et al.* (2006) verificaram que, pese embora se registe uma melhoria global no desempenho linguístico em níveis de proficiência mais avançados nos diferentes grupos de aprendentes analisados, os falantes cuja LM não possui um sistema de atribuição de género gramatical apresentam maiores dificuldades, quer em tarefas de associação do valor de género ao item lexical, quer em tarefas de estabelecimento de relações de concordância em género.

Relevante para a discussão dos dados apurados neste estudo é a distinção que se estabelece entre dois tipos de transferência linguística: a transferência de superfície (*surface transfer*) e a transferência profunda (*deep transfer*).

A transferência de superfície corresponde a “direct transfer of morphologically similar gender realization between L1 and L2” (Sabourin *et al.* 2006:10). Portanto, está intimamente correlacionada com o facto de o falante aprendente possuir na sua LM um sistema de atribuição de género nominal com características muito próximas do da LA de aprendizagem, nomeadamente no que concerne às propriedades formais subjacentes à atribuição dos valores de género aos nomes. Trata-se, por conseguinte, da possibilidade de

os aprendentes utilizarem, nas suas produções, a representação morfológica dos itens e dos respetivos valores de género que são, assim, diretamente importados da sua LM. Segundo os investigadores, a presença de cognatos entre as línguas potencia este tipo de transferência linguística e, dado o grau de semelhança entre as estruturas, auxilia o aprendente na tarefa de assimilação dos valores de género nominal da LA de aprendizagem.

Por sua vez, a transferência profunda diz respeito à transferência de estruturas sintáticas mais abstratas e manifesta-se quando não existe uma correspondência direta relativamente às marcas morfológicas associadas à realização dessas mesmas categorias gramaticais na LM e na LA de aprendizagem (Sabourin *et al.* 2006:3). Ou seja, a transferência profunda está implicada nas situações em que o falante aprendente possui na sua LM a categoria de género gramatical, mas com um sistema de atribuição de género nominal com propriedades formais de associação dos valores distintas do da LA de aprendizagem. No seu estudo, Sabourin *et al.* (2006) verificaram que este tipo de transferência também é relevante para a assimilação da categoria de género, já que o grupo de falantes de línguas românicas, pese embora a impossibilidade de transferirem as marcas morfológicas de atribuição dos valores de género da sua LM para a LA de aprendizagem, apresentam um desempenho globalmente superior ao que é registado pelo grupo de falantes cuja LM não possui um sistema de atribuição de género nominal, como é o caso dos falantes nativos de inglês. Estes últimos, por não terem representada na sua língua a categoria de género gramatical, apresentam piores resultados relativamente aos restantes grupos, quer em tarefas de associação de valor de género aos itens lexicais, quer em tarefas de concordância em género (Sabourin *et al.* 2006:23). Portanto, a partir da análise dos dados apurados, os autores concluem que (2006:24):

“the relatively good performance of the German group compared to the Romance group suggests that while transfer of abstract features plays a role in the L2 acquisition of gender, direct surface transfer of a congruent system is even more helpful, at least for acquiring gender assignment. Surface transfer, thus, probably accounted for some part of the advantage”

Porém, e como os próprios autores sublinham, não é possível assumir uma posição clara relativa ao papel desempenhado pela transferência (profunda ou de superfície) para a aquisição/aprendizagem dos sistemas de género gramatical, já que a assimilação desta estrutura pode depender da atuação de outros fatores que não estão intrinsecamente

relacionados com a configuração do conhecimento linguístico prévio (cf. Franceschina 2005; Sabourin *et al.* 2006).

### 2.3.2. Manifestações associadas à transferência linguística

Na literatura sobre transferência linguística, distinguem-se as manifestações associadas à: (i) transferência positiva; e (ii) transferência negativa (ou interferência)<sup>130</sup> (cf. Figura 2.3).

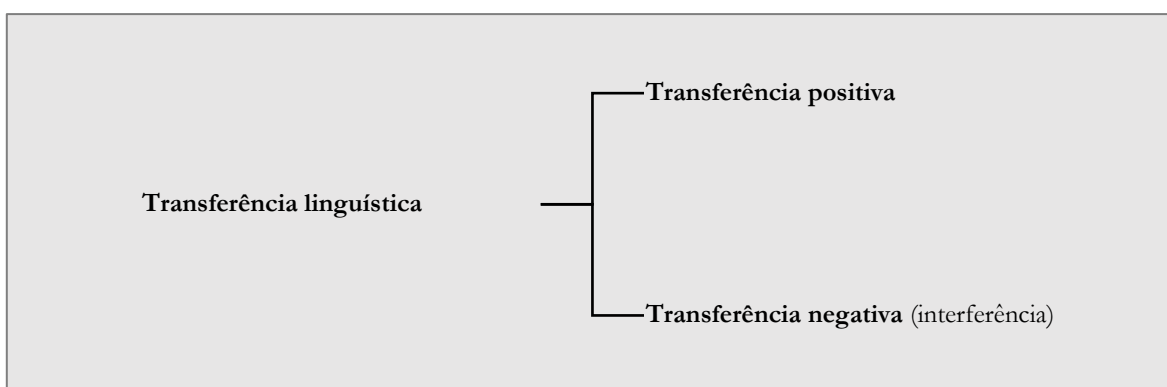


Figura 2.3 – Distinção dos efeitos positivos e negativos da transferência linguística

De um modo geral, entende-se que as similitudes estruturais entre a LM e a LNM facilitam a aprendizagem, porque os aprendentes podem transferir adequadamente, por transferência positiva, para a LNM, as estruturas linguísticas e os padrões culturais da sua LM (Lado 1957; Larsen-Freeman & Long 1991 [1994:56]; Ellis 2000:300; Suisse 2011:172). Em contrapartida, os conhecimentos linguísticos podem interferir (transferência negativa ou interferência) no processo de assimilação das estruturas de uma LNM conduzindo, por fim, os aprendentes ao desvio e/ou a outras manifestações não convergentes com a LA de aprendizagem (Ellis 2000:300).

#### 2.3.2.1. Transferência positiva

É geralmente aceite a ideia de que os conhecimentos linguísticos prévios dos aprendentes podem constituir uma ‘vantagem’ para o aprendente tardio durante o processo de assimilação dos diferentes planos estruturados da LNM (Odlin 1989; Schachter 1996;

---

<sup>130</sup> Assinale-se, porém, que as noções de ‘transferência positiva’, por um lado, e de ‘transferência negativa’, por outro, não implicam necessariamente dois processos cognitivos distintos, traduzindo apenas os efeitos (positivos e negativos) que a transferência linguística pode causar ao longo do processo de aquisição/aprendizagem de uma LNM (Gass & Selinker 2008:94).

Ellis 2000; Sabourin *et al.* 2006). Tal ‘vantagem’ está tipicamente associada ao grau de proximidade estrutural dos idiomas, i.e., da LM do falante aprendente e da LA de aprendizagem (Schachter 1996:161; Ellis 2000:304; Sabourin *et al.* 2006:23; cf. Secção 2.3.3.1.). Porém, nem sempre o facto de as línguas serem estruturalmente próximas implica, necessariamente, uma maior rapidez no processo de desenvolvimento da assimilação dos diferentes planos estruturados da LNM. Na verdade, uma proximidade relativa dos idiomas pode conduzir a uma maior incidência de desvios devidos a transferência linguística, já que nestes casos os aprendentes “are prepared to borrow but are less likely to be successful in avoiding error.” (Corder 1993:27).

Num estudo sobre o processo de aquisição da estrutura de negação por falantes nativos de espanhol, aprendentes de inglês como LNM, Zobl (1980:477) constatou que uma maior proximidade estrutural dos idiomas contribui para uma maior persistência dos desvios que, conseqüentemente, conduzem a um bloqueio do desenvolvimento do processo aquisitivo. Assim sendo, o autor admite que, em última instância, a transferência linguística pode conduzir o aprendente a um estágio de estabilização, i.e., a um progressivo bloqueio na aquisição/aprendizagem das estruturas de uma LNM<sup>131</sup> (Zobl 1980:477). Também Selinker & Lakshmanan (1993:197-216) consideram a transferência linguística como um dos fatores que contribui para a ocorrência de fenómenos linguísticos fossilizados (cf. Secção 2.2.2.).

No caso específico da assimilação do sistema de atribuição de valores de género nominal do português como LNM, poder-se-á supor que o facto de os aprendentes possuírem, na sua LM, um sistema de atribuição de género nominal com propriedades semelhantes às do português constitui, à partida, uma condição favorável para a assimilação desta estrutura gramatical, sobretudo em fases iniciais da aprendizagem (cf. Mariotto & Lourenço-Gomes 2013; Martins 2015). Em contrapartida, a presença de semelhanças pode ter um efeito inverso. Ou seja, à medida que o aprendente constata que o sistema de atribuição de género nominal da LA de aprendizagem apresenta características muito próximas das propriedades do sistema de atribuição de género nominal da sua LM, quer no que concerne ao número de valores de género que possui, quer relativamente às propriedades formais de atribuição dos valores de género aos nomes, pode provavelmente assumir que, na maior parte dos itens lexicais que vai adquirindo, os valores de género são

---

<sup>131</sup> Também Alonso (2002:235) observa que a transferência “acts together with other [inter]L[anguage] processes, such as fossilization, determining the degree of fossilized structures”.



correspondentes aos valores associados aos respetivos itens da sua LM, como consequência de uma transferência de superfície (Sabourin *et al.* 2006). Tais hipóteses serão consideradas na análise dos dados empíricos recolhidos para esta investigação.

### 2.3.2.2. Transferência negativa (ou interferência)

A transferência linguística pode conduzir o aprendente ao desvio, na medida em que interfere no desenvolvimento interlinguístico. P. Corder (1967 [1992]) considera que, e ao contrário do que é postulado pela Análise Contrastiva, os comportamentos desviantes não resultam unicamente da transferência linguística, e é sob este pressuposto que se baseia o trabalho de investigação desenvolvido pela Análise de Erros (Gass & Selinker 2008:102).

Visando, sobretudo, objetivos pedagógicos, a investigação conduzida sob o modelo da Análise de Erros permitiu a identificação de variados tipos de desvios, distinguindo-se os de natureza interlinguística, atribuíveis à influência dos conhecimentos linguísticos prévios, dos de natureza intralinguística, característicos dos próprios padrões de desenvolvimento da aquisição linguística e não dependentes, então, do sistema linguístico nativo ou de outras línguas previamente conhecidas pelo aprendente. Por conseguinte, estes tipos de desvio são expectáveis em diferentes grupos de aprendentes (Larsen-Freeman & Long 1991 [1994:61]; Gass & Selinker 2008:103).

Porém, é difícil determinar, de modo preciso, a causa das ocorrências desviantes, i.e., se estas se devem a razões de natureza interlinguística ou intralinguística<sup>132</sup>. Apesar de reconhecer esta dificuldade, Odlin (1989:37) considera que há comportamentos desviantes observáveis, sobretudo, no plano lexical e discursivo, nos quais é possível inferir a influência dos conhecimentos linguísticos prévios. Mais especificamente, observa que a transferência linguística pode estar associada a fenómenos desviantes de (i) substituição; (ii) decalque; e (iii) hipercorreção que, em certa medida, se correlacionam. No entendimento do autor, a substituição corresponde ao uso, por parte do aprendente, de itens lexicais ou de expressões provenientes da LM ou de outras línguas. Por sua vez, os desvios associados ao decalque refletem a interseção da estrutura do sistema linguístico nativo e verificam-se, por exemplo, quando, ao construir uma frase na LA de aprendizagem, o aprendente segue uma ordem de palavras não correspondente à da LA. Para Odlin (1989),

---

<sup>132</sup> Ellis (2000:62) ilustra este problema com o seguinte exemplo: “the ‘no’+verb error (as in \*No look my card) is universal, suggesting intralingual explanation, but Spanish learners of L2 English have been noted to make this error more frequently and for a longer period of time, suggesting that L1 pattern for negatives (*no* + verb) is also having influence”.

a hipercorreção pode também estar correlacionada com os efeitos negativos da transferência linguística, quando envolve a alteração da estrutura da LA em função da influência do conhecimento linguístico prévio. Para ilustrar este fenómeno, o investigador (1989:38) apresenta o seguinte exemplo:

“Arabic speakers occasionally make E[nglish] S[econd] L[anguage] spelling errors that involve substitutions of the letter b for the letter p, as in blaying (...); however, Arabic speakers also use p inappropriately (e.g., in hapit) in mistaken attempts to avoid b/p substitutions”.

Porém, como já referido, nem sempre é fácil averiguar se determinada ocorrência desviante na produção dos falantes resulta simplesmente da transferência linguística, na medida em que o desvio pode estar associado a outros fatores característicos da própria construção das interlínguas dos aprendentes (Zobl 1980; Ellis 1986; Selinker 1972 [1992]; Corder 1967 [1992]; Gass & Selinker 2008; veja-se ainda a Secção 2.2.2. do presente capítulo). A este respeito, observa Ellis (2000:334) que a transferência linguística interage com os princípios naturais do desenvolvimento da aquisição linguística, sendo que a investigação ainda precisa de apurar as condições promotoras de tal interação.

Portanto, a transferência linguística negativa pode ser direta, já que, e como se viu, poderá estar na base de diferentes tipos de comportamentos desviantes. Para além disso, a transferência linguística negativa pode ter manifestações indiretas, atuando, por exemplo, como um fator promotor de fenómenos como a inibição e a sobregeneralização, que passaremos a tratar em seguida.

Com efeito, a inibição (*avoidance*) surge tipicamente associada à transferência linguística (Odlin 1989; Ellis 2000; Gass & Selinker 2008). Em linhas gerais, a inibição caracteriza-se pelo evitamento do uso de determinadas estruturas da LA por parte do aprendente e pode estar associada à influência interlinguística na medida em que o aprendente evita o uso de determinadas estruturas gramaticais que, por não serem convergentes com as da sua LM, considera difíceis.

Ao analisar as frequências de uso das orações relativas em inglês num conjunto de 50 produções escritas por grupos de aprendentes que são falantes nativos de chinês, japonês, árabe e persa, Schachter (1974 [1992]) constatou que havia uma menor incidência de desvios nos aprendentes de LM chinesa e japonesa do que nos restantes grupos. A autora

verificou que estes resultados se compreendem pelo facto de os falantes nativos de chinês e de japonês evitarem o uso deste tipo de construções sintáticas<sup>133</sup> nos seus textos, uma vez que as respetivas LM apresentam estruturas relativas diferentes das existentes no inglês (Schachter 1974 [1992:201-202]).

Reconhecendo que a determinação dos nomes constitui uma área crítica no contexto da aquisição/aprendizagem tardia de PLNM, Lopes & Martins (2017) analisaram um conjunto de dados empíricos recolhidos em textos escritos por aprendentes tardios de PLNM que são falantes nativos de chinês e de espanhol. Neste estudo, as autoras verificaram que a evitação do uso de artigos pode estar correlacionada com a influência do conhecimento linguístico prévio proveniente da LM. Assim, Lopes & Martins (2017:173) consideram que as omissões:

“podem denunciar uma estratégia de fuga dos aprendentes que, confrontados com a complexidade dos padrões de uso dos artigos no PEC inferidos a partir do *input*, optam por não arriscar (...), mas serão também, no caso de grupos específicos de aprendentes, resultantes de um efeito da LM (...).”

Em contrapartida, a inibição poderá não depender unicamente da transferência, tal como Ellis (2000:305) afirma:

“[t]he extent of learner’s knowledge of the L2 and the attitudes learners hold toward their own and target-language act as factors that interact with L1 knowledge to determine avoidance behaviour”.

A transferência linguística poderá ser igualmente promotora de uma utilização recorrente do material linguístico da LNM (*overuse / overproduction*) (Odlin 1989:37; Ellis 2000:305; Gass & Selinker 2008:143). O fenómeno de *overuse* é visível quando, em diferentes contextos, o aprendente opta por recorrer sistematicamente às mesmas estruturas linguísticas. Segundo Odlin (1989:37), estas ocorrências sistemáticas podem estar correlacionadas com a própria atitude do falante que, para evitar a utilização de uma

---

<sup>133</sup> Schachter (1974 [1992]) admite ainda a possibilidade de a inibição ser mais frequente ao nível sintático do que fonológico, por exemplo. Segundo a investigadora, os alunos conseguem mais facilmente evitar o uso de determinadas construções sintáticas consideradas difíceis uma vez que podem parafraseá-las. Em contrapartida, já no plano fonológico e embora o aprendente possa, em determinados contextos, evitar o uso de palavras que considere mais difíceis de pronunciar, o que se verifica “es que el alumno se ve obligado a utilizar las palabras y, por lo tanto, a cometer errores al pronunciar los sonidos” (1974 [1992:204]).

determinada estrutura ou de uma forma linguística que considere mais difícil, recorre sistematicamente a estruturas com as quais sente mais ‘confiança’.

A transferência linguística apresenta, como vimos, impactos positivos e negativos para o desenvolvimento da aquisição não nativa com implicações várias no próprio ritmo e percurso aquisitivo (Odlin 1989; Gass 1996; Gass & Selinker 2008; Madeira 2017). Referidas as principais manifestações que, na literatura, são associadas à transferência linguística, consideramos em seguida os fatores promotores e inibidores da transferência, com especial enfoque para as condições de transferibilidade do género gramatical.

### **2.3.3. Condições promotoras de transferibilidade do género gramatical**

Uma parte substancial da investigação sobre a aquisição/aprendizagem dos diferentes planos estruturados de uma LNM procura compreender não só *como*, mas também *quando* é que a transferência linguística se manifesta ao longo do desenvolvimento interlinguístico dos aprendentes tardios (Gass 1979:328; Gass & Selinker 2008:137; Madeira 2017:314-315). Nas secções anteriores do presente capítulo, considerámos os aspetos relevantes para a definição de transferência linguística, tendo-se também dado conta do modo como este fenómeno interage com o desenvolvimento do processo aquisitivo dos diferentes planos estruturados da LA de aprendizagem à qual o falante aprendente está exposto.

A transferência linguística pode depender de variáveis relacionadas, quer com as propriedades da LA de aprendizagem, quer com o estágio de desenvolvimento da aquisição/aprendizagem (Odlin 1989; Ellis 2000, 2003; Murphy 2003). No que concerne aos fatores linguísticos, consideremos para a nossa análise as questões que se prendem com a proximidade tipológica dos idiomas. Para além disso, e atendendo aos objetivos estipulados para este trabalho, convocar-se-ão ainda um conjunto de hipóteses teóricas postuladas no âmbito do modelo generativo e da teoria da Gramática Universal para explicitar o potencial de transferibilidade dos diferentes planos linguísticos, com especial enfoque para as hipóteses relativas à configuração do ‘estádio inicial’ do aprendente tardio.

Para além disso, as condições subjacentes à transferibilidade das estruturas podem estar correlacionadas com o estágio de desenvolvimento do processo da aquisição/aprendizagem da LNM. A este respeito, ter-se-á em linha de conta as implicações

relativas ao nível de proficiência e de competência linguísticas do aprendente, já que esta será uma variável crucial para a análise dos dados empíricos recolhidos para este estudo.

### 2.3.3.1. Proximidade tipológica

Ao definir a transferência linguística, Ringbom (2013:396) observa que a proximidade estrutural dos idiomas, i.e., LM e LA de aprendizagem, é um fator determinante para as ocorrências de transferência linguística. Na mesma linha, Schachter (1996:161) considera que a aquisição plena dos planos estruturados de uma língua depende do grau de proximidade estrutural da LM do aprendente e da LA de aprendizagem, na medida em que: “[t]he closer two languages are in terms of syntax, phonology, and lexicon, the more likely it is that higher levels of completeness can be reached.”

De um modo semelhante, mas associando a idade à transferência linguística, Bialystok & Miller (1999), num estudo já citado neste capítulo, consideram que a presença de divergências estruturais entre a LM e a LA de aprendizagem tem impactos no processo de aquisição/aprendizagem por falantes aprendentes precoces e falantes aprendentes tardios, sobretudo quando a LM não possui uma categoria morfossintática representada na LNM. Para estes autores (1999:130):

“[y]ounger learners should readily construct the new L2 category, whereas older learners should have limited success at best. Older learners have neither access to language acquisition mechanisms nor knowledge of the structure, so these categories are presumably difficult to learn. For categories that correspond across the two languages, older learners could profit from the similarity while younger learners may use either analogy with the similar L2 construction, or guidance from language acquisition mechanisms, or both, to master the category. For similar categories, therefore, the gap between older and younger learners would be narrowed, but for reasons that are not necessarily clear.”

Já Kellerman (1977:80) assume que a transferência linguística não depende só da relação tipológica entre as línguas, mas, sobretudo, da distância estrutural que é ‘percebida’ pelos aprendentes e é com base nesse pressuposto que define o conceito de ‘psicotipologia’. A psicotipologia corresponde a um conjunto de percepções várias que os aprendentes possuem da distância estrutural das línguas e que ao longo do desenvolvimento interlinguístico pode promover, por um lado, ou inibir, por outro, a transferência

linguística<sup>134</sup> (Alonso 1999:144; Ellis 2000:327). Assim, e no entendimento de Kellerman (1977:93): “if the learner believes that there could be a relationship between N[ative] L[anguage] and T[arget] L[anguage] at a given point, he may well transfer.” Do mesmo modo, Krashen (1981) advoga que a transferência linguística resulta da representação mental que o aprendente tem das regras da LA e que considera mais próximas das da sua LM (1981:69, n.1.).

As questões relacionadas com a proximidade tipológica têm sido especialmente relevantes para os estudos sobre a aquisição/aprendizagem de uma LNM por falantes multilingues. Neste contexto, sugere-se que os aprendentes não transferem, necessariamente, as propriedades linguísticas da sua LM, mas de outras línguas previamente conhecidas que sejam tipologicamente mais próximas das da LA de aprendizagem (cf. DeAngelis 2005; Pinto 2012).

### **2.3.3.2. Configuração da gramática da interlíngua**

Uma parte substancial da investigação sobre a aquisição/aprendizagem da categoria de género gramatical numa LNM por aprendentes tardios toma como ponto de referência os pressupostos do modelo generativo e do pressuposto de que existe uma GU. Neste modelo teórico, são formuladas hipóteses várias relacionadas com a transferência linguística e com a sua relevância para a configuração e desenvolvimento da gramática da interlíngua. Consideram-se em seguida três dessas hipóteses que levantam questões distintas acerca do ‘estádio inicial’ do falante aprendente atendendo, igualmente, ao grau de acessibilidade (direto/indireto e total/parcial) à GU e à respetiva configuração da gramática do idioma nativo do falante, com consequências para o desenvolvimento da aquisição das categorias gramaticais na LNM.

#### **2.3.3.2.1. Full Transfer / Full Access Hypothesis**

Schwartz & Sprouse (1996) assumem que o ‘estádio inicial’ do aprendente tardio corresponde à representação total da gramática da sua LM e é com base neste pressuposto

---

<sup>134</sup> Também Ellis (2003:54) considera que a perceção dos falantes aprendentes relativamente à estrutura em aquisição é determinante para a transferência linguística, ao afirmar que: “transfer is governed by learner’s perceptions about what is transferable and by their stage of development.”

que formulam a hipótese *Full Transfer / Full Access* (FTFA)<sup>135</sup> (Schwartz & Sprouse 1996:41), segundo a qual:

“the starting point of L2 acquisition is quite distinct from that of L1 acquisition: in particular, it contends that all the principles and parameter values as instantiated in the L1 grammar immediately carry over as the initial state of a new grammatical system on first exposure to input from the target language (TL). This initial stage of the L2 system will have to change (...); that is, failure to assign a representation to input data will force some sort of restructuring of the system (‘grammar’), this restructuring drawing from options of U[niversal] G[rammar] (...). In some cases, this restructuring may occur quite rapidly; in others, much more time may be needed”

Assim, nesta hipótese teórica, é assumido que todas as propriedades da gramática do sistema linguístico nativo são potencialmente transferíveis para a LA<sup>136</sup> (*Full Transfer*). Todavia, e à medida que os aprendentes se vão deparando com estruturas não convergentes com as da sua LM, podem igualmente aceder, em simultâneo, à GU (*Full Access*), que continua disponível ao longo do desenvolvimento interlinguístico dos aprendentes (Schwartz & Sprouse 1996; White 2003:61). Consequentemente, a interação entre os conhecimentos da LM e a GU que caracteriza o estágio inicial do aprendente tardio verifica-se igualmente ao longo das diferentes etapas do desenvolvimento do processo de assimilação das estruturas de uma LNM.

Os resultados obtidos num estudo conduzido por White *et al.* (2004) sobre a aquisição/aprendizagem das categorias de género e de número por parte de aprendentes tardios de espanhol, falantes nativos de inglês e de francês, revelaram uma influência residual do conhecimento linguístico prévio, sobretudo quando se trata de aprendentes proficientes. White *et al.* (2004) assumem, então, os princípios subjacentes à hipótese da

---

<sup>135</sup> Originalmente, a hipótese FTFA teve a denominação de *Absolute L1 Influence Hypothesis*. Porém, Schwartz & Sprouse (1996) constataram que a denominação original gerou uma interpretação errada, transmitindo a ideia de que a influência linguística determina o desenvolvimento das interlínguas dos aprendentes. Na verdade, o que Schwartz & Sprouse (1996) defendem é que as propriedades gramaticais presentes na LM dos aprendentes são, em princípio, analisadas ao longo do desenvolvimento das interlínguas, sendo certo que cada estágio de desenvolvimento das interlínguas não deixa de ser compatível com os estádios de desenvolvimento das gramáticas permitidas pela GU (Schwartz & Sprouse 1996:40, n.1)

<sup>136</sup> A este respeito, afirma White (2008:47) que a transferência linguística “may be persistent or not, depending on particular linguistic properties and particular language combinations”.

*Full Transfer / Full Access*, ao observarem que “postpuberty learners are able to acquire gender agreement regardless the status of gender features in L1” (White *et al.* 2004:127).

Em contrapartida, para Franceschina (2005:61) um dos principais problemas da hipótese FTFA é o de ter assumido que, quer na aquisição/aprendizagem tardia de uma LNM, quer na aquisição de uma LM estão envolvidos os mesmos processos cognitivos (Schwartz & Sprouse 1996:42), ou seja, não se tem em conta os condicionamentos associados à possibilidade de haver uma ‘idade crítica’ para a assimilação das estruturas de uma língua não nativa. Ora, para a investigadora, a comparação entre o desempenho linguístico de falantes aprendentes precoces e falantes aprendentes tardios, sobretudo no que concerne à aquisição de propriedades gramaticais parametrizadas, i.e., instanciadas de forma distinta nas línguas, como o género gramatical, revela que as previsões do modelo da FTFA não se verificam, havendo, de facto, diferenças relativamente aos processos cognitivos envolvidos na assimilação das estruturas por falantes nativos, por um lado, e por aprendentes tardios, por outro. Assim, constrangimentos de natureza maturacional associados ao período cronológico em que se inicia a exposição à LA, verificáveis ao longo dos diferentes estádios de desenvolvimento interlinguístico dos aprendentes tardios, impedem, necessariamente, o pleno acesso aos processos cognitivos atuantes na aquisição linguística nativa (veja-se, a este respeito as considerações acerca da *Interpretability Hypothesis* descritas na Secção 2.2.1.).

Também Håkansson (2001), através da análise de um conjunto de dados empíricos recolhidos de aprendentes precoces (crianças e jovens adolescentes), que são falantes nativos de sueco, em diferentes estágios de aprendizagem de alemão como LNM, constatou que determinadas estruturas gramaticais, nomeadamente ao nível da sintaxe, não são transferíveis da LM dos aprendentes, havendo evidências de que os grupos de falantes em estudo constroem uma nova gramática a partir do léxico da LA. A investigadora assume, em alternativa, que há diferentes estágios de processamento da gramática da LA de aprendizagem, à luz do que é postulado no âmbito da ‘Teoria da Processabilidade’ (*Processability Theory*) de Pienemann (2008) (cf. Secção 2.2.2.2.).

Os resultados deste estudo revelam, ainda, um grau de variação assinalável entre os diferentes grupos de aprendentes estudado. Para a investigadora, o facto de a variação se registar em aprendentes com a mesma LM coloca, necessariamente, problemas relativamente à assunção de que, ao longo do desenvolvimento da aquisição/aprendizagem de uma LNM, há sempre ocorrências de transferência do sistema linguístico nativo (Håkansson 2001:80). Assim, segundo Håkansson (2001) a variação na interlíngua



justifica-se pelo facto de o sistema interlinguístico dos aprendentes se encontrar em diferentes estádios de desenvolvimento e não em função de propriedades potencialmente transferidas da LM.

#### 2.3.3.2.2. Minimal Trees Hypothesis

Vainikka & Young-Scholten (1994) assumem que o ‘estádio inicial’ do aprendente tardio de uma LNM corresponde a uma gramática particular com representações baseadas na gramática da LM do falante aprendente e é com base neste pressuposto que formulam a *Minimal Trees Hypothesis* (MTH) (Vainikka & Young-Scholten 1994).

Ao contrário do que é postulado pela hipótese FTFA, Vainikka & Young-Scholten consideram que, na aquisição/aprendizagem de uma LNM, o estágio inicial do aprendente tardio “consists solely of this transferred lexical projection” (1994:293), ou seja, no início da aprendizagem, só estão disponíveis as configurações relativas ao léxico do idioma nativo, não havendo, em contrapartida, acesso às categorias gramaticais/funcionais desse mesmo idioma. Assim “while functional categories are available in the U[niversal] G[rammar] inventory, initial grammars lack the full complement of functional categories, containing lexical categories and their projections” (White 2003:69).

Baseando-se em dados empíricos, as investigadoras defendem ainda que os processos de aquisição linguística nativa e não nativa são, em certa medida, similares, uma vez que os dois se iniciam a partir de projeções de natureza lexical e de as propriedades gramaticais/funcionais emergirem em fases posteriores do desenvolvimento da aquisição a partir dos dados do *input* linguístico (Vainikka & Young-Scholten 1994:266). Portanto, se as categorias lexicais e as suas propriedades configuram, então, o ‘estádio inicial’ do aprendente tardio, prevê-se que o estágio inicial de aprendentes com diferentes LM seja, necessariamente, diferente (White 2003:69).

No que diz respeito à configuração das propriedades gramaticais, tais como o género gramatical, Vainikka & Young-Scholten (1994) assumem que estas emergem a partir dos dados do *input* da LNM e não da LM, sendo este, talvez, um dos pontos mais problemáticos desta hipótese teórica. Ou seja, infere-se, a partir desta hipótese, que, ao longo do desenvolvimento linguístico dos falantes aprendentes, não haverá qualquer influência do conhecimento linguístico prévio. Ao analisar este pressuposto teórico, observa White (2003:69) que:

“[e]mergence of functional categories (...) in no way depends on properties of the L1 grammar; in other words, there is predicted to be no transfer in this domain, no stage or grammar in which properties of the mother-tongue functional categories are found, an assumption which differs from Full Transfer Full Access. Rather, the L2 learner acquires L2 functional categories, with L2 properties. Thus, L1 and L2 acquisition of any particular language are generally assumed to be identical with respect to functional categories and projections.”

Todavia, esta hipótese teórica apresenta problemas de validação empírica, já que, em boa verdade, estudos vários com dados sobre a aquisição não nativa revelam a presença da influência da língua nativa na configuração das propriedades gramaticais da LNM (cf. Schwartz & Sprouse 1996; veja-se ainda White 2003).

#### **2.3.3.2.3. Failed Functional Features Hypothesis**

Os proponentes da *Failed Functional Features Hypothesis* (FFFH) (Hawkins & Chan 1997) assumem que, apesar de no início da aquisição/aprendizagem de uma LNM os princípios da GU estarem disponíveis, quanto mais tardia é a exposição à LA, maiores dificuldades se observam na reconfiguração dos parâmetros da GU (Hawkins & Chan 1997:216). Esta hipótese teórica está intimamente associada à *Interpretability Hypothesis* (Tsimplici 2003; Tsimplici & Dimitrakopoulou 2007) na qual, e como vimos, se assume a impossibilidade da aquisição tardia de categorias gramaticais consideradas não-interpretáveis (cf. Secção 2.2.1).

A FFFH parte do pressuposto de que a GU dispõe de um conjunto de categorias funcionais, i.e., gramaticais, e que cada língua seleciona as categorias que irão integrar o seu inventário. Por conseguinte, as categorias funcionais podem ser classificadas como ‘universais’, existentes em todas as línguas, e como ‘parametrizadas’, i.e., selecionadas por algumas línguas (Franceschina 2003). Esta distinção é, então, fundamental, já que, na FFFH, se prevê que as categorias funcionais universais são adquiridas por aprendentes tardios de uma LNM, ao passo que as categorias funcionais parametrizadas, como o género gramatical, só são adquiridas quando se encontram representadas na LM do aprendente (Hawkins & Chan 1997; Franceschina 2003, 2005). Assim, e nas palavras de Franceschina (2003:97):

“universal features are predicted to be acquirable by all adult L2 learners in principle; by contrast, parameterized features are predicted to be fully acquirable only in cases where they were present in the learner's L1 functional feature inventory.”

Portanto, prevê-se que os princípios da GU constituem o estágio inicial dos aprendentes, não sendo possível, no entanto, e no decurso da aquisição linguística tardia, a reconfiguração dos parâmetros da GU.

Os fundamentos propostos no âmbito da FFFH têm sido particularmente relevantes em estudos empíricos sobre o processo de assimilação dos sistemas de atribuição dos valores de género em contexto de aquisição/aprendizagem de uma LNM (veja-se, por exemplo, para o espanhol como LNM, Franceschina 2003, 2005; para o holandês como LNM, Sabourin *et al.* 2006, e para o português como LNM, Lacsán 2015). Franceschina analisou um conjunto de dados recolhidos, em diferentes tipos de tarefas, de aprendentes de espanhol como LNM, falantes nativos de árabe, inglês, francês, alemão, grego e português (2005:121) em diferentes níveis de competência e proficiência linguísticas. A seleção dos grupos de informantes dependeu, assim, da presença ou ausência de um sistema de atribuição de género na respetiva LM. A partir da análise dos dados, Franceschina (2005) constatou que há um melhor desempenho global por parte dos aprendentes de LM com sistema de atribuição de género relativamente ao grupo de aprendentes cuja LM não possui um sistema de género. Por conseguinte, a investigadora afirma que “the functional feature makeup of the L1 has a crucial impact on L2 attainment” (Franceschina 2005:192), ou seja, os seus dados confirmam as previsões da FFFH, segundo as quais categorias que não estão instanciadas na LM não são plenamente adquiridas. Por fim, a autora (2005:196) conclui que:

“L1 exerts its influence in shaping the L2 end state not only in areas of grammar where the input available may be ambiguous, but also where there is unambiguous positive evidence in the L2 for the presence of particular features.”

Portanto, e de acordo com esta conclusão, infere-se que a GU não é suficiente para cumprir a tarefa de aquisição destas estruturas linguísticas. Porém, pese embora o facto de os desvios serem persistentes e sistemáticos nos falantes cuja LM não possui um sistema de género, Franceschina (2005:198) observa que, em termos globais, estes aprendentes apresentam um desempenho linguístico bom. A autora considera que tal performance poderá justificar-se pelo recurso destes aprendentes a estratégias compensatórias que lhes permitem uma maior aproximação às estruturas da LA de aprendizagem, apesar de não ser fácil a discriminação do tipo de estratégias mobilizadas pelos aprendentes tardios durante a assimilação das estruturas linguísticas. Infere-se, portanto, que fatores característicos do próprio desenvolvimento interlinguístico terão também um papel preponderante para

aquisição/aprendizagem da categoria de género gramatical. Para a discussão dos dados, a autora convoca também o papel que a exposição ao *input* poderá ter para a tarefa de assimilação da categoria de género, embora esta relação não seja direta nem fácil de determinar (2005:197-198):

“I shall assume there must be some minimum requirement on the quantity of input that a learner must be exposed to and also that the learner needs to be at right developmental stage in order to be able to acquire a given property, but I shall remain open about the exact details”

### **2.3.3.3. Nível de proficiência e de competência linguística**

Nos estudos sobre a transferência linguística, é geralmente aceite a ideia de que a transferência de determinados planos da estrutura linguística é mais proeminente nos níveis de proficiência linguística elementares do que nos níveis mais avançados (Odlin 1989; Ellis 1986:40, 2000:330; Murphy 2003:7). A prevalência dos efeitos da transferência nos níveis elementares está, segundo Odlin (1989:133), associada à presença de desvios que, nestes níveis, são tipicamente de origem interlinguística. A correlação dos níveis de proficiência mais baixos com maiores manifestações da transferência é suportada pela posição dos investigadores que assumem a transferência como uma estratégia adotada pelo aprendente a fim de colmatar as falhas relativas ao conhecimento da gramática da LNM (cf. Kellerman 1977; Krashen 1981).

Em contrapartida, se se tiver em conta os efeitos positivos da transferência, associados à promoção do desenvolvimento do processo aquisitivo, é possível detetar manifestações da transferência linguística mesmo em níveis de proficiência linguística mais avançados (Odlin 1989:133).

A correlação dos níveis de proficiência e de competência linguísticas com a transferência linguística não é evidente, já que o comportamento linguístico dos aprendentes em diferentes estádios do desenvolvimento pode não estar diretamente associado à influência interlinguística. Com efeito, dados empíricos recolhidos em estudos sobre a aquisição/aprendizagem da categoria de género em PLNM por aprendentes com diferentes perfis e em diferentes estádios de desenvolvimento (Cf. Capítulo 1, Secção 1.5.2.) revelam que, apesar da ocorrência de desvios relativos a esta estrutura em níveis de proficiência mais avançados, há uma melhoria no desempenho linguístico dos aprendentes à medida que vão avançando no processo aquisitivo. Porém,

é necessário averiguar se é possível estabelecer uma relação entre os padrões de desenvolvimento das interlínguas e a influência do conhecimento linguístico prévio e que tipo de manifestações sustentam essa relação. Neste estudo, procurar-se-á apurar, a partir da análise dos desvios de atribuição de valores de género aos nomes e de concordância nominal em género, em que medida a interação entre o conhecimento linguístico prévio e os padrões de desenvolvimento interlinguístico interagem na aquisição/aprendizagem desta estrutura gramatical no PLNM.

## 2.4. Síntese

Ao longo deste capítulo, consideraram-se os contributos das diferentes teorias que atuam no âmbito da ALNM para a análise do processo de aquisição/aprendizagem tardia de uma LNM. Para além disso, e atendendo aos objetivos estipulados para este trabalho de investigação, procedeu-se a uma análise cuidada das principais questões relativas ao papel da configuração do conhecimento linguístico prévio para o desenvolvimento da gramática interlinguística, com especial enfoque para o papel da transferência linguística na assimilação das propriedades do sistema de atribuição de valores de género aos nomes em português por aprendentes tardios. Assim, é em função das hipóteses teóricas aqui consideradas, no que respeita ao grau de transferibilidade da categoria de género gramatical, aos pressupostos relativos ao estágio inicial do aprendente e à conseqüente influência do conhecimento linguístico prévio, bem como às eventuais estratégias compensatórias adotadas pelos aprendentes ao longo das diferentes fases do desenvolvimento interlinguístico, que se procederá ao tratamento empírico dos dados.

O próximo capítulo é, então, dedicado às questões metodológicas adotadas no estudo empírico e no capítulo 4 apresentam-se os resultados apurados.









# Capítulo 3

## Metodologia

---

### 3.1. Introdução

No capítulo que agora se inicia, apresenta-se a metodologia adotada no estudo empírico. Assim, na parte introdutória (3.1.1.), apresentam-se as perguntas da investigação, bem como as variáveis tidas em conta para a análise dos dados. Na Secção 3.2., é apresentada a constituição da base empírica do estudo. Em 3.2.1. procede-se à descrição das bases de dados consultadas e das quais se seleccionaram as produções escritas que serão objeto de análise: o *Corpus* de Produções Escritas por Aprendentes de Português como L2 (PEAPL2) (coord. Cristina Martins) (Secção 3.2.1.1.); o Projeto de Recolha de Dados de Aprendizagem de Português Língua Estrangeira (RDAPLE) (coord. Isabel Leiria) (Secção 3.2.1.1.); e o *Corpus* de Aquisição de L2 (CAL2) (coord. Ana Madeira) (Secção 3.2.1.2.). Feita a descrição das bases de dados, na Secção 3.2.2. considera-se o processo subjacente à seleção dos textos que integram o *corpus* da presente investigação. Posteriormente, na Secção 3.3., apresenta-se o perfil dos sujeitos informantes, autores dos textos escritos que compõem o acervo empírico do estudo. Por fim, em 3.4., referem-se detalhadamente os procedimentos adotados para o tratamento dos dados seleccionados com a descrição das propriedades da tipologia de desvios elaborada para a presente investigação.

#### 3.1.1. Perguntas de investigação e variáveis em estudo

Como referido na **Introdução** da monografia, um dos principais objetivos do estudo consiste em determinar, com base na análise empírica de um acervo de textos com desvios de atribuição de valores de género aos nomes (AGN) e de concordância nominal em género (CNG) detetados dentro do sintagma nominal (SN), produzidos por aprendentes tardios de PLNM, falantes nativos de espanhol, italiano, alemão, inglês e chinês<sup>137</sup>, em diferentes

---

<sup>137</sup> O conjunto de falantes nativos de chinês integra informantes que são falantes nativos de mandarim e de cantonês. Todavia, dado o facto de nem sempre os aprendentes especificarem se a sua LM

níveis de aprendizagem, de que modo o conhecimento linguístico prévio, proveniente da LM, pode condicionar a assimilação dos valores de género dos itens nominais do português. Procura-se, então, averiguar se há padrões de comportamentos desviantes de atribuição de género aos nomes e de concordância que se correlacionem, quer com a LM do aprendente e respetivo nível de aprendizagem, quer com as propriedades do sistema de atribuição de género nominal do português, a LA de aprendizagem. Deste modo, em função dos objetivos, foram formuladas quatro perguntas de investigação. As duas primeiras perguntas correlacionam-se com a averiguação do (possível) papel da LM para a assimilação da categoria de género gramatical do português e dos seus respetivos valores. Nas duas últimas interrogamo-nos sobre se, e a partir dos padrões de comportamentos desviantes assinalados, é possível estabelecer uma relação entre os perfis dos aprendentes e as características do sistema de atribuição de género nominal do português e da concordância sintática dela decorrente.

Refira-se ainda que, para a formulação das quatro perguntas de investigação, teve-se em conta os objetivos estipulados para este estudo e as variáveis consignadas para a análise empírica, a saber: a LM do aprendente; o nível de proficiência dos informantes em português (nível QECRL); e as propriedades do sistema de atribuição de género nominal do português. Convém ainda salientar que a seleção das variáveis LM e nível de proficiência em português dependeu, por sua vez, da natureza dos dados disponíveis sobre os perfis dos informantes, aprendentes tardios de PLNM, nos acervos consultados, conforme se poderá constatar na Secção 3.3., dedicada à descrição do perfil dos sujeitos informantes.

Apresentam-se, em seguida, as perguntas delineadas para a investigação, bem como uma breve reflexão sobre a análise a empreender no tratamento dos dados empíricos selecionados em função de cada uma das questões colocadas.

---

correspondia ao mandarim ou ao cantonês, optou-se, neste trabalho, por adotar o termo ‘chinês’ para classificar a LM deste segmento da amostra.

**P1:** O desempenho linguístico do aprendente tardio relativamente à atribuição dos valores de género aos nomes e à concordância nominal em género no português beneficia do facto de ter representada na gramática da sua LM esta categoria gramatical?

Nesta questão, é equacionada a hipótese de haver um desempenho globalmente melhor na associação dos valores de género aos nomes por parte de aprendentes que são falantes nativos de um idioma com um sistema de classificação nominal baseado em género (espanhol, italiano e alemão) por oposição ao desempenho de aprendentes falantes nativos de uma língua sem a categoria de género gramatical (inglês e chinês). Com efeito, de acordo com os pressupostos da *Failed Functional Hypothesis* (Franceschina 2003, 2004; veja-se ainda o Capítulo 2 deste trabalho, Secção 2.3.3.2.3.), a aquisição de propriedades gramaticais parametrizadas, i.e., que não são seleccionadas por todas as línguas, como a categoria de género gramatical, só é possível quando estas se encontram representadas na LM do falante aprendente. Por conseguinte, poder-se-á esperar um melhor desempenho por parte dos aprendentes que são falantes nativos de espanhol, de italiano e de alemão, já que estas línguas possuem a categoria de género gramatical. Para além disso, entre estes informantes, coloca-se ainda a hipótese de haver um reduzido número de desvios de AGN e de CNG nas produções escritas por falantes nativos de espanhol e de italiano, relativamente ao número de desvios detetados em produções escritas por falantes nativos de alemão, visto que as línguas espanhola e italiana possuem sistemas de atribuição de valores de género nominal e de concordância nominal em género com propriedades similares às do português, quer no que respeita ao número de valores de género, masculino e feminino, quer no que concerne ao comportamento morfossintático das suas manifestações.

**P2:** O valor de género gramatical é uma propriedade transferível da LM para a LNM?

Na análise dos dados empíricos, procurar-se-á identificar, nos casos assinalados nas produções escritas por aprendentes cuja LM possui um sistema de classificação dos nomes baseado em género, se os valores relativos a esta categoria gramatical são uma propriedade potencialmente transferível da LM para a LA de aprendizagem. Ou seja, procurar-se-á aferir se é possível identificar padrões de comportamento correlacionados com a influência das propriedades formais e semânticas de atribuição dos valores de género aos nomes, existentes no idioma nativo do aprendente. Neste sentido, buscar-se-ão evidências de transferência de género, seja transferência de superfície ou profunda (cf. Sabourin *et al.*

2006), nos dados recolhidos dos segmentos da amostra dos sujeitos informantes que são falantes nativos de espanhol, italiano e alemão, já que estes idiomas exibem a categoria de género gramatical. Postula-se, conseqüentemente, que os aprendentes cuja LM apresenta um sistema de classificação dos nomes baseado em dois valores de género, masculino e feminino, como em espanhol e italiano, poderão, à partida, apresentar manifestações recorrentes de uma transferência direta dos valores da sua LM para a LA que não serão, em contrapartida, tão manifestamente visíveis nas produções escritas por falantes nativos de alemão, idioma com um sistema de atribuição de género nominal com particularidades distintas do português, pois apresenta três valores de género em oposição: masculino, feminino e neutro.

**P3:** Qual o grau de sensibilidade dos diferentes grupos de aprendentes relativamente aos indícios semânticos e formais de atribuição de valores de género aos nomes em português como LNM?

No âmbito desta questão, procurar-se-á compreender em que medida os diferentes grupos de informantes apresentam graus distintos de sensibilidade quanto aos indícios semânticos e formais de atribuição dos valores de género, masculino e feminino, aos itens nominais em português como LNM. Portanto, neste contexto, ter-se-á em conta, na análise dos comportamentos desviantes, as propriedades do *input* da LA e qual o seu (possível) papel ao longo das diferentes fases do processo de assimilação dos valores de género por parte dos aprendentes. Procura-se, assim, averiguar, nos vários segmentos da amostra que compõem o *corpus* em estudo, diferentes padrões de desenvolvimento da aquisição/aprendizagem da AGN e da CNG em PLNM correlacionados com as características do *input* da LA.

**P4:** Haverá, nos diferentes grupos de aprendentes, uma maior incidência de comportamentos desviantes de atribuição de valores de género aos nomes ou de concordância nominal em género nos especificadores (determinantes e quantificadores) ou nos modificadores (adjetivos) com função atributiva?

Para responder a esta última pergunta, procurar-se-á, na análise dos dados, averiguar que constituintes sintáticas dentro do SN e concordantes em género com o nome (especificadores e modificadores com função atributiva) apresentam uma maior incidência

de desvios de concordância nominal em género nos diferentes segmentos da amostra selecionada.

Em seguida, apresenta-se uma breve descrição das variáveis a ter em conta na análise dos dados empíricos:

▪ **LM do aprendente**

- ∇ A configuração da LM do aprendente no que concerne às propriedades dos respetivos sistemas de classificação nominal é tida em conta na análise dos dados empíricos. Distinguem-se, assim, os informantes cuja LM possui a categoria de género (espanhol, italiano e alemão) dos informantes que são falantes nativos de idiomas sem género gramatical (inglês e chinês). Entre as LM com género, distinguem-se ainda os idiomas com sistemas de dois valores de género em oposição (espanhol e italiano) e de três valores de género (alemão). Veja-se no Quadro 3.1 a síntese das principais características dos sistemas de classificação nominal das línguas selecionadas<sup>138</sup>, bem como o grau de proximidade destes relativamente ao sistema de atribuição de valores de género nominal do português, a LA de aprendizagem.

Grau de proximidade em relação ao sistema de atribuição de género do português (LA) ↑ + ↓ -	<b>LÍNGUAS ROMÂNICAS</b>
	<b>Espanhol</b>
	<ul style="list-style-type: none"> <li>■ Língua com dois valores de género em oposição (<b>masculino</b> vs. <b>feminino</b>).</li> <li>■ Possui um processo de marcação morfossintática e de concordância nominal muito semelhante ao da língua portuguesa.</li> </ul>
	<b>Italiano</b>
	<b>LÍNGUAS GERMÂNICAS</b>
	<b>Alemão</b>
	<ul style="list-style-type: none"> <li>■ Língua com três valores de género em oposição (<b>masculino</b> vs. <b>feminino</b> vs. <b>neutro</b>).</li> <li>■ Possui um sistema de concordância nominal complexo (género, número e caso)</li> </ul>
	<b>Inglês</b>
	<ul style="list-style-type: none"> <li>■ Língua sem sistema de atribuição de género nominal.</li> <li>■ Conserva vestígios de um primitivo sistema de género ao nível dos pronomes.</li> </ul>
	<b>LÍNGUA SINO-TIBETANA</b>
	<b>Chinês</b>
	<ul style="list-style-type: none"> <li>■ Língua sem sistema de classificação nominal baseado em género.</li> <li>■ Não admite a realização de relações de concordância sintática.</li> </ul>

**Quadro 3.1** – Síntese das características dos sistemas de classificação nominal das LM dos informantes, autores dos textos escritos que compõem o *corpus* da presente investigação (cf. Capítulo 1, Secção 1.4)

<sup>138</sup> Para uma descrição detalhada dos sistemas de classificação nominal dos idiomas selecionados, veja-se a Secção 1.4. do Capítulo 1 deste trabalho.

- **Nível de competência e de proficiência linguísticas no português (Nível QECRL)**

- ∇ Os textos seleccionados para a presente investigação encontram-se distribuídos por cinco níveis de proficiência linguística, desde o nível de iniciação, A1, até ao nível avançado, C1, atendendo ao nível da turma frequentada pelos informantes. Os descritores (A1, A2, B1, B2 e C1) correspondem aos estabelecidos pelo Conselho da Europa (2001) no âmbito do *Quadro Europeu Comum de Referência para as Línguas* (QECRL) e servirão de referência para a análise dos dados empíricos, já que permitem distinguir as diferentes fases do desenvolvimento linguístico dos aprendentes tardios.

- **Características de AGN do português**

- ∇ No tratamento dos dados empíricos ter-se-á em linha de conta as propriedades do *input* linguístico afetado. Por conseguinte, consideram-se como relevantes para a construção de uma tipologia de desvios, que servirá de base para a análise dos dados, as propriedades semânticas e formais de atribuição dos valores de género nominal do português, nomeadamente no que concerne à correlação entre os valores masculino e feminino e o carácter [+animado] do nome, por um lado, e, por outro, entre os valores de género masculino e feminino e os índices temáticos do item lexical *-o* e *-a*, respetivamente. Note-se ainda que na análise dos dados também se procurará averiguar se a própria estrutura formal dos nomes, i.e., se o facto de serem formas simples ou complexas (portadoras de afixos ou compostas) tem efeitos no número apurado de desvios de atribuição de valores de género aos nomes nos diferentes segmentos da amostra.

## 3.2. Constituição da base empírica do estudo

### 3.2.1. Descrição das bases de dados consultadas

A base empírica deste estudo é constituída por um acervo de textos escritos por aprendentes tardios de PLNM, selecionados em três bases de dados distintas<sup>139</sup>:

- ◆ **Corpus de Produções Escritas por Aprendentes de Português como L2** (PEAPL2) (coord. Cristina Martins), 1<sup>a</sup><sup>140</sup> e 2<sup>a</sup> fases de recolha, do Centro de Linguística Geral e Aplicada (CELGA-ILTEC) da Universidade de Coimbra<sup>141</sup>;
- ◆ **Projeto de Recolha de Dados de Aprendizagem de Português como Língua Estrangeira** (RDAPLE) (coord. Isabel Leiria) do Centro de Linguística da Universidade de Lisboa (CLUL)<sup>142</sup>; e
- ◆ **Secção “Produção Escrita de falantes aprendentes adultos” do Corpus de Aquisição de PL2** (CAL2) (coord. Ana Madeira) do Centro de Linguística da Universidade Nova de Lisboa<sup>143</sup>.

#### 3.2.1.1. Corpus PEAPL2 e Projeto Recolha de Dados de Aprendizagem de Português Língua Estrangeira (PLE)

O *Corpus* PEAPL2 disponibiliza, em regime de acesso livre e desde 2011, um conjunto de produções escritas por aprendentes tardios de PLNM que, entre maio de 2009 e maio de 2010 (fase 1 da recolha de dados) e entre janeiro e maio de 2011 (fase 2 da recolha de dados<sup>144</sup>), frequentaram cursos de português para estrangeiros na Faculdade de Letras da Universidade de Coimbra (FLUC), em turmas de diferentes níveis de aprendizagem, estipulados de acordo com o QECRL (Conselho da Europa 2011). Todos os textos que compõem o *Corpus* PEAPL2 estão devidamente identificados e organizados em

---

<sup>139</sup> Veja-se no Volume 2 da presente tese, o Anexo I que contém exemplos dos textos consultados nas três bases de dados selecionadas.

<sup>140</sup> Refira-se que os dados selecionados para este estudo, que integram a 1<sup>a</sup> fase do Projeto do *Corpus* PEAPL2, correspondem a um subconjunto de dados trabalhados por Ferreira (2011).

<sup>141</sup> Disponível para consulta, em regime de acesso livre, em <http://teitok2.iltec.pt/peapl2/>.

<sup>142</sup> Disponível para consulta, em regime de acesso livre, em <http://www.clul.ulisboa.pt/pt/24-recursos/350-recolha-de-dados-de-ple>.

<sup>143</sup> Disponível para consulta, mediante registo prévio, em <http://cal2.clunl.edu.pt/>.

<sup>144</sup> Os textos recolhidos na fase 2 do Projeto PEAPL2 foram elaborados, sobretudo, por aprendentes a frequentar turmas dos níveis intermédio (B2) e superior (C1).

função da LM do aprendente e do respetivo nível QECRL da turma frequentada, i.e., de A1 a C1.

Para além dos textos escritos, esta base de dados disponibiliza ainda um documento único com informações relevantes sobre os perfis dos informantes (Martins 2013). Estas informações foram recolhidas a partir dos dados fornecidos pelos próprios no momento em que respondiam aos inquéritos. A maioria dos textos do *Corpus* PEAPL2 foi escrita por alunos *Erasmus* e do *Curso Anual de Língua e Cultura Portuguesa para Estrangeiros* (CALCPE) que se encontravam a frequentar estes cursos num contexto de progressão da sua formação académica, permanecendo em Portugal durante um ou dois semestres. Há ainda um subconjunto de aprendentes imigrantes que, por viverem em território nacional, necessitam de frequentar cursos de português a fim de melhorar o conhecimento da língua e cultura portuguesas, sem deterem, necessariamente, formação académica superior.

O *Corpus* PEAPL2 resultou de uma parceria com Isabel Leiria, investigadora responsável pela organização e dinamização do Projeto de *Recolha de Dados de Aprendizagem de Português Língua Estrangeira*. Por conseguinte, as duas bases de dados partilham características que as tornam perfeitamente compatíveis, quer no que diz respeito aos estímulos utilizados nos inquéritos, quer no que concerne aos critérios adotados para o processo de recolha e de posterior transcrição dos textos.

À semelhança do *Corpus* PEAPL2, o Projeto *Recolha de Dados de Aprendizagem de Português Língua Estrangeira* disponibiliza, em regime de acesso livre, um acervo de materiais escritos por aprendentes de PLNM, bem como informação relativa aos respetivos perfis sociolinguísticos. As produções deste *corpus* foram recolhidas entre outubro de 2008 e outubro de 2010, tendo sido elaboradas por aprendentes tardios a frequentar cursos de PLNM de diferentes níveis de aprendizagem. No entanto, importa salientar que, ao contrário do que ocorre no *Corpus* PEAPL2, os autores dos textos escritos disponibilizados neste Projeto não se encontravam, no momento da recolha das produções, em contexto de imersão linguística. Ou seja, na altura da produção dos textos, estes aprendentes estavam a frequentar cursos de PLNM ministrados em 18 universidades estrangeiras espalhadas pelo mundo, mais propriamente nos seguintes países: Alemanha, Áustria, Bulgária, Coreia do Sul, Espanha, EUA, França, Índia, Itália, Polónia, Reino Unido, República Checa e Roménia<sup>145</sup>.

---

<sup>145</sup> Cf. <http://www.clul.ulisboa.pt/pt/24-recursos/350-recolha-de-dados-de-ple>.



Refira-se ainda que os textos recolhidos no âmbito do Projeto *Recolha de Dados de Aprendizagem de PLE* se encontram organizados por LM do informante e por nível QECRL da turma frequentada. Tendo em conta o nível, os textos foram distribuídos por três grupos: o grupo 1 integra textos de aprendentes a frequentar os níveis A1 e A2; o grupo 2 integra produções escritas por aprendentes nos níveis B1 e B2; e o grupo 3 engloba textos de aprendentes dos níveis superiores, i.e., C1 e C2<sup>146</sup>.

Cada produção das duas bases de dados resulta de uma resposta do aprendente a um estímulo previamente apresentado por escrito, tendo o informante de responder de acordo com o estímulo selecionado e sem poder recorrer a um dicionário ou a outro tipo de material de apoio<sup>147</sup>. Os estímulos utilizados, quer no *Corpus PEAPL2*, quer no Projeto de *Recolha de Dados de Aprendizagem de PLE*, foram selecionados a partir de um conjunto maior de estímulos. Este conjunto de base resultou, por sua vez, de um processo de revisão e de ampliação das várias propostas de redação elaboradas para a tese de doutoramento de Isabel Leiria<sup>148</sup> (Leiria 2006).

No Projeto de *Recolha de Dados de Aprendizagem de PLE* colocou-se à disposição dos professores colaboradores do Projeto uma lista com 83 estímulos, devidamente organizados por três grandes áreas temáticas contempladas no Projeto *Português Fundamental* (1984):

- (i) o indivíduo;
- (ii) a sociedade; e
- (iii) o meio ambiente.

Em função dessa lista, os professores selecionaram um subconjunto de 41 estímulos. Por sua vez, no *Corpus PEAPL2* foi selecionado, do acervo total, um subconjunto de 9 estímulos, equitativamente distribuídos pelas áreas temáticas acima referidas (cf. Martins 2013).

---

<sup>146</sup> Cf. <http://www.clul.ulisboa.pt/pt/24-recursos/350-recolha-de-dados-de-ple>.

<sup>147</sup> Refira-se ainda que, por vezes, um mesmo aprendente produziu mais do que um texto, em função de dois ou mais estímulos diferentes.

<sup>148</sup> Cf. Secção dedicada à Metodologia do Projeto *Recolha de Dados de Aprendizagem de PLE* (<http://www.clul.ulisboa.pt/pt/24-recursos/350-recolha-de-dados-de-ple>).

O processo de transcrição dos dados, quer no Projeto *Recolha de Dados de Aprendizagem de PLE*, quer no *Corpus PEAPL2*, obedeceu a um protocolo previamente estabelecido, tendo-se adotado as convenções propostas em Leiria (2006:181). Como se pode verificar na Figura 3.1, os critérios de transcrição dos documentos manuscritos permitem a identificação, por parte do investigador, de segmentos rasurados e/ou acrescentados pelos aprendentes (veja-se também, no Volume 2 deste trabalho, o Anexo I).

<xxx>	segmentos riscados
<(...)>	segmentos riscados ilegíveis
/xxx/	segmentos acrescentados
/*xxx/	leituras conjeturadas

**Figura 3.1** – Convenções adotadas durante o processo de transcrição dos documentos manuscritos que compõem o *Projeto de Recolha de Dados de Aprendizagem de PLE* e o *Corpus PEAPL2* (retirado de Leiria (2006:181))

### 3.2.1.2. *Corpus* de Aquisição de L2 (CAL2)

O *Corpus de Aquisição de L2* (CAL2) integra um Projeto mais abrangente intitulado “Morfologia e Sintaxe na Aquisição de L2”, desenvolvido, em 2004, no Centro de Linguística da Universidade Nova de Lisboa (CLUNL) sob a coordenação de Ana Madeira. O CAL2 disponibiliza, através de registo prévio, um vasto acervo de dados escritos e orais transcritos, produzidos por aprendentes adultos e infantis de PLNM, para além de dados produzidos por falantes nativos do português que constituem, assim, o grupo de controlo.

Atendendo aos objetivos da presente investigação, consideraram-se apenas as produções escritas por falantes adultos, aprendentes de PLNM a frequentar cursos de PLE em universidades portuguesas, em turmas com diferentes níveis de aprendizagem. Estas produções foram recolhidas entre o ano 2001 e o ano 2010<sup>149</sup>.

As produções disponibilizadas no CAL2 encontram-se organizadas por LM do aprendente e nível de proficiência em português da turma frequentada, tendo sido feita a distinção dos textos produzidos por aprendentes dos níveis de “Iniciação”, “Intermédio” e “Avançado”.

A informação relativa aos perfis dos informantes, dados pessoais (idade, nacionalidade, sexo), percurso linguístico (língua materna e outras línguas que o aprendente conheça além do português) e percurso de aquisição/aprendizagem do português, com referência ao ano de início de exposição à língua portuguesa, é apresentada no cabeçalho

---

<sup>149</sup> A informação relativa à data de recolha dos textos é facultada no cabeçalho de cada produção escrita. Veja-se, no Volume 2 deste trabalho, o Anexo I.

do documento relativo a cada texto, existindo um código associado a cada tipo de informação, conforme se explicita na Figura 3.2:

<T>	- tipo de texto
<D>	- data de produção
<DN>	- data de nascimento
<I>	- idade
<S>	- sexo
<LN>	- local de nascimento
<LM>	- língua materna
<PP>	- profissão dos pais
<L~M>	- línguas não maternas
<NCP>	- nível de proficiência linguística

Figura 3.2 – Códigos utilizados no cabeçalho dos textos disponibilizados no CAL2

Para o processo de transcrição dos documentos manuscritos<sup>150</sup> disponibilizados no CAL2 foram adotados critérios distintos dos seguidos no *Corpus* PEAPL2 e no Projeto *Recolha de Dados de Aprendizagem de PLE*. Nestes textos, os segmentos rasurados graficamente no documento *word* (como, por exemplo, ~~essa~~) correspondem aos segmentos apagados pelo sujeito informante, não se tendo, contudo, feito a sinalização dos segmentos acrescentados, à semelhança das bases de dados anteriores.

O CAL2 disponibiliza diferentes tipos de texto (descritivos, argumentativos, expositivos, epistolares, etc.) que resultam de respostas dos aprendentes a vários tipos de tarefas escritas e nas quais se incluem tarefas propostas em contexto de avaliação de conhecimentos. Assim sendo, estes textos não resultam, necessariamente, de estímulos previamente estipulados para o Projeto, tal como se verificou nas bases de dados anteriormente descritas (cf. Secção 3.2.1.1.). Consequentemente, os vários textos abordam temáticas muito diversas, o que não inviabiliza, segundo cremos, a compatibilidade destas produções escritas relativamente às que constam no Projeto de *Recolha de Dados de Aprendizagem de PLE* e no *Corpus* PEAPL2.

---

<sup>150</sup> As normas de informatização seguidas no âmbito da constituição do CAL2 foram gentilmente cedidas pela Professora responsável do Projeto, a Professora Doutora Ana Madeira.

### 3.2.2. Seleção do *corpus* da investigação

Para a constituição do *corpus* deste estudo, procedeu-se, num primeiro momento, à análise do conjunto total de textos recolhidos das bases de dados anteriormente descritas (*Corpus* PEAPL2 – 1<sup>a</sup> <sup>151</sup> e 2<sup>a</sup> fases de recolha; Projeto de *Recolha de Dados de Aprendizagem de PLE*; e CAL2), escritos por falantes nativos das LM selecionadas em diferentes níveis QERCL.

Relativamente à distribuição dos textos por nível QECRL em português, constatou-se que as três bases de dados possuem um sistema de organização distinto e, por esse motivo, foi necessário uniformizar os dados disponibilizados. Com efeito, os textos recolhidos do *Corpus* PEAPL2 apresentam-se distribuídos por cinco níveis de proficiência linguística, atendendo ao nível da turma frequentada pelos informantes e que são referidos de acordo com os descritores estabelecidos pelo Conselho da Europa (2001:49) no âmbito do *Quadro Europeu Comum de Referência para as Línguas* (QECRL): A1, A2, B1, B2 e C1.

Por sua vez, os textos do Projeto de *Recolha de Dados de Aprendizagem de PLE* encontram-se distribuídos por três grupos, aos quais foi atribuído um código numérico: o código 1 integra textos de níveis A1 e A2; o código 2 integra textos de níveis B1 e B2, e, por fim, o código 3 engloba textos dos níveis C1 e C2.

No *Corpus* de Aquisição de L2, os textos foram agrupados em função de uma classificação de tipo qualitativo, i.e., estão organizados pelo nível «iniciação», «intermédio» e «avançado», a que corresponderão os níveis QECRL, A1-A2, B1-B2 e C1, respetivamente.

Assim, no sentido de compatibilizar os dados que constituem o acervo textual da presente investigação, optou-se pela sua organização em três grupos distintos, identificados com os descritores propostos no QECRL: (i) A1-A2; (ii) B1-B2; e (iii) C1 <sup>152</sup>.

Veja-se no Quadro 3.2 a distribuição do número total de textos consultados nas três bases de dados durante o processo de constituição da base empírica da presente investigação, distribuídos por LM dos informante e nível QECRL da turma frequentada <sup>153</sup>.

---

<sup>151</sup> Conforme referido na nota 140, os textos selecionados da 1<sup>a</sup> fase de recolha de dados do *Corpus* PEAPL2 correspondem a um conjunto de produções analisadas por Ferreira (2011).

<sup>152</sup> Atendendo às informações disponibilizadas nos textos analisados, verificou-se que não existem informantes do nível C2.

<sup>153</sup> Refira-se que, por vezes, um mesmo aprendente produziu mais do que um texto, o que levou à diferença entre o número de informantes e o número de textos.

Informantes				Produções escritas
Língua Materna	Nível QECRL	Base de dados <sup>154</sup>	#	#
Espanhol	A1-A2	PEAPL2	19	20
		RDAPLE	35	36
		CAL2	71	103
	B1-B2	PEAPL2	33	41
		RDAPLE	21	30
		CAL2	51	75
	C1	PEAPL2	5	7
		RDAPLE	0	0
		CAL2	37	77
$\Sigma$			<b>272</b>	<b>389</b>
Italiano	A1-A2	PEAPL2	27	40
		RDAPLE	49	65
		CAL2	95	151
	B1-B2	PEAPL2	29	43
		RDAPLE	32	35
		CAL2	51	73
	C1	PEAPL2	0	0
		RDAPLE	5	6
		CAL2	31	53
$\Sigma$			<b>319</b>	<b>466</b>
Alemão	A1-A2	PEAPL2	17	21
		RDAPLE	40	41
		CAL2	85	102
	B1-B2	PEAPL2	47	68
		RDAPLE	0	0
		CAL2	34	40
	C1	PEAPL2	0	0
		RDAPLE	0	0
		CAL2	41	69
$\Sigma$			<b>264</b>	<b>341</b>
Inglês	A1-A2	PEAPL2	7	12
		RDAPLE	20	20
		CAL2	43	58
	B1-B2	PEAPL2	27	45
		RDAPLE	8	12
		CAL2	37	37
	C1	PEAPL2	5	6
		RDAPLE	2	4
		CAL2	29	57
$\Sigma$			<b>178</b>	<b>251</b>
Chinês	A1-A2	PEAPL2	19	23
		RDAPLE	0	0
		CAL2	32	40
	B1-B2	PEAPL2	11	11
		RDAPLE	0	0
		CAL2	36	48
	C1	PEAPL2	10	12
		RDAPLE	0	0
		CAL2	11	14
$\Sigma$			<b>119</b>	<b>148</b>
$\Sigma$			<b>1 152</b>	<b>1 595</b>

**Quadro 3.2** – Acervo total de produções escritas analisadas nas três bases de dados, distribuídas por LM e nível QECRL da turma frequentada pelos informantes

<sup>154</sup> Bases de dados consultadas: (1) PEAPL2 (*Corpus* de Produções Escritas por Aprendentes de PL2); (2) RDAPLE (Projeto Recolha de Dados de Aprendizagem de PLE) e (3) *Corpus* de Aquisição de L2.

Como se verifica no Quadro 3.2, há um maior número de informantes de LM italiana e espanhola, sobretudo nos níveis A1-A2, relativamente ao número registado nos restantes segmentos da amostra. Por sua vez, o segmento da amostra de aprendentes que são falantes nativos de chinês é o mais reduzido e, conseqüentemente, regista um menor número de textos.

Em termos globais, constata-se que o acervo consultado é considerável, constituído, aproximadamente, por 1600 produções escritas. Por razões de natureza metodológica e atendendo à dimensão do acervo consultado, optou-se pela seleção de textos com desvios de AGN e de CNG dentro do SN.

Portanto, tendo em conta que o foco da presente investigação é o desvio de AGN e de CNG no SN, selecionaram-se textos que apresentam, precisamente, situações nas quais os aprendentes atribuem erradamente o valor de género ao nome, sendo que este se repercute, nos termos definidos por Corbett (1991), através do mecanismo da concordância sintática. Assim, em função deste pressuposto, consideraram-se os casos em que há uma atribuição de valores de género desviante visível, quer pela forma dos especificadores (determinantes e quantificadores), quer pela forma dos modificadores (adjetivos) integrados no SN, como se ilustra nos exemplos seguintes:

- (1) «\*Fui **um viagem** não muito **barato**» (Italiano.A1-A2);
- (2) «\*pagar um apartamento sem **o ajuda** <dos meus> do meu pai» (Inglês.B1-B2);
- (3) «\*Não temos **um televisão**» (Alemão.C1)

Os casos dos desvios de AGN e de CNG que incidem sobre modificadores com função de predicativo do sujeito ficaram fora do escopo da análise empírica, i.e., não se trataram, nesta investigação, as situações em que apenas se verifica a atribuição desviante de valor de género no adjetivo com a função de predicativo de sujeito, como, por exemplo:

- (4) \*O meu carro é **branca**.

Esta opção decorreu do facto de se ter preferido observar, para esta investigação, a concordância em género dentro do SN, uma vez que o adjetivo que assume a função de predicativo do sujeito se encontra, tipicamente, mais distanciado do núcleo nominal com o

qual tem de concordar, o que pode implicar uma maior dificuldade de processamento e um maior ónus para a memória de curto prazo do aprendente tardio<sup>155</sup>.

Convém ainda ressaltar que, embora a base empírica deste estudo seja composta por produções com desvios de atribuição dos valores de género (masculino ou feminino) assinalados nas formas dos elementos que compõem o SN, no tratamento dos dados apurados também se terá em linha de conta, para além das ocorrências de nomes afetados por desvios de AGN e de CNG, as ocorrências convergentes com a LA desses mesmos nomes. Pretende-se, assim, obter, a partir da observação conjunta dos dados, convergentes e não convergentes, uma descrição, o mais detalhada possível, do desenvolvimento do processo de assimilação dos valores de género nominal em PLNLM por aprendentes tardios.

Apresentados os pressupostos subjacentes à seleção dos dados empíricos para o estudo, no Quadro 3.3 regista-se, por segmentos da amostra, a distribuição do número de informantes e de produções escritas selecionadas que constituem a base empírica deste estudo.

---

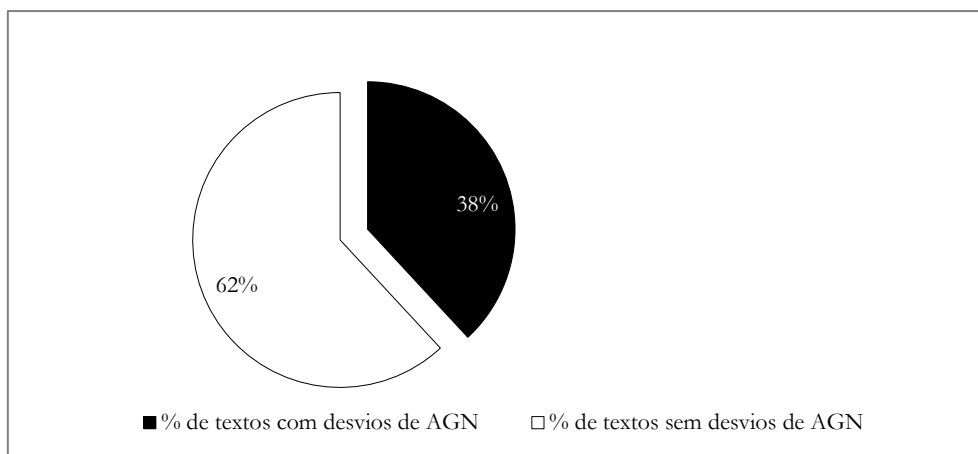
<sup>155</sup> Note-se, porém, que apesar de, no presente estudo, se ter tomado esta opção, em futuras investigações pretendemos dar conta, precisamente, da relação entre a posição do adjetivo face ao nome para o processamento dos valores de género nominal em português como LNM.

Informantes			Produções escritas	
Língua Materna	Nível QECRL da turma frequentada	#	#	#de palavras
Espanhol	A1-A2	42	46	8 029
	B1-B2	24	26	5 853
	C1	9	11	3 479
$\Sigma$		<b>76</b>	<b>83</b>	<b>17 361</b>
Italiano	A1-A2	88	104	19 280
	B1-B2	35	40	9 353
	C1	12	12	9 249
$\Sigma$		<b>136</b>	<b>156</b>	<b>37 882</b>
Alemão	A1-A2	70	74	11 915
	B1-B2	48	60	15 598
	C1	16	20	5 563
$\Sigma$		<b>134</b>	<b>154</b>	<b>33 076</b>
Inglês	A1-A2	38	46	6 166
	B1-B2	42	54	16 624
	C1	20	29	6 505
$\Sigma$		<b>100</b>	<b>129</b>	<b>29 295</b>
Chinês	A1-A2	28	37	6 423
	B1-B2	25	34	7 705
	C1	9	10	2 093
$\Sigma$		<b>62</b>	<b>81</b>	<b>16 221</b>
$\Sigma$		<b>506</b>	<b>603</b>	<b>133 835</b>

**Quadro 3.3** – Acervo textual com desvios de AGN e de CNG que compõem o *corpus* da investigação

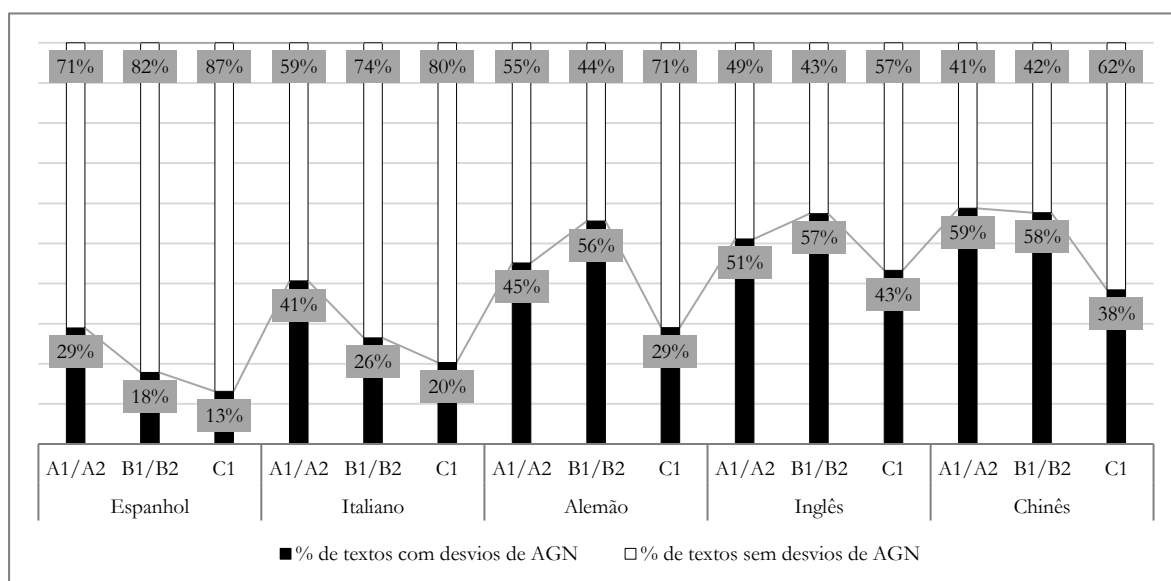
Assim, do conjunto total analisado de **1595** produções escritas por **1152** informantes (cf. Quadro 3.2), constatou-se que **603** textos escritos por **506** informantes apresentavam desvios de AGN e de CNG dentro do SN (cf. Quadro 3.3). Ora, em termos proporcionais, e como é ilustrado no Gráfico 3.1, 38% dos textos analisados nas três bases de dados apresentam desvios de AGN e de CNG, sendo este o acervo que constitui a base empírica do presente estudo.





**Gráfico 3.1** – Percentagem de textos consultados nas três bases de dados com desvios de AGN e de CNG em relação à percentagem de textos sem desvios

O Gráfico 3.2 ilustra, em valores percentuais (numa escala de 0 a 100), a proporção de produções escritas com desvios de AGN e de CNG no SN apurada por cada segmento da amostra, sendo que esta proporção foi contabilizada em função do número absoluto de textos produzido nos diferentes segmentos (cf. Quadro 3.2).



**Gráfico 3.2** – Proporção (numa escala de 0 a 100) de textos com desvios de AGN e de CNG no SN, face ao número de textos sem desvios, consultados nas três bases de dados

Segundo os dados cartografados no Gráfico 3.2, é sobretudo nos níveis QECRL A1-A2 e B1-B2 que se regista a maior proporção de textos com desvios de AGN e de CNG, sobretudo nas produções escritas por falantes nativos de línguas sem a categoria de género gramatical (inglês e chinês) e também por falantes nativos de alemão. Regista-se, em contrapartida, uma menor proporção de textos com desvios entre os informantes de LM românicas, especialmente no segmento da amostra de falantes nativos de espanhol. Porém,

é igualmente assinalável a taxa de acerto por todos os segmentos da amostra, ou seja, estes dados parecem evidenciar que os aprendentes, no momento de atribuir um valor de género a um nome, núcleo de SN, selecionam adequadamente as formas dos elementos do SN que, com o nome, estabelecem relações de concordância sintática, acertando, em geral, mais do que erram.

### 3.3. Perfil dos informantes

Como se referiu nas subsecções anteriores, o acervo textual que compõe o *corpus* da presente investigação resultou da recolha de um conjunto de produções escritas por 506 aprendentes tardios de PLNM (cf. Quadro 3.3), disponibilizadas em três bases de dados distintas:

- *Corpus* PEAPL2;
- Projeto de *Recolha de Dados de Aprendizagem de PLE*; e
- *Corpus* de Aquisição de L2 (CAL2).

Como também vimos, os *corpora* consultados são perfeitamente compatíveis e utilizáveis para o efeito previsto no presente estudo, pelo que deles se extraíram os textos aqui analisados. Porém, no que concerne às informações disponibilizadas acerca dos perfis dos aprendentes, assinalam-se algumas particularidades distintas entre as bases de dados selecionadas, o que é compreensível dado o facto de a metodologia adotada por cada Projeto não ter sido a mesma. Ainda assim, foi possível organizar e tratar os dados de um modo que permitiu o seu cotejo.

Portanto, atendendo à natureza deste estudo, apresentam-se as informações relativas aos perfis dos informantes, nomeadamente, no que concerne à sua distribuição por:

- ∇ língua materna (Secção 3.3.1.);
- ∇ outras línguas não maternas (Secção 3.3.2.);
- ∇ nível QECRL em português (Secção 3.3.3.);
- ∇ idade (Secção 3.3.4.);
- ∇ período de exposição ao português (Secção 3.3.5.); e
- ∇ experiência de imersão<sup>156</sup> (Secção 3.3.6.).

---

<sup>156</sup> Como já referido anteriormente neste capítulo (cf. Secção 3.2.1.1.), nem todos os informantes se encontravam, no momento da produção dos textos, em contexto de imersão linguística, havendo, então, um subconjunto de informantes que não frequentaram cursos de PLNM em Portugal. Pese embora o facto de no tratamento dos dados a imersão linguística não constituir uma variável de análise, considerámos que a sua referência é relevante para a caracterização dos perfis dos informantes.

### 3.3.1. Língua Materna

Os 506 informantes distribuem-se de forma diferenciada pelos vários grupos de LM selecionados para este estudo. Conforme se pode constatar no Quadro 3.4, não há uma distribuição equitativa dos informantes por grupo de LM, havendo um maior número de informantes que são falantes nativos de alemão e de italiano. Em contrapartida, o segmento da amostra de informantes de língua chinesa é o mais reduzido. Esta mesma tendência já se verificava, aliás, no acervo global de dados disponível nos *corpora* textuais analisados (cf. Quadro 3.2).

Língua Materna	# Informantes
Espanhol	75
Italiano	135
Alemão	134
Inglês	100
Chinês	62
$\Sigma$	506

Quadro 3.4 – Distribuição do número de informantes por LM

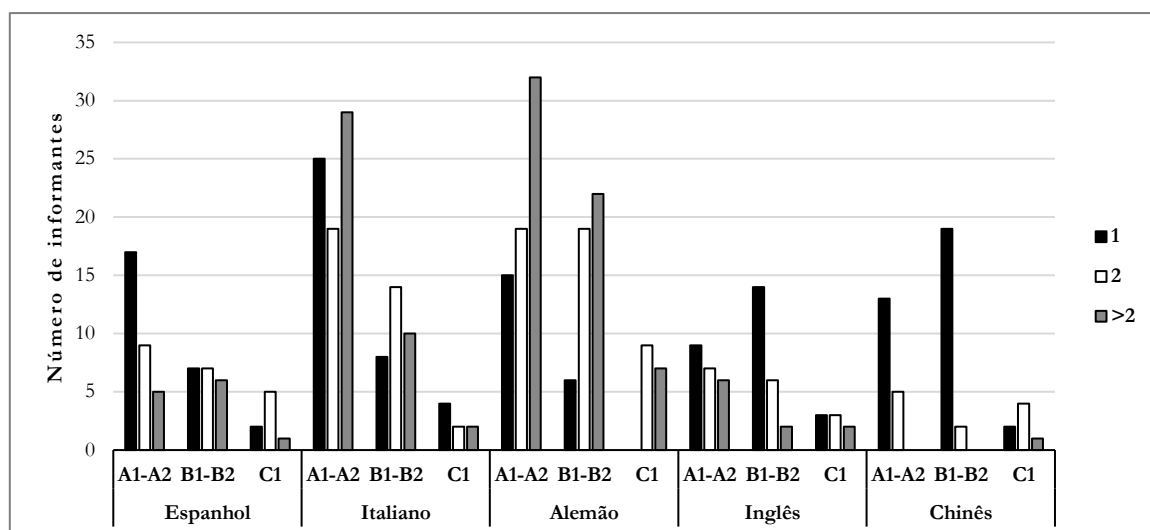
### 3.3.2. Outras línguas não maternas

Para a descrição dos perfis dos informantes, é igualmente relevante assinalar a informação relativa ao seu percurso linguístico e, de acordo com as respostas facultadas pelos próprios aprendentes<sup>157</sup>, foi possível identificar outros idiomas previamente conhecidos, para além da LM e do português.

Assim, considere-se o Gráfico 3.3, no qual é possível verificar a distribuição dos informantes por número de LNM conhecida(s) (1, 2 ou mais do que 2).

---

<sup>157</sup> Dos 506 informantes que compõem o *corpus*, 109 não facultaram quaisquer dados relativamente ao conhecimento de outras LNM.



**Gráfico 3.3** – Distribuição do número de informantes por número de outra(s) LNM conhecida(s) (1, 2 ou mais do que 2)

A partir das informações disponibilizadas, constata-se um predomínio de informantes com conhecimentos de uma outra LNM. Note-se ainda que no segmento da amostra de falantes nativos de italiano e, sobretudo, no segmento de falantes nativos de alemão, se regista um número elevado de informantes que considera ter conhecimentos em mais do que duas outras LNM.

Entre os idiomas assinalados nas respostas, a maioria dos aprendentes refere ter conhecimentos de inglês. Com efeito, grande parte dos aprendentes refere ser proficiente neste idioma. No caso dos falantes de LM inglesa, a maioria refere ter conhecimentos de espanhol.

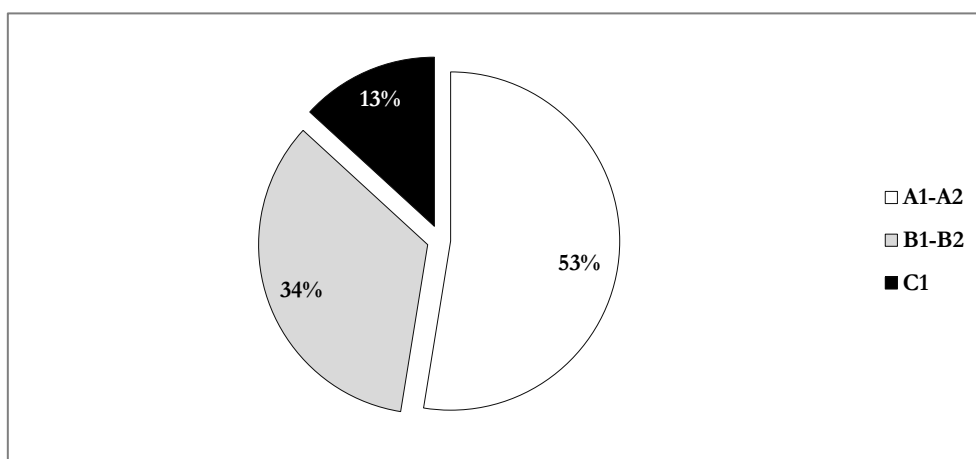
O conhecimento de outras LNM pode igualmente influenciar o desenvolvimento linguístico dos falantes aprendentes de PLNM e, mais especificamente, a aquisição/aprendizagem da atribuição dos valores de género nominal em português. Há, com efeito, vários estudos que avaliam, precisamente, o papel do conhecimento linguístico prévio das LNM para a assimilação dos valores de género nominal em PLNM (cf. Lacsán 2015; Pinto 2015). Porém, nos dados recolhidos para este estudo, não é possível obter uma indicação precisa da LNM (ou das LNM) que cada falante aprendente conhece, nem dos níveis QECRL dos informantes nesses mesmos idiomas. Por conseguinte, e pese embora a relevância desta informação, esta variável não será tida em conta na análise.

### 3.3.3. Nível QECRL em português

Para além da LM, outra variável importante de análise dos dados é o nível QECRL em português da turma frequentada pelos informantes. Como se verificou, para a organização dos textos foi necessário compatibilizar os dados (Cf. Secção 3.2.2.), tendo-se agrupado as diferentes produções escritas em três grupos distintos:

- (i) A1-A2;
- (ii) B1-B2;
- (iii) C1.

A partir da informação representada no Gráfico 3.4, é possível concluir que os informantes se encontram, maioritariamente, em turmas dos níveis A1 e A2 (53% do total de sujeitos da amostra) e o nível C1 é o menos representado. Mais uma vez, esta tendência também já se verificava no conjunto global de textos analisado (cf. Quadro 3.2), havendo nos níveis A1 e A2 um maior número de produções escritas do que nos restantes níveis.



**Gráfico 3.4** – Distribuição (em valores percentuais) do conjunto total de informantes por nível QECRL da turma frequentada

O Gráfico 3.5 regista a distribuição do número de informantes por LM e o nível QECRL em português da turma frequentada.

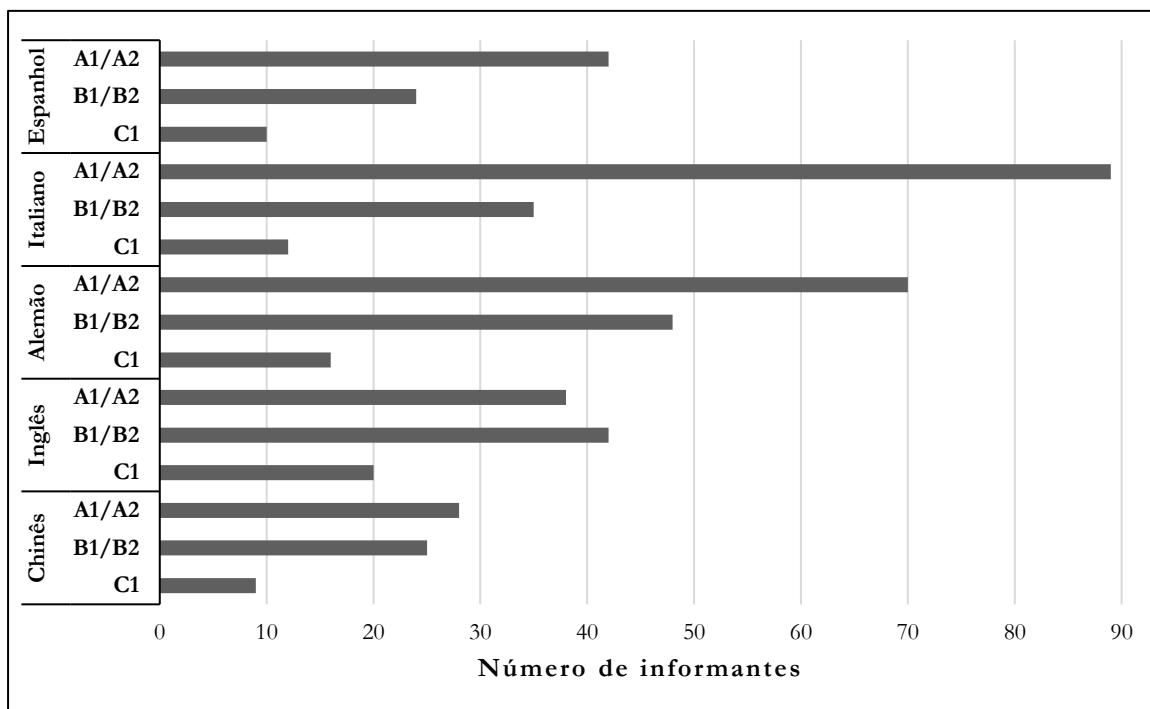


Gráfico 3.5 – Distribuição do número de informantes por nível QECRL em português da turma frequentada

### 3.3.4. Idade

Segundo foi possível apurar, o valor médio de idade do conjunto dos 506 informantes no momento da produção dos textos é de 24 anos<sup>158</sup>. No Gráfico 3.6, apresenta-se a distribuição dos informantes pelas diferentes faixas etárias.

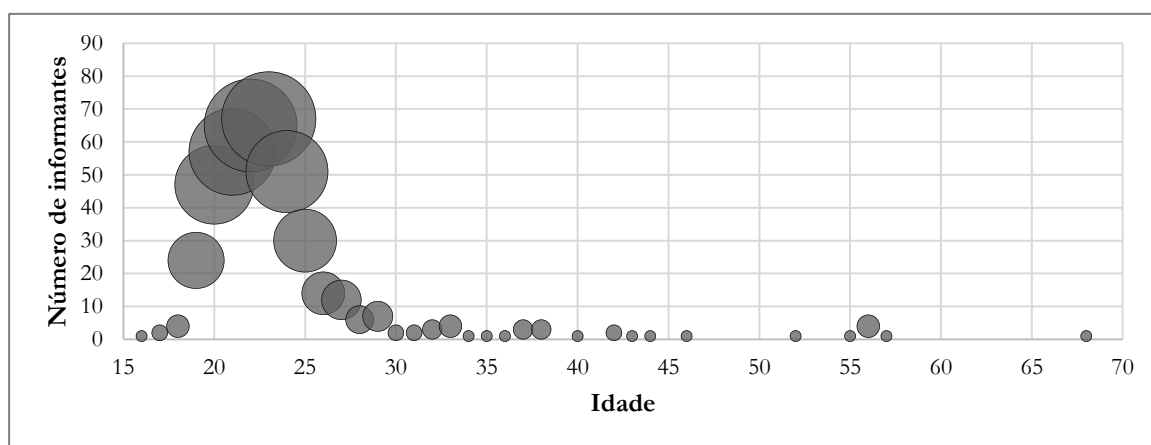


Gráfico 3.6 – Distribuição do número de informantes por faixa etária

Como se pode verificar, há uma maior concentração do número de informantes com idades compreendidas entre os 20 e os 25 anos.

<sup>158</sup> Do conjunto dos 506 informantes, 87 não forneceram informação relativamente à idade.

### 3.3.5. Período de exposição ao português

Para além da idade, importa ter em conta o período de exposição dos informantes à LA, o português. Ora, de acordo com a informação disponibilizada em cada base de dados consultada, foi possível aferir o período de tempo durante o qual os falantes estiveram expostos ao português, até ao momento da produção dos textos<sup>159</sup>.

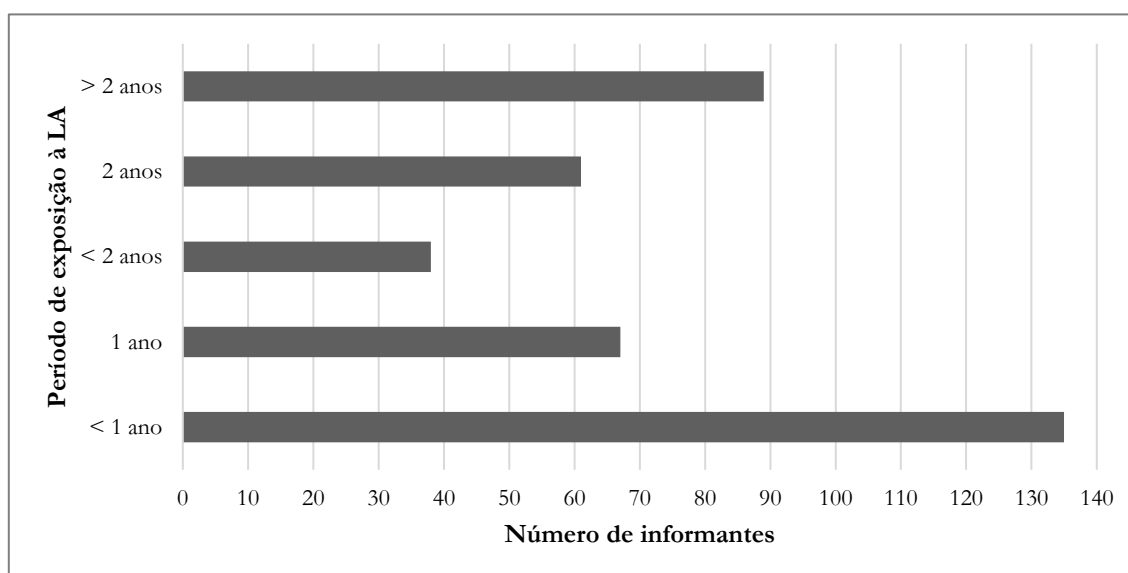


Gráfico 3.7 – Distribuição do número de informantes por período de exposição à LA

A partir dos dados cartografados no Gráfico 3.7, regista-se uma maior concentração do número de informantes com um período de exposição ao português inferior a um ano, que se explica dada a maior proporção de informantes dos níveis A1-A2 (cf. Gráficos 3.5 e 3.8).

Assim sendo, tendo em conta que a maioria dos aprendentes se situa entre os 20 e os 25 anos de idade (cf. Gráfico 3.6), é possível confirmar que se trata, em todos os casos, de aprendentes tardios de PLNM e que nenhum teve exposição ao *input* da LA antes dos 16 anos.

<sup>159</sup> Do conjunto dos 506 informantes, 118 não forneceram dados que permitissem averiguar o tempo de exposição à LA.



Partindo das várias informações disponibilizadas e vistas até ao momento, averiguou-se a distribuição conjunta dos informantes em função de: (i) LM; (ii) nível QECRL da turma frequentada; e (iii) período de exposição à LA de aprendizagem. Essas informações encontram-se cartografadas, em conjunto, no Gráfico 3.8.

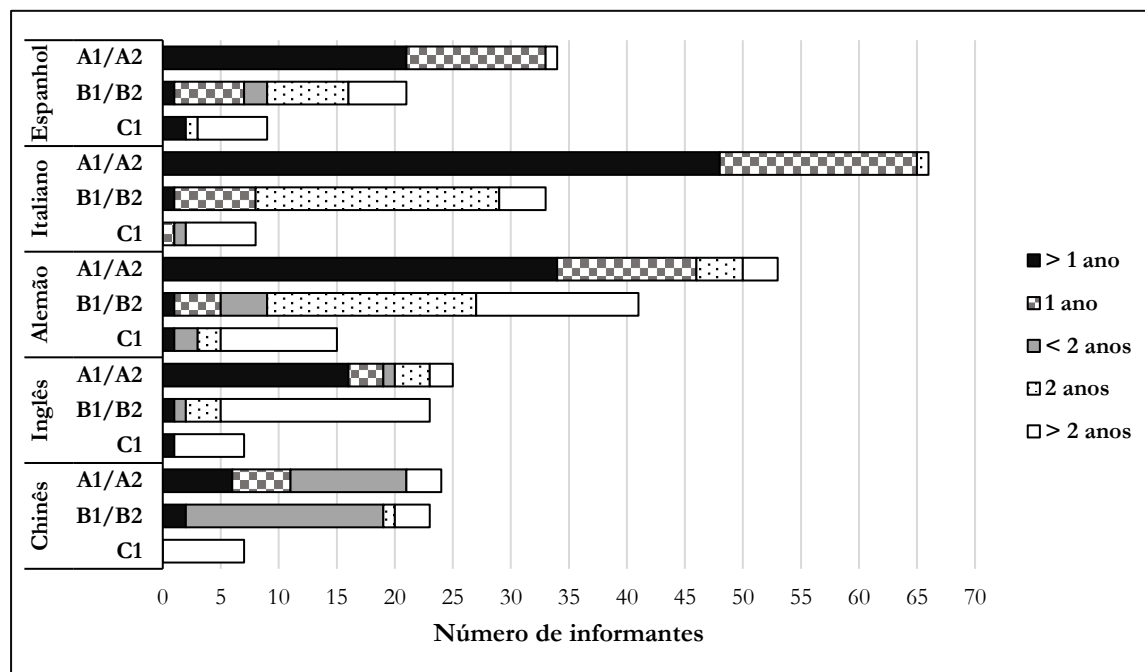


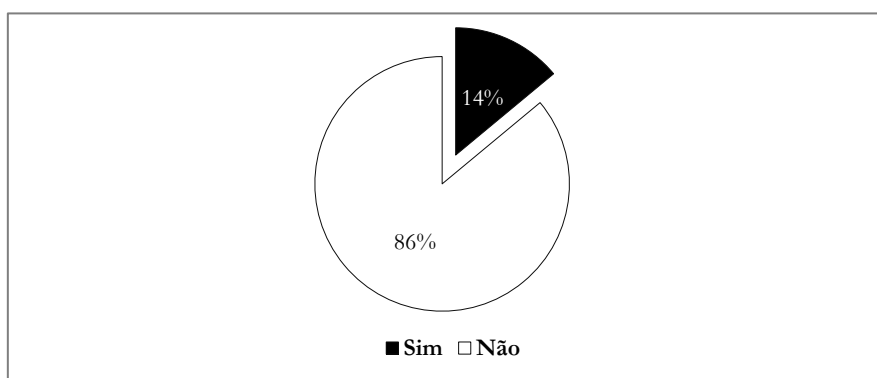
Gráfico 3.8 – Distribuição dos informantes por LM, nível QECRL e período de exposição à LA

Partindo dos dados cartografados no Gráfico 3.8, verifica-se que, de um modo global e em todos os segmentos da amostra, os informantes com um maior período de exposição à LA (superior a 2 anos) encontram-se, e como seria de esperar, nos níveis QECRL mais avançados em PLNM. Ou seja, um período de exposição à LA mais longo associa-se tipicamente ao nível QECRL mais avançado. Esta tendência é, aliás, notória nos falantes nativos de chinês de C1, já que todos estes informantes apresentam um período de exposição ao português superior a 2 anos, enquanto nos restantes grupos de aprendentes (espanhol, italiano, alemão e inglês) há, no C1, vários informantes com períodos menores de exposição. Estes indicadores sugerem, então, que nestes segmentos da amostra poderá haver um efeito de LM ou até de outras LNM que justifiquem a sua inserção numa turma de nível avançado.

### 3.3.6. Experiência de imersão

No momento da recolha dos dados, nem todos os 506 informantes se encontravam em contexto de imersão linguística. Com efeito, conforme já referido, os textos recolhidos no âmbito do Projeto de *Recolha de Dados de Aprendizagem de PLE* foram produzidos por aprendentes a frequentar cursos de PLNM em diferentes universidades estrangeiras (cf. Secção 3.2.1.1.). Na presente investigação não se terá em conta o peso da imersão linguística para o tratamento dos dados, uma vez que a consideração desta variável implicaria uma análise mais fina das questões pertinentes à imersão linguística, nomeadamente no que concerne à descrição exaustiva das situações de contacto e de uso do português por cada informante. Ora, os dados disponibilizados nos acervos consultados não permitem extrair informações detalhadas e claras a este respeito, pelo que a consideração da imersão linguística para a análise dos comportamentos desviantes não ficaria suficientemente sustentada<sup>160</sup>. Assim sendo, apresentamos, neste ponto, a distribuição dos informantes atendendo à sua presença em Portugal e fora de Portugal no momento da recolha dos textos (Gráfico 3.9), fazendo ainda uma breve referência aos índices de desvios produzidos por conjunto de informantes, com e sem imersão linguística. Considerou-se pertinente a apresentação destes dados nesta subsecção do trabalho, uma vez que, e como se verá, tal informação vai permitir, igualmente, justificar o facto de não se ter considerado a imersão linguística como variável de análise.

Face ao número total de informantes (506), verifica-se que, em termos proporcionais, 14% dos aprendentes não se encontram em contexto de imersão linguística, sendo que este valor percentual corresponde, em termos absolutos, a um total de 71 informantes (cf. Gráfico 3.9).



**Gráfico 3.9** – Distribuição, em valores percentuais, dos informantes em contexto de imersão linguística (SIM) face aos informantes que não se encontram em imersão linguística (NÃO)

<sup>160</sup> Sobre a possível influência da imersão linguística para a assimilação dos valores de género nominal em PLNM, veja-se, por exemplo, o estudo de Lacsán (2015).

No que diz respeito à distribuição dos desvios por informantes, em imersão linguística e sem imersão linguística, constata-se que, pelo menos em termos quantitativos, não são observáveis divergências consideráveis a este nível entre estes dois segmentos da amostra, como se pode verificar no Quadro 3.5, já que, por segmentos da amostra de informantes, em imersão e sem imersão linguística, as percentagens relativas de desvios apuradas, calculadas em função do número absoluto de nomes em SN produzido em cada segmento, são muito próximas, 5,7% e 6,7%, respetivamente.

	Informantes		$\Sigma$
	em imersão linguística	sem imersão linguística	
# de ocorrências de nomes em SN	19 483	1980	21 463
# de desvios	1 108	133	1 241
% de desvios	5,7	6,7	5,8

**Quadro 3.5** – Distribuição do número de nomes em SN, de desvios de AGN e de CNG e da respetiva percentagem relativa por segmentos de amostra de informantes, em imersão linguística e sem imersão linguística



### 3.4. Procedimentos de tratamento dos dados

#### 3.4.1. Contabilização de itens nominais

Numa primeira fase do tratamento dos dados, foi necessário determinar o número das ocorrências dos itens lexicais que, em cada texto, poderiam considerar-se nomes. De acordo com este pressuposto, contabilizou-se como um único nome os compostos sintáticos como, por exemplo, *fim de semana* e *ponto de vista*.

Para além disso, na contabilização dos itens nominais, estabeleceram-se critérios de exclusão da contagem, optando-se por não contabilizar as ocorrências de:

- (i) topónimos, tais como *Coimbra*, *Portugal*, *Pavia*, *Milão* ou *Europa*, já que este tipo de itens apresenta algumas particularidades relativamente à sua distribuição por valores de género masculino ou feminino. Em boa verdade, o facto de haver topónimos que, tipicamente, não exigem a presença de artigo faz com que não seja claro o valor de género que lhes está associado (cf. Raposo & Nascimento 2013:1011).
- (ii) antropónimos, dado que estes nem sempre surgem identificados nos textos e são substituídos por um código que não permite a sua identificação, como “XXX”. A omissão destes itens prende-se, precisamente, com as questões éticas relacionadas com a omissão da identificação dos autores dos textos escritos; e
- (iii) nomes que correspondem a importações diretas de outras línguas e que não integram o léxico da língua portuguesa, como, por ex., nomes de pratos gastronómicos típicos do país de origem dos aprendentes (*lasagna*, *canneloni*) ou de festividades tradicionais, como *L'abuffata deli strazzati*, nome de uma festa típica de Itália.

Estipulados os critérios de exclusão dos itens<sup>161</sup>, procedeu-se à contagem das ocorrências dos itens nominais por texto, excluindo-se todos os segmentos rasurados pelos aprendentes, já que estes não correspondem ao produto final pretendido pelo informante. O Quadro 3.6 apresenta a distribuição do número de palavras e de nomes por texto, com a indicação, não só dos números absolutos e dos valores da média ( $\bar{X}$ ) de palavras e de nomes

---

<sup>161</sup> A estipulação destes critérios conduziu a uma recontagem dos dados analisados por Ferreira (2011:44-45) que integram o acervo empírico da presente investigação.

apurados por produção escrita, mas também da percentagem de ocorrências de nomes em relação ao total de palavras.

Informantes			Produções escritas			
LM	Nível QECRL	#	#	#de palavras	#de ocorrências de nomes	% de ocorrências de nomes em relação ao # de palavras
Espanhol	A1-A2	42	46	8 029	1 623	20,21
	B1-B2	24	26	5 853	1 204	20,57
	C1	9	11	3 479	692	19,89
$\Sigma$		<b>75</b>	<b>83</b>	<b>17 361</b>	<b>3 519</b>	20,27
$\bar{X}$				<b>209</b>	<b>42</b>	
Italiano	A1-A2	88	104	19 280	3 917	20,32
	B1-B2	35	40	9 353	1 750	18,71
	C1	12	12	9 249	2 049	22,15
$\Sigma$		<b>135</b>	<b>156</b>	<b>37 882</b>	<b>7 716</b>	20,37
$\bar{X}$				<b>242</b>	<b>49</b>	
Alemão	A1-A2	70	74	11 915	2 198	18,45
	B1-B2	48	60	15 598	3 007	19,28
	C1	16	20	5 563	1 096	19,70
$\Sigma$		<b>134</b>	<b>154</b>	<b>33 076</b>	<b>6 301</b>	19,05
$\bar{X}$				<b>216</b>	<b>41</b>	
Inglês	A1-A2	38	46	6 166	1 144	18,55
	B1-B2	42	54	16 624	3 355	20,18
	C1	20	29	6 505	1 447	22,24
$\Sigma$		<b>100</b>	<b>129</b>	<b>29 295</b>	<b>5 946</b>	20,30
$\bar{X}$				<b>227</b>	<b>46</b>	
Chinês	A1-A2	28	37	6 423	1 268	19,74
	B1-B2	25	34	7 705	1 578	20,48
	C1	9	10	2 093	442	21,12
$\Sigma$		<b>62</b>	<b>81</b>	<b>16 221</b>	<b>3 288</b>	20,27
$\bar{X}$				<b>200</b>	<b>41</b>	
$\Sigma$		<b>506</b>	<b>603</b>	<b>133 835</b>	<b>26 770</b>	20,00
$\bar{X}$				<b>221</b>	<b>44</b>	

**Quadro 3.6** – Distribuição do total e média de palavras e de ocorrências de nomes por texto produzido por LM e nível QECRL dos informantes

Identificado o número de ocorrências de itens nominais por produção escrita, procedeu-se, posteriormente, à contabilização das ocorrências de nomes que são núcleos de sintagmas nominais não reduzidos (SN), como *a casa bonita*, e das ocorrências de nomes em sintagmas nominais reduzidos, como, por exemplo, *beber café* e *fazer compras* (cf. Quadro 3.7). Portanto, e atendendo ao modo como se fará a análise dos dados empíricos, foram quantificadas as ocorrências de nomes em SN constituídos por especificadores (determinantes e quantificadores) e/ou modificadores (adjetivos) com função atributiva, por um lado, e, por outro, as ocorrências de sintagmas nominais que contêm apenas o nome como elemento nuclear.

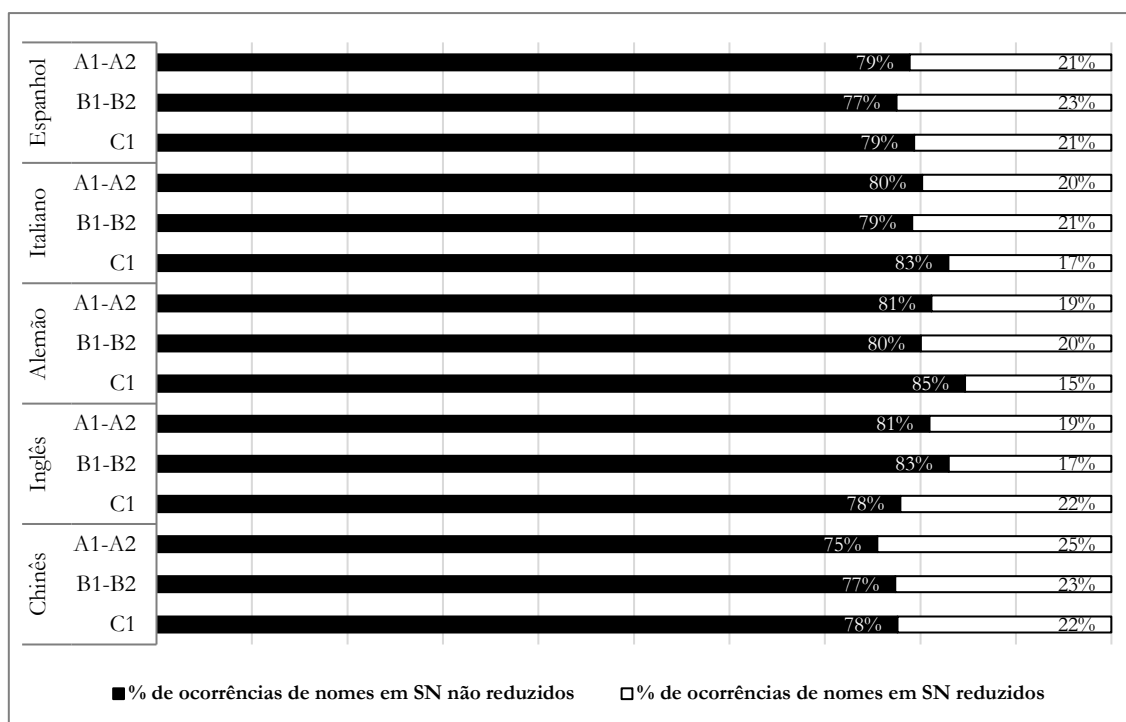
Assim sendo, o número total de ocorrências de nomes em sintagmas nominais não reduzidos (SN) servirá, posteriormente, de referência para o cálculo dos desvios apurados, uma vez que só estes contextos permitem inferir o valor de género nominal atribuído pelo aprendente e as formas dos respetivos constituintes sintáticos que, dentro do SN, concordam em género com o nome.

Informantes		Produções escritas	
LM	Nível QECRL	#de ocorrências de nomes em Sintagmas nominais não reduzidos	#de ocorrências de Sintagmas Nominais reduzidos
Espanhol	A1-A2	1 280	343
	B1-B2	933	271
	C1	549	143
<b>Σ</b>		<b>2 762</b>	<b>757</b>
Italiano	A1-A2	3 141	775
	B1-B2	1 385	365
	C1	1 699	350
<b>Σ</b>		<b>6 225</b>	<b>1 490</b>
Alemão	A1-A2	1 784	414
	B1-B2	2 407	600
	C1	928	168
<b>Σ</b>		<b>5 119</b>	<b>1 182</b>
Inglês	A1-A2	926	218
	B1-B2	2 783	572
	C1	1 127	320
<b>Σ</b>		<b>4 836</b>	<b>1 110</b>
Chinês	A1-A2	957	311
	B1-B2	1 221	357
	C1	343	99
<b>Σ</b>		<b>2 521</b>	<b>767</b>
<b>Σ</b>		<b>21 463</b>	<b>5 306</b>

**Quadro 3.7** – Distribuição no *corpus* de ocorrências de nomes em SN não reduzidos por oposição ao número de ocorrências de SN reduzidos

No Gráfico 3.10 é feito o registo dos dados apurados a partir da contabilização das ocorrências de nomes em SN, em valores percentuais (numa escala proporcional de 0 a 100%) e, como se pode aí verificar, ainda que apenas em termos quantitativos, a maioria dos segmentos da amostra de informantes por LM e nível QECRL regista um padrão idêntico relativamente às ocorrências de nomes em sintagmas nominais não reduzidos (cerca de 80%) face às ocorrências de nomes em sintagmas nominais reduzidos que atinge, aproximadamente, e na maior parte dos segmentos da amostra, os 20%.





**Gráfico 3.10** – Percentagem de ocorrências de nomes em SN não reduzidos por oposição à percentagem de ocorrências de nomes em SN reduzidos

O registo dos contextos de uso dos nomes em SN não reduzidos e de nomes em SN reduzidos pode ainda contribuir para a averiguação de eventuais estratégias adotadas pelos aprendentes ao longo do desenvolvimento do processo de aquisição/aprendizagem da AGN e da CNG em PLNM. Com efeito, e à semelhança do que é postulado por Martins (2015), as ocorrências de SN reduzidos poderão estar correlacionadas com uma eventual estratégia de evitação de uso de nomes em estruturas de concordância por parte dos diferentes grupos de aprendentes.

A título de exemplo, considerem-se os seguintes casos extraídos do *corpus*:

- (5) «Normalmente Ø **mulheres** têm mais responsabilidades (...) Depois estas funções, Ø **sociedade** tem (...)» (Inglês.B1-B2)
- (6) «(...) saí de casa e passei por Ø **rua**.» (Chinês.B1-B2)
- (7) «Eu não tenho Ø **irmão**.» (Chinês.A1-A2)
- (8) «(...) caminhamos por Ø **cidade** (...) fomos por Ø **rio** (...)» (Inglês.C1)
- (9) «Se este filtro falta, talvez Ø **crianças** sentiriam-se (...)» (Alemão.B1-B2)

Uma parte substancial destes casos pode, eventualmente, estar associada à dificuldade dos aprendentes na atribuição do valor de género ao núcleo do SN. Poder-se-á, então, supor que, ao não saber identificar corretamente o valor de género do nome, o

aprendente opta por utilizá-lo isoladamente, adotando uma estratégia que lhe permite, por conseguinte, ‘fugir’ ao desvio.

De facto, e como se viu no Capítulo 2, a inibição pode associar-se a diferentes fatores, como, por exemplo, o conhecimento linguístico prévio da LM (cf. Capítulo 2, Secção 2.3.2.2.; veja-se ainda o estudo de Lopes & Martins (2017)). Todavia, a relação entre a LM do aprendente e os fenómenos de inibição não é direta nem fácil de aferir e, dadas as características e diversidade dos dados apurados neste estudo, também não nos será possível averiguar, com rigor, a possível relação entre as ocorrências de SN reduzidos e o conhecimento linguístico prévio da LM do aprendente.

### 3.4.2. Contabilização dos desvios

No processo de seleção dos textos que integram a base empírica da presente investigação, consideraram-se os desvios de AGN e de CNG no SN constituídos por um núcleo nominal, e demais constituintes sintáticos (especificadores e/ou modificadores com função atributiva), que estabelecem com o nome relações de concordância, visto que esta estrutura permite a identificação dos valores atribuídos aos nomes pelos aprendentes tardios de PLNM.

Pelas razões evocadas em 3.2.2., não se considerou, neste estudo, a análise dos comportamentos desviantes que incidem sobre os adjetivos com função de predicativo do sujeito. Assim sendo, em função deste pressuposto, foram assinalados os desvios que incidiam, quer sobre a forma morfológica do nome, quer sobre as demais palavras que com ele devem concordar em género (especificadores e modificadores com função atributiva), conforme se pode observar nos seguintes exemplos recolhidos do *corpus*:

- (10) «\*Também procuro estudar, falar com os companheiros e ir **à**<sub>DETERMINANTE</sub> **cinema**.» (Espanhol.A1-A2)
- (11) «\*A universidade tem **o**<sub>DETERMINANTE</sub> **longo**<sub>ADJETIVO</sub> **híсторia** na europa, ainda no mundo.» (Chinês.B1-B2)

Assinale-se que, por vezes, um mesmo nome origina mais do que um desvio. Veja-se, por exemplo, os seguintes casos:

- (12) «\*Lembras-te **da**<sub>DETERMINANTE</sub> **corteja**<sub>NOME</sub>, aquela terça-feira de loucura?» (Inglês.B1-B2)
- (13) «\*(...) e foi **um**<sub>DETERMINANTE</sub> **viagem** **muito** **pesado**<sub>ADJETIVO</sub>» (Espanhol.A1-A2)

Em (12) o aprendente, para além de alterar a forma do nome masculino ‘cortejo’, atribuiu-lhe o valor de género feminino, visível pela seleção da forma do especificador. Já em (13), ao item feminino ‘viagem’ o aprendente atribuiu incorretamente o valor de género masculino, marcado, quer no especificador quer no modificador que, neste caso, é o adjetivo ‘pesado’. Por conseguinte, o número de ocorrências de desvios de AGN e de CNG no *corpus* é ligeiramente superior ao número de ocorrências de nomes afetados, conforme se constata no Quadro 3.8.

Informantes			Produções escritas		
LM	Nível QERCL	#	#	#de ocorrências de nomes com desvios	#de desvios
Espanhol	A1-A2	42	46	69	74
	B1-B2	24	26	31	34
	C1	9	11	13	13
$\Sigma$		75	83	113	121
Italiano	A1-A2	88	104	168	181
	B1-B2	35	40	73	78
	C1	12	12	15	15
$\Sigma$		135	156	256	274
Alemão	A1-A2	70	74	161	175
	B1-B2	48	60	97	103
	C1	16	20	31	35
$\Sigma$		134	154	288	313
Inglês	A1-A2	38	46	103	113
	B1-B2	42	54	143	147
	C1	20	29	59	61
$\Sigma$		100	129	305	321
Chinês	A1-A2	28	37	92	94
	B1-B2	25	34	101	105
	C1	9	10	13	13
$\Sigma$		62	81	206	212
$\Sigma$		506	603	1 169	1 241

**Quadro 3.8** – Distribuição, por segmentos da amostra (LM e nível QERCL), dos números absolutos de ocorrências de nomes com desvios e do número de desvios, apurados em cada texto

Como se verifica no Quadro 3.8, apurou-se no *corpus* um total de **1169 ocorrências de nomes** que deram origem a **1241 desvios** de AGN e de CNG.

A averiguação dos desvios e, sobretudo, dos constituintes sintáticos nos quais incidem as marcas de desvio dependeu da leitura global e da interpretação do texto em que se manifestam as diferentes ocorrências desviantes. Com efeito, só a partir do contexto é que é possível inferir o constituinte do SN no qual se manifestam as marcas de desvio de AGN e de CNG. Tal procedimento foi especialmente relevante para a identificação dos casos desviantes sobre os nomes, i.e., para a identificação dos itens nominais pretendidos pelos informantes. Neste sentido, considere-se o seguinte exemplo:

- (14) «\*Eu tive um emprego temporário com **duas amigos** na Bobadela.»  
(Chinês.B1-B2)

Em (14) poder-se-ia assumir que se trata de um desvio de atribuição do valor de género nominal marcado na forma do quantificador ‘duas’. Todavia, em outros momentos do texto é possível inferir que se trata, objetivamente, da seleção inadequada da forma do nome. Ou seja, o item lexical pretendido pelo informante é *amigas* e, por conseguinte, é sobre este nome que se verifica a marca de desvio.

Por fim, refira-se ainda que foi assinalado no *corpus* selecionado um conjunto variado de situações nas quais se verifica a ambivalência na seleção dos valores de género. De facto, em alguns SN observa-se que os aprendentes usam indiferenciadamente as formas masculinas e femininas dos seus vários elementos. Também Pinto (2015) verificara, nos seus dados, ocorrências semelhantes, referindo-se à ambivalência como uma “estratégia de “seleção mista” do género” (2015:105), adotada pelos aprendentes. Assim sendo, também assinalaremos na discussão dos resultados, e sempre que for pertinente, as ocorrências destes fenómenos.



### 3.4.3. Tipologia de desvios

Como se viu no Capítulo 1 deste trabalho (cf. Figura 1.5), na atribuição dos valores de género aos nomes em português atuam critérios de natureza semântica, nos casos em que o valor de género gramatical (masculino / feminino) corresponde ao género natural do referente (macho / fêmea), e critérios de natureza formal, observando-se, a este respeito, uma correlação, ainda que parcial, entre valores de género (masculino e feminino) e índices temáticos ‘-o’ e ‘-a’, respetivamente.

Deste modo, é possível identificar no português:

- (i) nomes cujo valor de género é dedutível, simultaneamente, a partir de critérios de natureza semântica [SEM +] e formal [FORM +], como, por exemplo: *o menino, a menina, a rapariga*;
- (ii) nomes cujo valor de género se infere a partir de critérios de natureza semântica [SEM +], mas não de critérios de natureza formal [FORM -]: *o homem, a mulher, o rapaz*;
- (iii) nomes cujo valor de género se deduz a partir de critérios formais [FORM +], mas não de critérios semânticos [SEM -], como é o caso de alguns nomes sobrecomuns (*a pessoa, a criança*) e de nomes com referentes não-sexuados, masculinos de tema em -o (*o carro*) e femininos de tema em -a (*a casa*); e
- (iv) itens nominais cujo valor de género não se deduz nem a partir de critérios semânticos [SEM -], nem de critérios formais [FORM -]. Nestes casos, somente se infere o valor de género do nome a partir de constituintes sintáticos (especificadores – determinantes e quantificadores – e adjetivos) que coocorrem com o nome. É o caso dos nomes masculinos de tema em -a (*o dia, o programa*), femininos de tema em -o (*a tribo, a (estação de) rádio*); de nomes masculinos e femininos de tema em -e (*o dente, a ponte*); de nomes masculinos e femininos de tema Ø (morfema zero, em nomes que terminam em consoante)<sup>162</sup> (*o mar, a catedral*); de nomes masculinos e femininos atemáticos (*o café, a canção*); de alguns epicenos (*a serpente*); de alguns nomes sobrecomuns (*o cônjuge*) e de nomes comuns de dois (*o/a turista; o/a estudante*).

---

<sup>162</sup> Mattoso Câmara Jr. (1994:87) identifica estes casos como nomes de tema em -e teórico.

O Quadro 3.9 apresenta uma breve síntese da atribuição de género nominal em português, com a distribuição dos itens relativamente ao tipo de critérios ou de indícios de atribuição dos valores de género.

Em função das características de atribuição valores de género nominal em português, distinguiram-se, na tipologia criada, quatro categorias de desvios, devidamente identificadas com um código numérico, de 1 a 4.

	<b>Indício semântico [SEM]</b> <i>género gramatical</i> (masculino / feminino) → <i>género natural</i> (macho / fêmea)	<b>Indício formal [FORM]</b> (não inteiramente fiável) Índice Temático -o = Masculino Índice Temático -a = Feminino	<b>Exemplos</b>
<b>Categoria 1</b>	<b>+</b>	<b>+</b>	<i>o menino, a menina</i>
<b>Categoria 2</b>	<b>+</b>	<b>-</b>	<i>o rapaz, o homem, a mulher</i>
<b>Categoria 3</b>	<b>-</b>	<b>+</b>	<i>o carro, a casa a criança, a pessoa</i>
<b>Categoria 4</b>	<b>-</b>	<b>-</b>	<i>o dia, a tribo o pente, a lente o papel, a catedral o coração, a mão  o/a motorista, o/a estudante o cônjuge</i>

**Quadro 3.9** – Proposta de caracterização do sistema de género nominal em português, atendendo aos indícios de atribuição de género nominal

A **categoria 1** integra os desvios que incidem sobre nomes animados cujo valor de género é corretamente dedutível a partir de critérios semânticos [SEM +] e formais [FORM +] (*o menino; a menina*); a **categoria 2** compreende os desvios relativos a nomes animados cujo valor de género se infere a partir de critérios semânticos [SEM +], mas não de critérios formais [FORM +] (*o rapaz, a mulher*); a **categoria 3** representa os comportamentos desviantes relativos a nomes não-animados, sendo que a atribuição dos valores de género não se deduz a partir de critérios semânticos [SEM -], mas de critérios formais [FORM +], podendo o aprendente recorrer à ‘pseudorregra’ de atribuição de valores de género: “se nome termina em -o é de género masculino; se termina em -a é de género feminino” (*o carro, a casa*). Por fim, a **categoria 4** integra os desvios que incidem sobre nomes cujo valor de género não se infere nem a partir de critérios semânticos [SEM -] nem de critérios formais [FORM -], nos quais somente se infere determinado valor de género a partir dos constituintes sintáticos que coocorrem com o nome: (*o mapa, a tribo, o pente, a ponte*).

Distinguiram-se ainda, para cada categoria, quatro subtipos de desvio estipulados em função dos constituintes do SN afetados. Assim, identificaram-se os casos em que o desvio incide nos determinantes – quer constituintes preposicionados, quer constituintes não preposicionados – (det.); nos quantificadores (quant.); nos adjetivos com função



atributiva – (adj.) e nos nomes (nom.)<sup>163</sup>. Note-se que nos casos em que um nome origina mais do que um desvio, são identificados os diferentes constituintes afetados. O Quadro 3.10 apresenta alguns exemplos retirados do *corpus* distribuídos por categoria e constituintes do SN nos quais se evidenciam as marcas de desvio<sup>164</sup>.

Categorias	Constituintes afetados		Exemplos
Categoria 1 [SEM + FORM +]	Especificadores	Determinantes (det.)	«*(...) com meus familiares, meus pais e <b>meus amigos</b> » (Italiano.A1-A2) «* Gozar muito tempo com <b>meu filha</b> » (Espanho.A1-A2) «* Um <b>destes “amigas”</b> (teve) tinha chegado de Busapest.» (Alemão.A1-A2)
		Quantificadores (quant.)	«* Como tenho <b>algumas amigos</b> que moram em Lisboa» (Italiano.B1-B2) «* Eu tive um emprego temporário com <b>duas amigos</b> » (Chinês.B1-B2)
	Adjetivos (adj.)		«*Ela teve uma <b>madrasta</b> muito <b>mau</b> » (Alemão.A1-A2)
	Nomes (nom.)		-
Categoria 2 [SEM + FORM -]	Especificadores	Determinantes (det.)	«* <b>As rapazes</b> e as raparigas do curso são simpáticos também» (Italiano.A1-A2) «*(...) ela tinha de fazer um funeral em sua coração para <b>o seu irmã</b> » (Chinês.B1-B2)
		Quantificadores (quant.)	«*Temos uma casa grande (...) duas irmãs e <b>duas irmãos</b> » (Inglês.A1-A2)
	Adjetivos (adj.)		«*Sou a <b>irmão</b> mais <b>nova</b> » (Espanhol.A1-A2)
	Nomes (nom.)		-
Categoria 3 [SEM - FORM +]	Especificadores	Determinantes (det.)	«*Tu ainda lembras <b>da nossa tempo</b> junto» (Alemão.B1-B2) «*(...) nós precisamos ir <b>ao praia</b> » (Chinês.A1-A2)
		Quantificadores (quant.)	«*(...) ela lê <b>muitos historias</b> de amor» (Italiano.C1) «* (...) Na minha casa temos <b>dois casas de banho</b> » (Espanhol.A1-A2)
	Adjetivos (adj.)		«*Bões filmes passam em <b>horários estranhas</b> » (Alemão.A1-A2) «*Para o <b>próxima ano</b> gostaria imenso de voltar» (Alemão.C1)
	Nomes (nom.)		«*Em 2003 o governo lançou o <b>proposto</b> de introduzir cartões» (Inglês.C1)
Categoria 4 [SEM - FORM +]	Especificadores	Determinantes (det.)	«* <b>O passagem</b> do ano foi menos relaxante para mim» (Italiano.A1-A2) «*(...) chegaram no Lisboa <b>nas dias</b> passadas» (Alemão.A1-A2)
		Quantificadores (quant.)	«*Há peixes de <b>muitos (cores) cores</b> » (Inglês.A1-A2) «*Eu não tinha <b>nenhuma guarda-chuva</b> » (Chinês.B1-B2)
	Adjetivos (adj.)		«*(...) chegaram no Lisboa <b>nas dias passadas</b> » (Alemão.A1-A2) «*A Suíça é um <b>país montanhosa</b> » (Inglês.B1-B2)
	Nomes (nom.)		«*Andigo via férias vir mito <b>nevo</b> » (Chinês.A1-A2) «*Eu adoro as <b>pastelas</b> de nata» (Italiano.A1-A2)

Quadro 3.10 – Tipologia de desvios com exemplos extraídos do *corpus*

<sup>163</sup> Note-se que os casos que afetam a forma morfológica do nome foram considerados desvios sempre que a formatação do nome resulta no uso de um valor de género gramatical oposto ao da forma original, como por exemplo, *\*uma feriado* em vez de *um feriado*; ou *\*umas passeias* em vez de *uns passeios*. Por conseguinte, estes casos são identificados, quer como desvios relativos à forma morfológica do nome, quer como desvios relativos ao uso do especificador e/ou do adjetivo que coocorre(m) com o nome dentro do SN.

<sup>164</sup> Todos os dados analisados encontram-se disponíveis no Anexo II do Volume 2 do presente trabalho.

Assim, é em função da tipologia de desvios criada que se vai proceder no Capítulo 4 ao tratamento, análise e discussão dos dados empíricos.

# Capítulo 4

## Resultados e Discussão

---

### 4.1. Introdução

O presente capítulo é dedicado à apresentação e discussão dos resultados obtidos a partir da análise dos desvios de atribuição de valores de género aos nomes (AGN) e de concordância nominal em género (CNG), no sintagma nominal (SN), detetados no conjunto de produções escritas que compõem o *corpus* selecionado para este estudo.

Este capítulo encontra-se organizado em duas partes essenciais. Referem-se, em primeiro lugar, os resultados globais (Secção 4.2.1.), passando-se, em seguida, aos resultados apurados por variável de análise (4.2.2.), LM dos informantes (4.2.2.1.) e nível QECRL da turma frequentada (4.2.2.2.). Por fim, cruzam-se as duas variáveis de análise, o que permite, assim, obter uma visão conjunta dos resultados globais quantitativos apurados (4.2.2.3.).

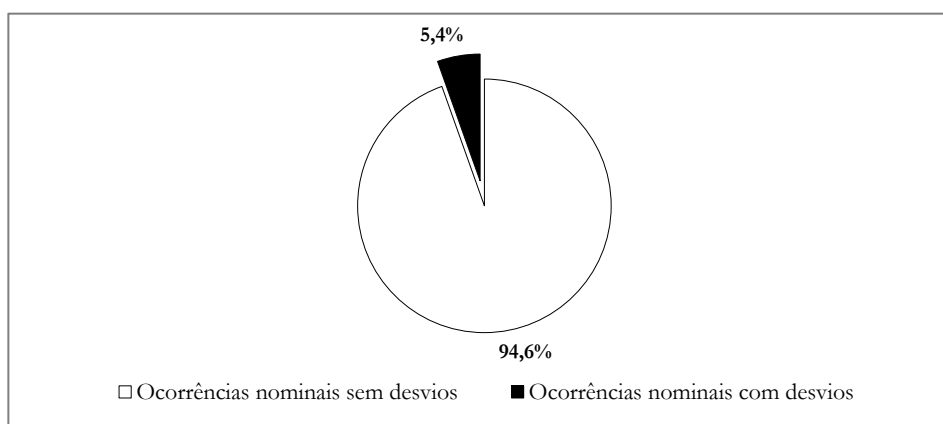
Na segunda parte deste capítulo (Secção 4.2.3.), fornecem-se os dados relativos à análise dos comportamentos desviantes tendo em conta a tipologia de desvios criada para esta investigação (veja-se o Capítulo 3, Secção 3.4.3.). A apresentação detalhada dos resultados obtidos (Secção 4.2.3.1.) abrange não só a referência dos itens nominais com maior incidência de desvios por informantes, distribuídos por LM e respetivo nível QECRL, mas também o registo dos constituintes sintáticos do SN nos quais se evidenciam as marcas de desvio.

## 4.2. Resultados

### 4.2.1. Resultados globais

A base empírica do presente estudo é composta por **603 textos** escritos por **506 informantes**. No conjunto global deste acervo textual, verificou-se que 1169 ocorrências de nomes originaram 1241 desvios de AGN e de CNG (cf. Capítulo 3, Quadro 3.8).

Tendo em conta o número absoluto de ocorrências nominais (cf. Capítulo 3, Quadro 3.7), verificou-se que, em valores percentuais, os desvios de AGN e de CNG afetam 5,4% de todas as ocorrências nominais (cf. Gráfico 4.1).



**Gráfico 4.1** – Distribuição das ocorrências nominais com desvios de AGN e de CNG, por oposição às ocorrências nominais sem desvios

Estes dados revelam, assim, que, durante o processo de elaboração dos textos, uma grande parte dos sujeitos que compõem a amostra analisada escolhe adequadamente os valores de género dos nomes que usam.

## 4.2.2. Resultados por variáveis de análise

### 4.2.2.1. Distribuição dos resultados por LM

Para averiguar os comportamentos desviantes de aprendentes tardios de PLNM que são falantes nativos de espanhol, italiano, alemão, inglês e chinês, procedeu-se à quantificação dos desvios por segmentos de amostra. Assim, a partir da totalidade de ocorrências nominais produzidas em SN não reduzidos pelos aprendentes das diferentes LM (Quadro 4.1), calculou-se a percentagem relativa de desvios de AGN e de CNG (cf. Gráfico 4.2).

LM	Espanhol	Italiano	Alemão	Inglês	Chinês	$\Sigma$
# de ocorrências de nomes	2 762	6 225	5 119	4 836	2 521	21 463
# de desvios (AGN e CNG)	121	274	313	321	212	1 241

Quadro 4.1 – Distribuição do número total de ocorrências de nomes e de desvios de AGN e de CNG por LM

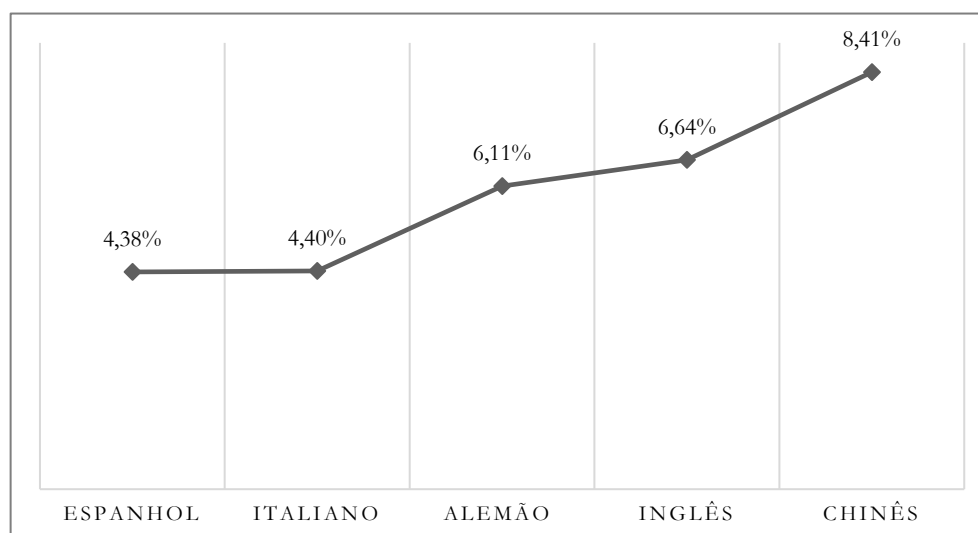


Gráfico 4.2 – Distribuição das percentagens relativas de desvios de AGN e de CNG por LM

Segundo os dados apresentados no Gráfico 4.2, regista-se uma maior percentagem relativa de desvios de AGN e de CNG nas produções escritas por falantes nativos de língua chinesa relativamente à registada nos textos dos falantes nativos das demais línguas seleccionadas e, em particular, das línguas românicas. Com efeito, a percentagem relativa de desvios apurada nas produções escritas por aprendentes falantes nativos de chinês (8,41%) é quase o dobro da registada nos textos de falantes aprendentes nativos de italiano, que é de 4,40%, e nos textos produzidos por falantes de LM espanhola (4,38%). Para além disso, entre estes dois últimos segmentos da amostra de informantes por LM, as

percentagens relativas de desvio são muito próximas. Também num patamar aproximado de valores de percentagens relativas de desvios situam-se os informantes cuja LM é uma língua germânica, sendo que os desvios de AGN e de CNG produzidos por aprendentes que são falantes nativos de inglês apresentam uma percentagem relativa ligeiramente mais elevada (6,64%) face à registada entre o segmento da amostra de informantes falantes nativos de alemão (6,11%).

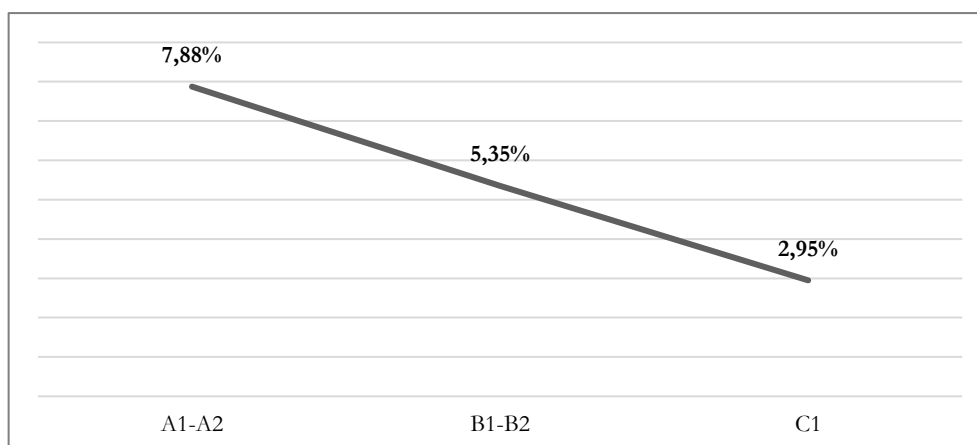
Parece, pois, haver uma tendência para um aumento gradual da percentagem relativa de desvios de AGN e de CNG em função da distância entre a configuração do sistema de classificação nominal das LM dos aprendentes relativamente à LA de aprendizagem, o português, i.e., quanto mais distantes são as LM (chinês, inglês, alemão e línguas românicas) face ao sistema de atribuição de valores de género nominal do português, maior é a proporção dos desvios. Ora, estes dados parecem indiciar que o facto de a LM do aprendente possuir um sistema de atribuição de género nominal com características próximas das do sistema do português favorece a assimilação da categoria de género gramatical.

#### 4.2.2.2. Distribuição dos resultados por nível QECRL

Considere-se, agora, a distribuição dos desvios em função de outra variável de análise: o nível QECRL da turma frequentada pelos informantes. Assim, a partir do número total de ocorrências nominais em SN não reduzidos, produzidas por segmentos de amostra – de A1 a C1 – (cf. Quadro 4.2), calculou-se a percentagem relativa de desvios de AGN e de CNG (cf. Gráfico 4.3).

Nível QECRL	A1-A2	B1-B2	C1	$\Sigma$
# de ocorrências de nomes	8 088	8 729	4 646	21 463
# de desvios (AGN e CNG)	637	467	137	1 241

**Quadro 4.2** – Distribuição do número total de ocorrências de nomes e de desvios de AGN e de CNG por nível QECRL



**Gráfico 4.3** – Distribuição das percentagens relativas de desvios de AGN e de CNG por nível QECRL.

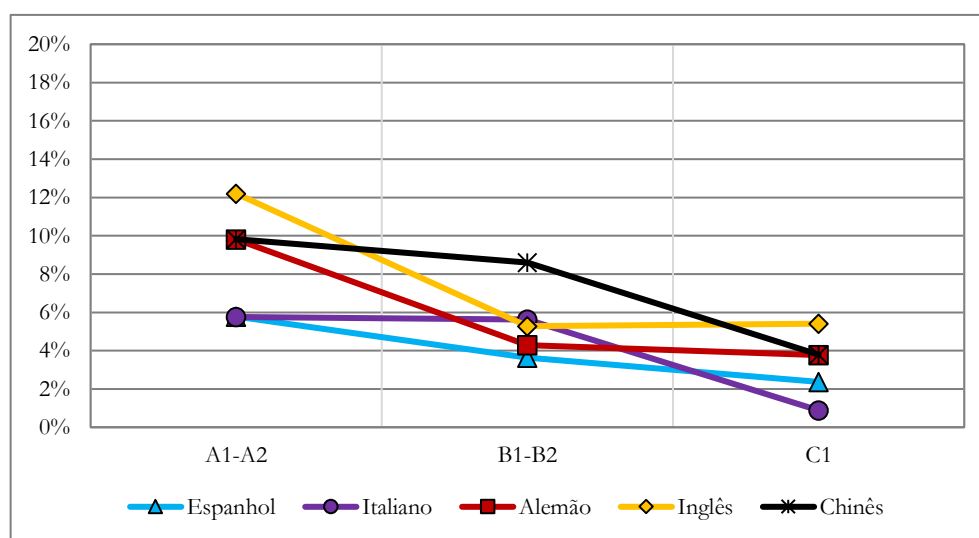
A partir dos dados apresentados no Gráfico 4.3, é possível constatar que há, no *corpus*, uma maior incidência de desvios de AGN e de CNG nos níveis A1 e A2, havendo uma progressiva descida das percentagens relativas de desvios nos níveis mais avançados, sendo que os aprendentes de nível C1 registam uma percentagem relativa de desvios consideravelmente baixa (2,95%). Estes dados corroboram, assim, as conclusões de outros estudos (Ferreira 2011; Mariotto & Lourenço-Gomes 2013; Pinto 2015; Martins 2015), nos quais também se regista uma melhoria significativa no desempenho dos aprendentes à medida que progredem na aprendizagem do português. No entanto, é certo que, mesmo em C1, os desvios de AGN e de CNG nunca chegam a ser totalmente erradicados das produções textuais dos aprendentes.

#### 4.2.2.3. Distribuição dos resultados por LM e nível QECRL

Procede-se, em seguida, ao tratamento dos dados, tendo em conta a variável da LM e o nível QECRL da turma frequentada pelos aprendentes. Tomando como referência o número absoluto de ocorrências nominais em SN não reduzidos, produzido por segmentos de amostra (cf. Quadro 4.3), calculou-se a percentagem relativa de desvios de AGN e de CNG (Gráfico 4.4). No Quadro 4.3 apresentam-se também os valores da média ( $\bar{X}$ ) de desvios produzidos por texto em cada segmento da amostra, já que estes valores também servirão de referência para a análise dos dados.

Informantes		Produções escritas		Desvios		
LM	Nível QERCL	#	# de ocorrências nominais	# de ocorrências	$\bar{X}$	%
Espanhol	A1-A2	46	1 280	74	1,6	5,78
	B1-B2	26	933	34	1,3	3,64
	C1	11	549	13	1,2	2,37
Italiano	A1-A2	104	3 141	181	1,7	5,76
	B1-B2	40	1 385	78	2,0	5,63
	C1	12	1 699	15	1,3	0,88
Alemão	A1-A2	74	1 784	175	2,4	9,81
	B1-B2	60	2 407	103	1,7	4,28
	C1	20	928	35	1,8	3,77
Inglês	A1-A2	46	926	113	2,5	12,20
	B1-B2	54	2 783	147	2,7	5,28
	C1	29	1 127	61	2,1	5,41
Chinês	A1-A2	37	957	94	2,5	9,82
	B1-B2	34	1 221	105	3,1	8,60
	C1	10	343	13	1,3	3,79
$\Sigma$		603	21 463	1 241	2,1	5,78

**Quadro 4.3** – Distribuição do número total de desvios de AGN e de CNG, e respetivas percentagens relativas, por segmentos da amostra (LM e nível QERCL)



**Gráfico 4.4** – Percentagens relativas de desvios de AGN e de CNG por segmentos da amostra (LM e nível QERCL)

Tendo em conta os dados representados no Gráfico 4.4, constata-se que, em todos os segmentos de informantes por LM, os níveis A1-A2 registam uma maior percentagem relativa de desvios. Na passagem para os níveis B1-B2 e C1, a percentagem relativa diminui consideravelmente, sobretudo nas produções escritas pelos aprendentes falantes nativos de inglês e de alemão. De facto, nos textos de falantes nativos de inglês, regista-se uma descida



acentuada do peso relativo de desvios nos níveis A1-A2 para os níveis B1-B2, de 12,2% para 5,3%. Já no nível C1, a percentagem relativa dos comportamentos desviantes neste segmento da amostra mantém-se nos mesmos valores, pese embora se registre uma ligeira subida, já que ronda os 5,4%.

Entre as produções escritas por falantes nativos de alemão observa-se, igualmente, uma descida acentuada da percentagem relativa de desvios na passagem dos níveis A1-A2 para B1-B2, de 9,8% para 4,3%. A tendência decrescente que aqui se assinala mantém-se, embora não seja tão acentuada, em C1, no qual se apurou, tendo em conta o total de ocorrências de nomes produzidos, cerca de 3,8% de percentagem relativa de desvios de AGN e de CNG.

Quanto aos falantes nativos de chinês, observa-se que, apesar de os níveis A1-A2 e B1-B2 registarem percentagens relativas consideravelmente próximas (cerca de 9,8% e 8,6%, respetivamente), é no nível C1 que se regista uma descida assinalável dos comportamentos desviantes face ao número absoluto de ocorrências nominais neste segmento, 3,8%, sendo este valor muito próximo do registado nas produções de aprendentes de LM alemã (cf. Gráfico 4.4). Portanto, neste segmento da amostra, parece haver indícios de que, mesmo não tendo a categoria de género gramatical representada na gramática da sua LM, a maioria dos informantes do nível C1 consegue, no momento da produção dos nomes, atribuir-lhes adequadamente o valor de género.

No que concerne aos resultados apurados entre os falantes nativos de línguas românicas e, em particular, os falantes nativos de espanhol, verifica-se que, apesar de logo nos níveis A1-A2 se registarem índices de desvio baixos (5,8%), nos restantes níveis se verifica apenas uma ligeira descida deste valor (3,7% em B1-B2), atingindo-se, no nível C1, uma percentagem relativa de desvios que atinge aproximadamente os 2,4%, ou seja, quase metade do valor percentual apurado em A1-A2. Estes dados parecem, assim, sugerir uma certa estabilização na aprendizagem neste segmento da amostra, i.e., um efeito *plateau* no processo de assimilação da AGN e de CNG, já que, mesmo em fases posteriores do desenvolvimento das suas interlínguas, os falantes de LM espanhola não apresentam um desempenho consideravelmente melhor daquele que se regista, por exemplo, nos falantes nativos de chinês, idioma que (cf. Capítulo 1, Secção 1.4.5.) não apresenta qualquer sistema de atribuição de género nominal nem admite a realização de relações de concordância sintática.

Já no segmento da amostra de informantes que são falantes nativos de italiano, verifica-se, no Gráfico 4.4, que há uma ligeira descida da percentagem relativa de desvios

da passagem dos níveis A1-A2 para B1-B2, de 5,8% para 5,6%. Observa-se ainda que, nos níveis B1-B2, se regista um maior afastamento das percentagens relativas de desvios registadas entre o segmento da amostra de informantes que são falantes nativos de italiano e o segmento da amostra de informantes que são falantes nativos de espanhol. Com efeito, a percentagem relativa de desvios nos níveis B1-B2 no segmento da amostra de informantes de LM espanhola é inferior comparativamente à registada nestes níveis no segmento de aprendentes de LM italiana. Contudo, no nível C1, esta tendência inverte-se, já que nas produções escritas pelos falantes nativos de italiano se regista uma clara descida da percentagem relativa de desvios de AGN e de CNG (0,9%), ao passo que nos textos de falantes nativos de espanhol se regista uma certa estabilização do número de desvios, não havendo uma descida significativa da percentagem relativa de desvios face à registada nas fases iniciais da aprendizagem do português neste segmento.

Ainda que estes resultados nos permitam obter uma perspetiva global e de natureza quantitativa dos dados apurados, havendo, por isso, a necessidade de uma análise qualitativa mais fina, que será empreendida nas próximas Secções, as tendências aqui assinaladas colocam, desde já, em evidência alguns aspetos relativos à progressão da aquisição/aprendizagem da AGN e de CNG por diferentes grupos de informantes, que convém assinalar.

Em primeiro lugar, verifica-se que os aprendentes que são falantes nativos de idiomas românicos apresentam um desempenho globalmente mais satisfatório face ao que se regista nos aprendentes falantes nativos de idiomas que não pertencem à mesma família linguística do português. Assim sendo, poder-se-á admitir que uma maior proximidade dos sistemas linguísticos contribui positivamente, pelo menos no início da aprendizagem, para a assimilação dos valores de género dos nomes em português. Porém, nota-se que, embora no início da aprendizagem haja um reduzido número de desvios nos textos dos informantes de LM românicas, ao longo do desenvolvimento das interlínguas destes aprendentes, essa aparente ‘vantagem’ parece desvanecer-se, dando lugar a um certo efeito de estabilização da aprendizagem, já que os índices de desvio no nível C1 apurados nos segmentos de informantes de LM românicas são muito próximos dos registados nos restantes segmentos da amostra de informantes por LM (alemão, inglês e chinês). Tal efeito de estabilização é visível, sobretudo, se se tiver em conta os resultados apurados entre os informantes que são falantes nativos de espanhol a frequentar turmas de C1 (cf. Gráfico 4.4). Contudo, é preciso também assinalar que a proximidade das estruturas dos idiomas não conduz necessariamente a um efeito de estabilização da aprendizagem, já que, no segmento da

amostra de aprendentes de LM italiana, se regista uma clara melhoria do desempenho no nível C1 relativamente ao registado nos restantes níveis. Por outras palavras, estes dados demonstram que provavelmente outros fatores, para além da proximidade estrutural das línguas, poderão contribuir para o efeito *plateau* registado no segmento de falantes nativos de espanhol.

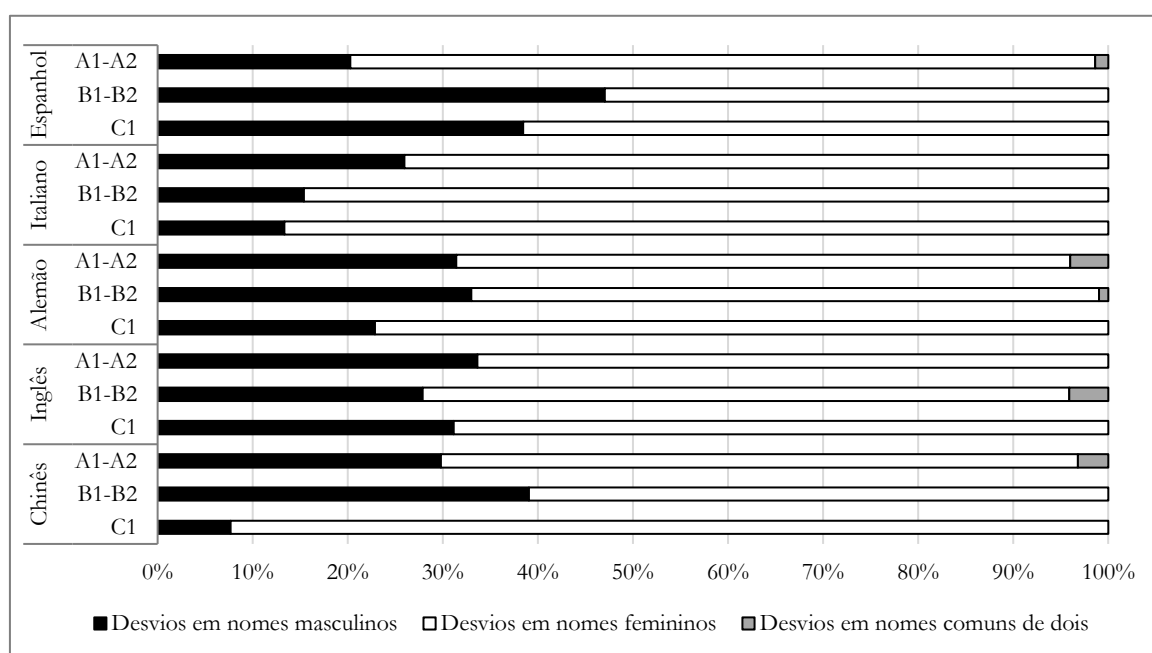
Os dados apurados indiciam ainda que, embora os falantes nativos de línguas sem sistemas de atribuição de género, como o chinês e o inglês, apresentem, inicialmente, maiores dificuldades na AGN e na CNG, ao longo do seu desenvolvimento (inter)linguístico conseguem registar uma considerável melhoria do seu desempenho, aproximando-se dos valores percentuais de desvios registados nos outros grupos de LM. Ou seja, é possível inferir, nestes aprendentes, uma certa relação entre o desenvolvimento da aprendizagem do português e a diminuição dos comportamentos desviantes de AGN e de CNG, podendo-se admitir que a assimilação desta estrutura parece ser muito favorecida com a progressão do conhecimento da língua.

Relativamente aos dados do desempenho dos falantes nativos de alemão, os níveis A1-A2 registam índices de desvios consideravelmente altos. Por conseguinte, poder-se-á admitir que o facto de este segmento da amostra ter na sua LM a categoria gramatical de género, embora com particularidades distintas do sistema de atribuição de género nominal do português, não favorece, pelo menos inicialmente, o processo de aquisição/aprendizagem dos valores de género nominal. Todavia, é de assinalar que, na passagem de A1-A2 para B1-B2, se regista uma diminuição considerável dos comportamentos desviantes, com valores percentuais muito próximos dos registados nas produções textuais de outros segmentos de LM. Ou seja, regista-se também neste segmento uma relação positiva entre o desenvolvimento da aprendizagem e a assimilação desta estrutura gramatical.

Após se identificarem as tendências de desvios por segmentos da amostra de informantes atendendo à LM, procurou-se ainda determinar, no conjunto global de comportamentos desviantes assinalado, se estes resultam preferencialmente do uso generalizado de um valor de género em detrimento de outro valor. Com efeito, de acordo com os dados apurados em outros trabalhos já referidos na presente monografia (Martins 2015; Mariotto & Lourenço Gomes 2013), a maior parte dos desvios resulta da adoção, pelo aprendente, de uma estratégia de atribuição por *default* da forma masculina. Por outras palavras, verifica-se que, aquando da atribuição do valor ao género do item nominal, os

aprendentes tendem a selecionar, preferencialmente, nos diferentes elementos concordantes com o nome, o valor de género masculino, a forma não-marcada (*default*). Por este motivo, espera-se que haja uma maior incidência de desvios em formas nominais de género feminino.

Com o intuito de averiguar, no *corpus* selecionado, uma possível estratégia de atribuição de género por *default*, procedeu-se à organização do conjunto total de desvios apurados por segmentos da amostra tendo em conta o valor de género da forma nominal afetada. O Gráfico 4.5 ilustra, em valores percentuais, a proporção dos desvios por valor de género (masculino e feminino) do nome afetado.



**Gráfico 4.5** – Distribuição, em valores percentuais (numa escala proporcional de 0 a 100), dos desvios em função do valor de género do item nominal afetado

A leitura do Gráfico 4.5 permite constatar que em todos segmentos da amostra se regista o mesmo padrão, i.e., a maior parte dos desvios incide sobre formas nominais femininas, havendo, portanto, uma preferência pela atribuição por *default* do valor de género masculino. Observa-se, então, em todos os segmentos, o uso da forma masculina dos especificadores e adjetivos em SN com um nome feminino, como atestam os seguintes exemplos:

- (1) «\* (...) **meu morada** fica perto da universidad (...)» (Espanhol.A1-A2)
- (2) «\*A decisão de morrer é a decisão de fugir **ao dor**.» (Italiano.B1-B2)
- (3) «\*Há **dois semanas** que foi Queima das fitas, foi fantástica!» (Inglês.B1-B2)

- (4) «\*(...) por isso, eu faço **este actividade** nos meus tempos livres cada vez mais frequente.» (Chinês.A1-A2)
- (5) «\*(...) na cidade passa por ter mais escolhas de lazer, (...) /os/ bares, /os/ parques, /os/ **discotecas** já ocupam um lugar» (Chinês.C1)

Em contrapartida, são menos frequentes os desvios que afetam itens nominais de género masculino como, por exemplo:

- (6) «\*(...) vamos a Las Vegas a **uma Casino** » (Alemão.A1-A2)
- (7) «\***Todas as dias** de noite com os meus novos amigos espanhóis, que são muito simpáticos» (Italiano.A1-A2)
- (8) «Eu tenho **muitas problemas** com as comidas.» (Italiano.A1-A2)

A marcação do valor de género feminino nos demais constituintes que compõem o SN cujo núcleo é um nome masculino regista-se, sobretudo, em nomes como *dia*, *sistema*, *problema*, ou seja, itens masculinos com índice temático *-a* e nos quais a aplicação da ‘pseudoregra’ de atribuição de valor de género (“se nome termina em *-o* é de género masculino, se nome termina em *-a* é feminino”) resulta em desvio. Ora, poder-se-á, então, supor que os aprendentes assumem estes nomes como sendo femininos e é por esse motivo que recorrem à forma feminina dos demais elementos concordantes. Retomaremos a discussão destes dados mais adiante, na Secção 4.3.1.4.

Os desvios que incidem sobre formas nominais comuns de dois são muito menos frequentes<sup>165</sup> e resultam, sobretudo, da opção do uso do valor de género feminino em contextos nos quais o masculino, pela leitura genérica que lhe está tipicamente associada, seria o valor de género adequado:

- (9) «?esperei por **as clientes** (Chinês.A1-A2)
- (10) «?Há **muitas turistas** no Centro de Londres (Inglês.B1-B2)».

---

<sup>165</sup> De facto, nos nomes comuns de dois valores de género admite-se, como a própria designação indica, a associação aos dois valores de género disponíveis em português (masculino e feminino): *o/a estudante*, *o/a colega*). Por conseguinte, facilmente se compreendem os índices de desvio muito baixos neste tipo de itens. Assim sendo, nestes casos considerou-se desvio quando, num contexto de referência genérica, o aprendente optou pela utilização do feminino ao contrário do masculino que, prototipicamente, é o valor de género que permite a leitura genérica. Por exemplo, para designar, genericamente, um grupo de estudantes diz-se *os estudantes*, já que a opção pelo feminino, *as estudantes*, não permite essa leitura, antes pressupondo que o grupo a que se refere o falante é apenas composto por *estudantes* do sexo feminino.

Há ainda a assinalar, entre os nomes comuns de dois, a ocorrência de alguns itens nos quais se observa, por parte do informante, uma certa hesitação na seleção do valor de género das formas dos elementos concordantes com o nome, como atesta o seguinte exemplo retirado de uma produção escrita por um falante nativo de chinês de A1-A2:

(11) «\*Eu com (...) os **minhas colegas** (Chinês.A1-A2).».

Com efeito, e como já referido no Capítulo 3, Secção 3.4.2., no *corpus* em estudo verificam-se ocorrências de casos de ambivalência de atribuição do valor de género nominal, i.e., situações nas quais os aprendentes usam indiferenciadamente as formas masculinas e femininas dos vários elementos que integram o SN. Estes casos podem resultar da dificuldade, por parte do aprendente, em atribuir adequadamente o valor de género ao núcleo do SN. Nos exemplos seguintes, poder-se-á verificar que, apesar da seleção da forma concordante do especificador, o adjetivo não apresenta marcas do valor de género adequado:

(12) «\*Eu acho que tenho uma **alimentação adequado**» (Inglês. C1)

(13) «\* (...) ele viveu e vive fora do seu **país materna**» (Alemão.C1)

(14) «\*(...) No ano pasado descubriu a **comida japonês**» (Italiano.B1-B2)

Há ainda casos em que, em dois especificadores, apenas um se apresenta na forma adequada quanto ao valor de género, tal como atesta o seguinte exemplo:

(15) «\*Na expectativa de que a minha candidatura suscite interesse da **vosso parte**.» (Espanhol.B1-B2)

### 4.2.3. Resultados apurados em função da tipologia de desvios

Como se verificou no capítulo anterior (Capítulo 3, Secção 3.4.3.), para a análise dos desvios de atribuição dos valores de género aos nomes e de concordância nominal em género, foi criada uma tipologia, organizada em função dos indícios de atribuição de género nominal do português (cf. Capítulo 1, Secção 1.3.3.), com o intuito de averiguar o grau de sensibilidade dos diferentes grupos de informantes relativamente aos indícios semânticos e/ou formais de atribuição dos valores de género aos nomes em português. A tipologia de desvios criada está organizada em quatro categorias, por sua vez subdivididas em quatro subcategorias cada, que correspondem aos constituintes nos quais se manifestam as marcas de desvio (cf. Capítulo 3, Quadro 3.11).

Recorde-se que a categoria 1 abrange os desvios sobre nomes cujo valor de género é inferível, simultaneamente, a partir de indícios semânticos [SEM +] e formais [FORM +], como, por exemplo, ‘*o menino, a menina*’.

A categoria 2 integra os desvios sobre nomes em que o valor de género é dedutível a partir de critérios semânticos [SEM +], sendo, no entanto, possível através de critérios formais [FORM -], tal como em ‘*o homem*’ e ‘*a mulher*’.

Na categoria 3, encontram-se os desvios em nomes cuja atribuição de género não se deduz a partir de critérios semânticos [SEM -], mas de critérios formais [FORM +]. Integram esta categoria os desvios relativos a nomes sexuais, cujo valor de género não corresponde, necessariamente, ao sexo da entidade designada: ‘*o indivíduo, a pessoa*’, bem como os desvios sobre formas nominais sem referente sexual em que observa uma correspondência entre os valores de género masculino e feminino e respetivos índices temáticos *-o* e *-a*: ‘*o carro*’ e ‘*a casa*’.

Por fim, a categoria 4 abrange os desvios relativos a formas nominais cujo valor de género não se infere nem a partir de critérios semânticos [SEM -] nem de critérios formais [FORM -]. Encontram-se aqui os comportamentos desviantes que afetam nomes com o traço semântico [+ sexual], dado que em português existem formas nominais com referentes sexuais cujo valor de género gramatical não corresponde, necessariamente, ao género natural da entidade designada, como é o caso dos nomes epicenos (*o elefante*) e de alguns nomes sobrecomuns (*o cônjuge*). Integram ainda esta categoria os desvios relativos a nomes comuns de dois, i.e., a nomes que apresentam uma única forma, quer para o masculino, quer para o feminino (*o/a estudante*). Fazem também parte desta categoria os desvios relativos a nomes não sexuais masculinos terminados em *-a* (*o programa*) e femininos

terminados em *-o* (*a tribo*), para além dos casos desviantes relativos a nomes masculinos e femininos de outras classes temáticas (*o pente, a ponte, o coração, a paisagem*) e nos quais não é possível estabelecer uma associação entre a forma do nome e respetivo valor de género nominal.

Como se disse, para cada categoria de desvios foram ainda criadas subcategorias, em função dos constituintes do sintagma nominal afetados. Assim, identificaram-se os casos em que o desvio incide:

- (i) nos **especificadores – determinantes e quantificadores**;
- (ii) nos **modificadores – adjetivos** (com função atributiva<sup>166</sup>); e
- (iii) no **nome**.

No Quadro 4.4 apresenta-se a distribuição dos **1241 desvios** de atribuição de género nominal detetados no *corpus* em estudo, em função das categorias criadas. Para além dos valores absolutos, apresentam-se igualmente os valores da média ( $\bar{X}$ ) de desvios em cada categoria apurados por texto. No Gráfico 4.6, apresenta-se a distribuição dos desvios em valores percentuais, a fim de se ter uma ideia da proporção de desvios apurada por categoria.

Categorias de desvio	# de ocorrências de desvios	$\bar{X}$ de desvios
Categoria 1 [SEM +, FORM +]	35	1,1
Categoria 2 [SEM +, FORM -]	23	1,4
Categoria 3 [SEM -, FORM +]	435	1,5
Categoria 4 [SEM -, FORM -]	748	1,8
$\Sigma$	1 241	2,1

Quadro 4.4 – Número absoluto de desvios, e dos valores da média, por categorias de desvio

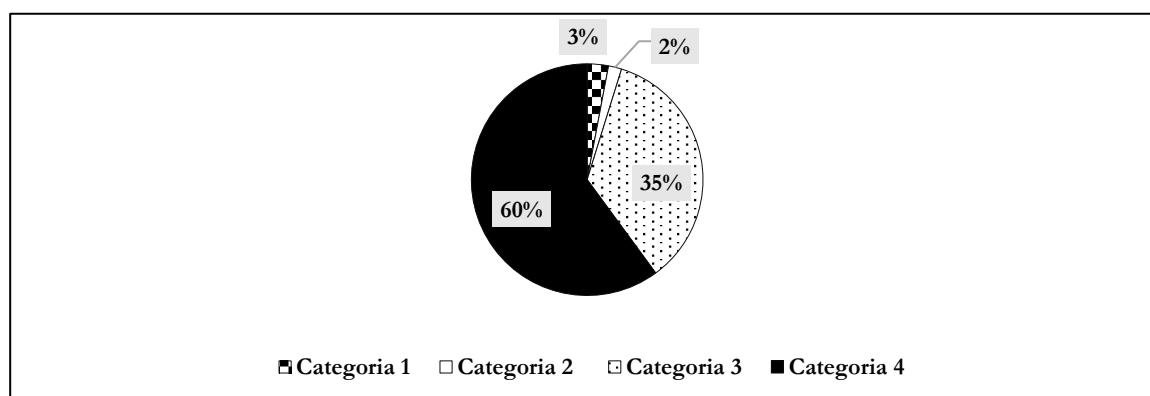


Gráfico 4.6 – Percentagens de desvios, distribuídas por categorias

<sup>166</sup> Para compreender a razão pela qual apenas se considerou para a análise empírica os desvios de AGN em adjetivos em uso atributivo, veja-se no Capítulo 3 a Secção 3.2.2.



A distribuição dos desvios por categoria revela uma maior incidência destes em nomes cujo valor de género não se infere a partir de indícios semânticos, nem a partir de indícios formais, i.e., na categoria 4. É também nesta categoria que se registam os valores médios de desvios por texto mais elevados. Com efeito, regista-se uma média aproximada de 2 desvios de AGN e de CNG pertencentes à categoria 4 por produção escrita, relativamente aos valores apurados (1,1, 1,4 e 1,5) nas restantes categorias de desvio.

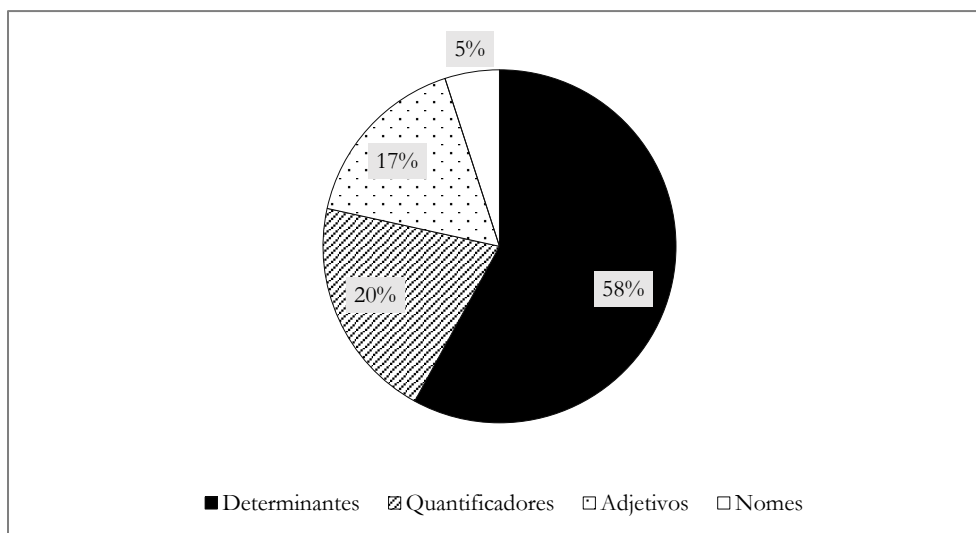
Em contrapartida, a proporção de desvios sobre nomes cujo valor de género se infere a partir de indícios semânticos é muito baixa, i.e., as categorias 1 e 2, atingindo os 2% e os 3% respetivamente. Quanto à média de desvios por texto, as categorias 1 e 2 registam 1,1 e 1,4 desvios, respetivamente, sendo que estes valores são inferiores ao registados nas demais categorias.

Ainda nesta primeira análise dos dados é observável também uma proporção de desvios relativamente alta sobre nomes cujo valor de género é inferível a partir de indícios formais, i.e., os integrados na categoria 3. Como se verifica, os desvios desta categoria correspondem a 35% do conjunto global de desvios apurado no estudo. Esta tendência é também observada em outros trabalhos sobre aquisição/aprendizagem da atribuição de género nominal em PLNM, nomeadamente em Ferreira (2011), Martins (2015) e Pinto (2015).

Veja-se, em seguida, no Quadro 4.5 a distribuição do número de desvios por constituintes afetados, bem como o registo dos valores da média ( $\bar{X}$ ) de desvios apurada por texto. O Gráfico 4.7 ilustra a distribuição dos desvios em valores percentuais.

Subcategorias de desvio		# de ocorrências de desvios	$\bar{X}$ de desvios
Especificadores	Determinantes ( <b>det.</b> )	725	1,7
	Quantificadores ( <b>quant.</b> )	255	1,2
Adjetivos ( <b>adj.</b> )		198	1,1
Nomes ( <b>nom.</b> )		63	1,1
$\Sigma$		1 241	2,1

**Quadro 4.5** – Número absoluto de desvios, e dos valores da média, por subcategorias de desvio



**Gráfico 4.7** – Percentagens de desvios, distribuídas por categorias

Em termos globais, são os especificadores, sobretudo os determinantes, os constituintes sintáticos mais afetados, por oposição aos adjetivos e nomes. A maior incidência dos comportamentos desviantes nos especificadores também já tinha sido registada noutros trabalhos (Ferreira 2011; Martins 2015; Pinto 2015).

Este resultado não é, contudo, corroborado pelo que foi observado noutros estudos. Com efeito, uma parte substancial da investigação disponível sobre a aquisição/aprendizagem da AGN e da CNG por aprendentes tardios de uma LNM aponta para uma maior incidência de desvios de atribuição dos valores de género nos adjetivos (veja-se, a este respeito, os contributos dos trabalhos de Bruhn de Garavito & White 2002 e White *et al.* 2004). No contexto da aquisição/aprendizagem da categoria de género gramatical por aprendentes tardios de PLN, Lacsán (2015) verificou que há uma maior incidência de desvios nos adjetivos do que nos determinantes, especialmente quando aqueles se encontram mais afastados do núcleo. A partir dos dados apurados, a autora considera que “gender assignment of determiners is more easily acquired than in adjectives” (2015:69).

O facto de os dados recolhidos para a presente investigação não apresentarem uma maior incidência de desvios de concordância nominal em género nos adjetivos pode estar, em parte, correlacionado com a metodologia adotada para o tratamento empírico. Em primeiro lugar, no presente estudo teve-se apenas em conta os desvios sobre adjetivos com função atributiva, enquanto o estudo de Lacsán abrange também dados relativos ao uso de adjetivos com função predicativa. Para além disso, os dados analisados por Lacsán (2015) foram extraídos de diferentes tarefas de produção, tais como nomeação de figuras, preenchimento de questionários, entre outras, especificamente desenvolvidas para conduzir

os aprendentes ao uso de estruturas que envolvessem relações de concordância nominal, com a manipulação de diferentes variáveis linguísticas envolvidas na atribuição dos valores de género aos nomes em português. Em contrapartida, neste trabalho, os dados empíricos foram recolhidos de textos autênticos, produzidos por aprendentes tardios de PLNM, não tendo havido uma orientação específica, quer sobre a avaliação do desempenho dos aprendentes relativamente à marcação de género nominal, quer sobre a utilização de especificadores e de adjetivos em diferentes contextos. Por conseguinte, o facto de se não terem controlado as condições de produção dos enunciados analisados terá condicionado o nível de incidência de desvios marcados nos diferentes elementos concordantes.

Referidos os resultados globais apurados segundo as características da tipologia de desvios, nas próximas secções procede-se a uma análise cuidada dos tipos e subtipos de desvios apurados, quer numa perspectiva quantitativa, quer qualitativa, tendo-se em linha de conta o desempenho de cada segmento da amostra, i.e., analisar-se-ão os diferentes tipos e subtipos de desvio em função da variável LM, bem como do nível QECRL da turma frequentada pelo aprendente. Segundo cremos, esta abordagem permitirá um maior conhecimento dos padrões de desenvolvimento da aquisição/aprendizagem do sistema de classificação nominal em PLNM, atendendo não só às características relativas ao perfil do aprendente tardio, nomeadamente no que concerne à configuração da gramática da LM e ao respetivo nível de aprendizagem em português, mas também às particularidades do *input* linguístico afetado.



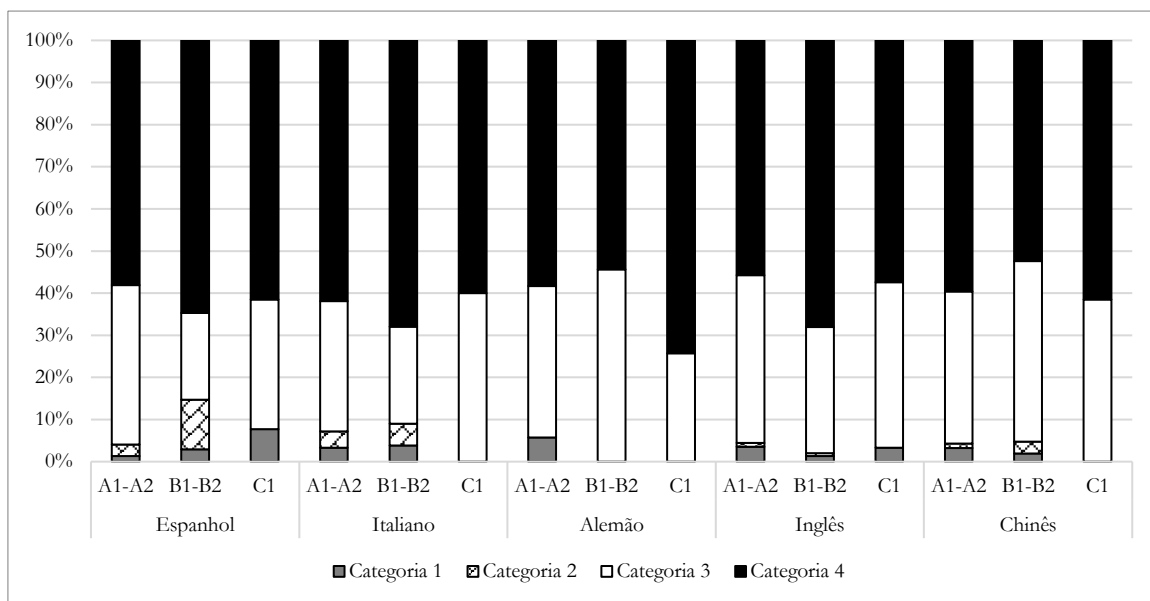
#### 4.2.3.1. Categorias e Subcategorias de desvios de AGN

No Quadro 4.6 encontra-se a distribuição do número absoluto de desvios, apurados por categoria, e dos valores da média ( $\bar{X}$ ) de desvios por texto, registados por cada segmento de amostra (LM e nível QECRL).

Informantes		Produções Escritas	Desvios								
LM	Nível QECRL		#	Categoria 1		Categoria 2		Categoria 3		Categoria 4	
		#		$\bar{X}$	#	$\bar{X}$	#	$\bar{X}$	#	$\bar{X}$	
Espanhol	A1-A2	46	74	1	0,02	2	0,04	28	0,61	43	0,93
	B1-B2	26	34	1	0,04	4	0,15	7	0,27	22	0,85
	C1	11	13	1	0,09	0	0,00	4	0,36	8	0,73
Italiano	A1-A2	104	181	6	0,06	7	0,07	56	0,54	112	1,08
	B1-B2	40	78	3	0,08	4	0,10	18	0,45	53	1,33
	C1	12	15	0	0,00	0	0,00	6	0,50	9	0,75
Alemão	A1-A2	74	175	10	0,14	0	0,00	63	0,85	102	1,38
	B1-B2	60	103	0	0,00	0	0,00	47	0,78	56	0,93
	C1	20	35	0	0,00	0	0,00	9	0,45	26	1,30
Inglês	A1-A2	46	113	4	0,09	1	0,02	45	0,98	63	1,37
	B1-B2	54	147	2	0,04	1	0,02	44	0,81	100	1,85
	C1	29	61	2	0,07	0	0,00	24	0,83	35	1,21
Chinês	A1-A2	37	94	3	0,08	1	0,03	34	0,92	56	1,51
	B1-B2	34	105	2	0,06	3	0,09	45	1,32	55	1,62
	C1	10	13	0	0,00	0	0,00	5	0,50	8	0,80
$\Sigma$		<b>603</b>	<b>1 241</b>	<b>35</b>	<b>0,06</b>	<b>23</b>	<b>0,04</b>	<b>435</b>	<b>0,72</b>	<b>748</b>	<b>1,24</b>

**Quadro 4.6** – Distribuição do número absoluto e dos valores da média de ocorrências de desvios por categoria apurados por texto produzido em cada segmento da amostra (LM e nível QECRL)

Tendo em conta o número absoluto de desvios por segmento, procedeu-se, depois, ao cálculo das percentagens de desvios de AGN e de CNG em função das categorias de desvios estipuladas para a presente investigação (cf. Gráfico 4.8).



**Gráfico 4.8** – Percentagem de desvios por categoria em função da LM e do nível QECRL da turma frequentada pelos informantes

No que concerne à distribuição dos resultados atendendo ao tipo de desvio produzido, é possível constatar que em todos os segmentos da amostra se observam padrões idênticos, i.e., os itens nominais cujo valor de género não se infere a partir de critérios semânticos [SEM -] nem a partir de critérios formais [FORM -] (categoria 4) apresentam uma maior incidência de desvios comparativamente ao que se verifica em itens nominais cujo valor de género se infere a partir de critérios de natureza semântica (categorias 1 e 2).

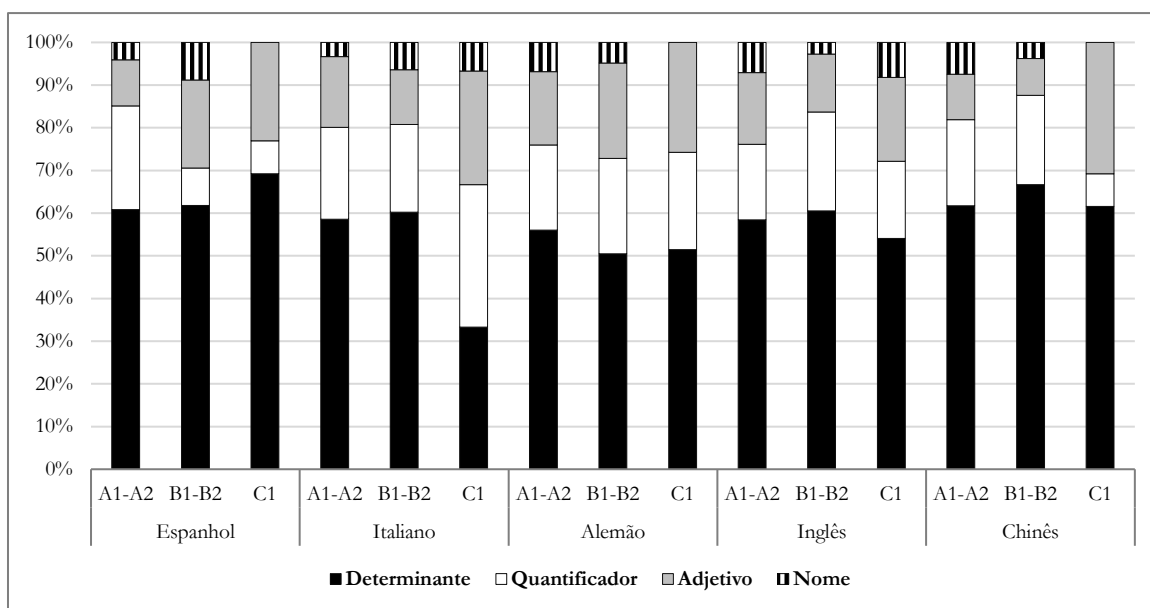
A maior incidência de desvios na categoria 4 face ao que se regista nas categorias 1 e 2 (Quadro 4.6) poderá também estar relacionada com o número reduzido de ocorrências de nomes pertencentes a estas duas últimas categorias nos textos seleccionados (cf. Quadros 4.8 e 4.10). Por conseguinte, num momento posterior da análise empírica, ter-se-á em conta o peso relativo dos desvios de AGN e de CNG em função do número de ocorrências dos tipos de nomes, registado por diferentes segmentos da amostra (cf. subsecções seguintes).

Da observação destes resultados globais é de assinalar ainda, nos vários segmentos, o número avultado de desvios em nomes cujo valor de género é corretamente dedutível a partir dos índices temáticos *-o* e *-a*, i.e., na categoria 3. Deste modo, poder-se-á supor que, face à ambiguidade do *input*, os aprendentes demonstram uma certa ‘desconfiança’ perante o poder preditivo dos indicadores morfológicos para a atribuição do valor de género nominal.

No que diz respeito aos constituintes em que se evidencia(m) as marcas de desvio, o Quadro 4.7 apresenta a distribuição do número de desvios produzido em cada segmento da amostra e dos valores da média ( $\bar{X}$ ) apurados por texto. No Gráfico 4.9 é possível verificar a proporção dos desvios em valores percentuais.

Informantes		Produções Escritas	Desvios								
LM	Nível QERCL		#	#	Determinante		Quantificador		Adjetivo		Nome
		#			$\bar{X}$	#	$\bar{X}$	#	$\bar{X}$	#	$\bar{X}$
Espanhol	A1-A2	46	74	45	0,98	18	0,39	8	0,17	3	0,07
	B1-B2	26	34	21	0,81	3	0,12	7	0,27	3	0,12
	C1	11	13	9	0,82	1	0,09	3	0,27	0	0,00
Italiano	A1-A2	104	181	106	1,02	39	0,38	30	0,29	6	0,06
	B1-B2	40	78	47	1,18	16	0,40	10	0,25	5	0,13
	C1	12	15	5	0,42	5	0,42	4	0,33	1	0,08
Alemão	A1-A2	74	175	98	1,32	35	0,47	30	0,41	12	0,16
	B1-B2	60	103	52	0,87	23	0,38	23	0,38	5	0,08
	C1	20	35	18	0,90	8	0,40	9	0,45	0	0,00
Inglês	A1-A2	46	113	66	1,43	20	0,43	19	0,41	8	0,17
	B1-B2	54	147	89	1,65	34	0,63	20	0,37	4	0,07
	C1	29	61	33	1,14	11	0,38	12	0,41	5	0,17
Chinês	A1-A2	37	94	58	1,57	19	0,51	10	0,27	7	0,19
	B1-B2	34	105	70	2,06	22	0,65	9	0,26	4	0,12
	C1	10	13	8	0,80	1	0,10	4	0,40	0	0,00
$\Sigma$		<b>603</b>	<b>1 241</b>	<b>725</b>	<b>1,20</b>	<b>255</b>	<b>0,42</b>	<b>198</b>	<b>0,33</b>	<b>63</b>	<b>0,10</b>

**Quadro 4.7** – Distribuição do número absoluto e dos valores da média de ocorrências de desvios por constituinte afetado, apurados por texto escrito em cada segmento da amostra (LM e nível QERCL)



**Gráfico 4.9** – Distribuição das percentagens de desvios por constituinte afetado em função da LM e do nível QECL da turma frequentada pelos informantes

Conforme se pode verificar no Gráfico 4.9, em todos segmentos da amostra a maior proporção de desvios concentra-se nos especificadores, sobretudo nos determinantes, sendo que a proporção de desvios relativos a nomes é a mais baixa.

Verifica-se ainda que é especialmente em C1 que se regista uma maior proporção de desvios de AGN e de CNG nos adjetivos. Com efeito, uma parte substancial das ocorrências desviantes registada neste nível incide sobre os adjetivos, sendo particularmente frequentes as situações de ambivalência de atribuição do valor de género, i.e., situações nas quais os aprendentes selecionam adequadamente a forma do especificador de acordo com o valor de género do nome, embora a forma do adjetivo não concorde em género com o nome, conforme ilustram os exemplos seguintes, extraídos do *corpus*:

- (16) «\*(...) onde encontra\va/mos muitas **pessoas ingleses**.» (Alemão.C1)
- (17) «\*Eu acho que tenho uma **alimentação adequado**.» (Inglês.C1)
- (18) «\*Nas nossas férias, só passeámos nas **ruas limpos**, visitámos o símbolo da cidade.» (Chinês.C1)
- (19) «\*A gente, depois de ter experimentado a extraordinária facilidade do seu utilizo, dificilmente decide regressar ao utilizo dos **meios de comunicação usadas** anteriormente» (Italiano.C1)
- (20) «\*Sei que não eres uma **mau rapariga** mas depois do acontecido não sei como olhar para ti.» (Espanhol.C1)



Nas próximas subsecções, procede-se a uma análise detalhada dos desvios de atribuição dos valores de género e de concordância nominal em género, por categoria e atendendo aos constituintes sintáticos afetados. Procurar-se-á, assim, averiguar os padrões dos comportamentos desviantes dos diferentes segmentos da amostra ao longo do desenvolvimento da aquisição/aprendizagem da atribuição de género aos nomes em PLNM.

#### 4.2.3.1.1. Categoria 1 [SEM +, FORM +] (*o menino, a menina*)

Como referido anteriormente, sob a categoria 1 registaram-se os desvios relativos a nomes cujo valor de género se infere, simultaneamente, a partir de critérios de natureza semântica [SEM +] e formal [FORM +].

Os desvios assinalados incidem, essencialmente, sobre nomes referentes a pessoas e, em particular, sobre itens lexicais que nomeiam relações interpessoais, conforme ilustram os exemplos seguintes:

(21) «\*O meu melhor **amiga** chama-se XXX.» (Inglês.A1-A2)

(22) «\*uma hora depois de sua **marido** quem chega à casa as 17h45 (...)»  
(Inglês.C1)

(23) «\*Como tenho algumas **amigos** que moram em Lisboa (...)»  
(Italiano.B1-B2)

Registam-se apenas duas ocorrências desviantes em nomes que referem animais:

(24) «\*Ela mora cum o seus pais e um **cadela**.» (Italiano.A1-A2)

(25) «\*Houve uma altura em que gostava de tirar fotos as (a) **gatos...**»  
(Italiano.A1-A2)

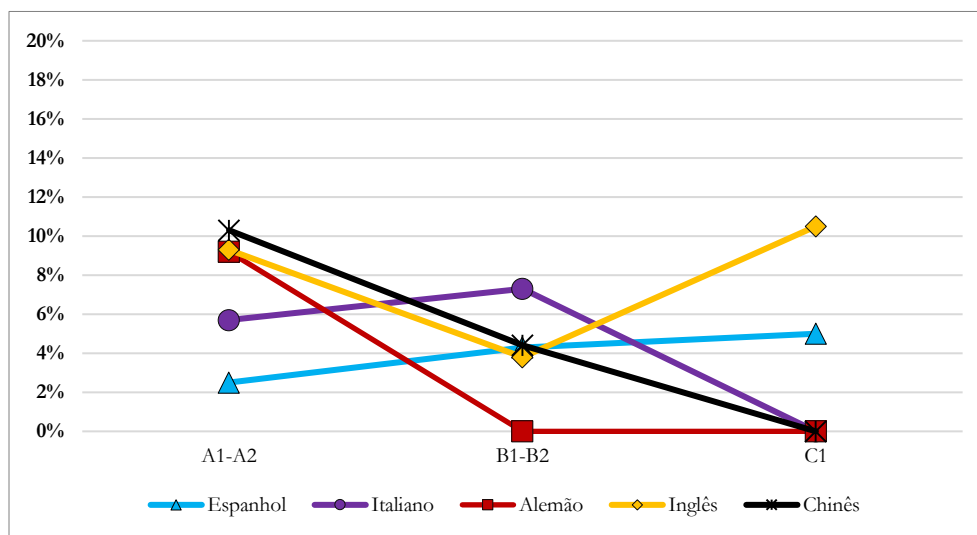
A fim de identificar o peso relativo dos desvios da categoria 1, contabilizou-se no *corpus* o número absoluto de ocorrências deste tipo de itens em SN não reduzidos (cf. Quadro 4.8). Ou seja, foram contabilizadas as ocorrências de nomes como *menino*, *menina*, *professora* e *rapariga*, por exemplo.

Para além dos valores absolutos, no Quadro 4.8 apresentam-se também os valores da média ( $\bar{X}$ ) de nomes e de desvios da categoria 1 produzidos por texto selecionado. Conforme é possível constatar, a média de ocorrências de nomes desta categoria em cada texto é aproximadamente de 1 item por texto, sendo que a média de desvios de AGN no conjunto global de textos é de 0,06, tratando-se, então, de um valor muito baixo.

Informantes		Produções Escritas		Categoria 1				
LM	Nível QERCL	#	# de ocorrências de nomes	Nomes		Desvios		
				# de ocorrências	$\bar{X}$	# de ocorrências	$\bar{X}$	%
Espanhol	A1-A2	46	1 280	40	0,87	1	0,02	2,5
	B1-B2	26	933	23	0,88	1	0,04	4,3
	C1	11	549	20	1,82	1	0,09	5
Italiano	A1-A2	104	3 141	106	1,02	6	0,06	5,7
	B1-B2	40	1 385	41	1,03	3	0,08	7,3
	C1	12	1 699	74	6,17	0	0	0
Alemão	A1-A2	74	1 784	109	1,47	10	0,14	9,2
	B1-B2	60	2 407	43	0,72	0	0	0
	C1	20	928	31	1,55	0	0	0
Inglês	A1-A2	46	926	43	0,93	4	0,09	9,3
	B1-B2	54	2 783	52	0,96	2	0,04	3,8
	C1	29	1 127	19	0,66	2	0,07	10,5
Chinês	A1-A2	37	957	29	0,78	3	0,03	10,3
	B1-B2	34	1 221	45	1,32	2	0,06	4,4
	C1	10	343	2	0,20	0	0	0
$\Sigma$		603	21 463	677	1,12	35	0,06	5,17

**Quadro 4.8** – Distribuição do número absoluto e dos valores da média de ocorrências de nomes da categoria 1 produzidos no *corpus* e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QERCL)

Atendendo ao número absoluto de ocorrências de nomes pertencentes à categoria 1 produzido por segmento da amostra, fez-se o cálculo das percentagens relativas de desvios que se apresentam no Gráfico 4.10.



**Gráfico 4.10** – Percentagem relativa de desvios da categoria 1 por segmento da amostra (LM e nível QERCL dos informantes)

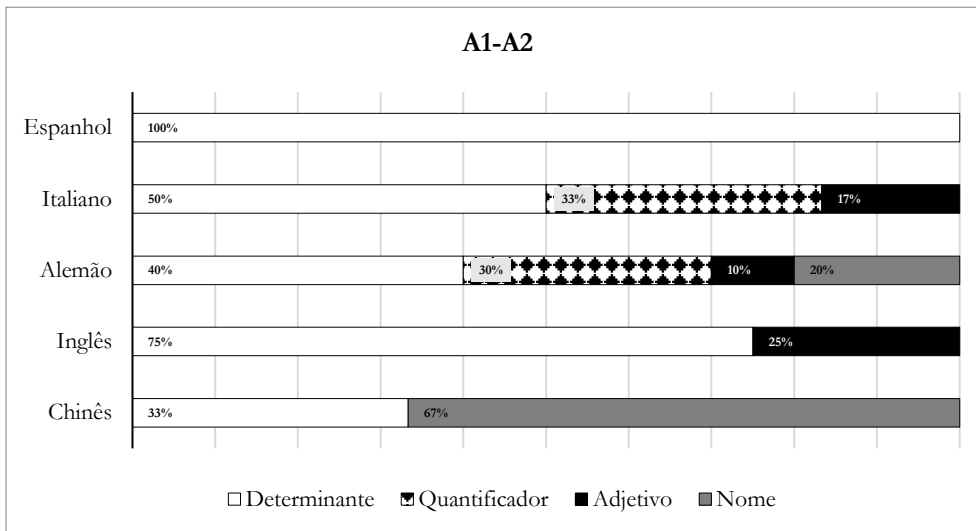
Conforme ilustra o Gráfico 4.10, neste tipo de itens nem todos os segmentos da amostra registam uma diminuição gradual da percentagem relativa de desvios à medida que progredem na aprendizagem do português, ou seja, à medida que avançam no nível QECRL em português. Aliás, apenas nas produções escritas por falantes nativos de chinês e de alemão se observa uma tendência decrescente das percentagens relativas de desvios desde os níveis QECRL A1-A2 até C1, sendo que, neste nível, não se registam quaisquer casos desviantes da categoria 1 de informantes falantes destas duas LM. Também no segmento da amostra de informantes que são falantes nativos de italiano não se registam quaisquer desvios da categoria 1 no nível C1.

Em contrapartida, nas produções escritas por falantes nativos de espanhol e de inglês regista-se uma tendência crescente de desvios nesta categoria de nomes, sendo que, no segmento de informantes de LM espanhola, o nível C1 apresenta, aproximadamente, o dobro da percentagem relativa de desvios registada nos níveis A1-A2, bem como um número mais elevado do valor da média de desvios de AGN e de CNG da categoria 1 por texto (cf. Quadro 4.8).

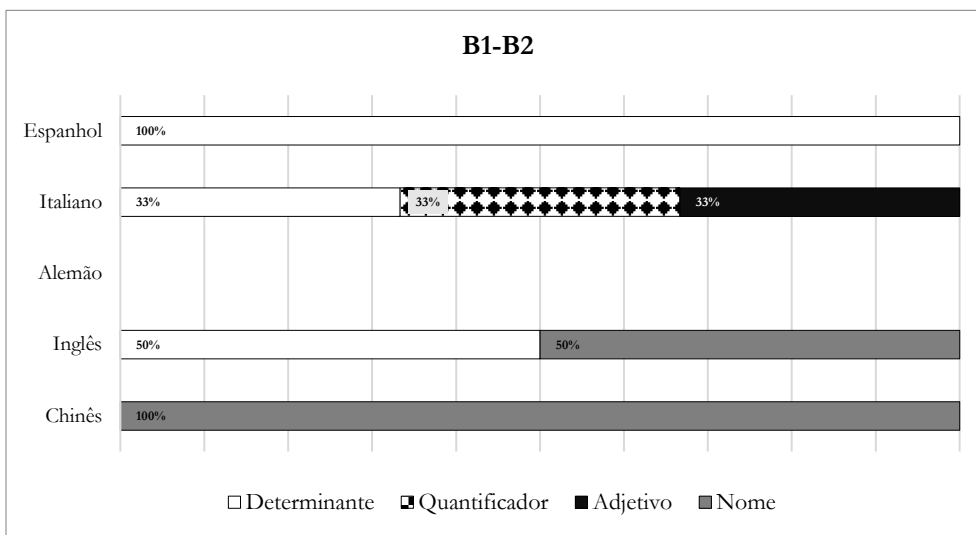
Convém, no entanto, relembrar que os dados no Gráfico 4.10 estão correlacionados com o número de ocorrências dos itens pertencentes à categoria 1 presentes nos textos produzidos pelos diferentes segmentos de amostra, repercutindo-se, então, a proporção do peso dos comportamentos desviantes da categoria 1 face às ocorrências deste tipo de itens nos textos selecionados. Sendo assim, o facto de haver um aumento gradual do peso relativo dos desvios não é representativo de um maior número de desvios; o que se verifica é um valor proporcional maior em função do número absoluto de ocorrências de itens da categoria 1 por nível QECRL.

No caso dos falantes nativos de espanhol, por exemplo, o que se verifica é uma redução gradual do número de ocorrências destes itens nos textos produzidos nos níveis B1-B2 e C1, e, em resultado disso, o peso relativo das formas desviantes vai sendo proporcionalmente maior (cf. Quadro 4.8). Situação idêntica verifica-se também no segmento da amostra de informantes de LM inglesa que, em C1, regista uma percentagem relativa de desvios da categoria 1 consideravelmente mais elevada do que a registada nos níveis B1-B2.

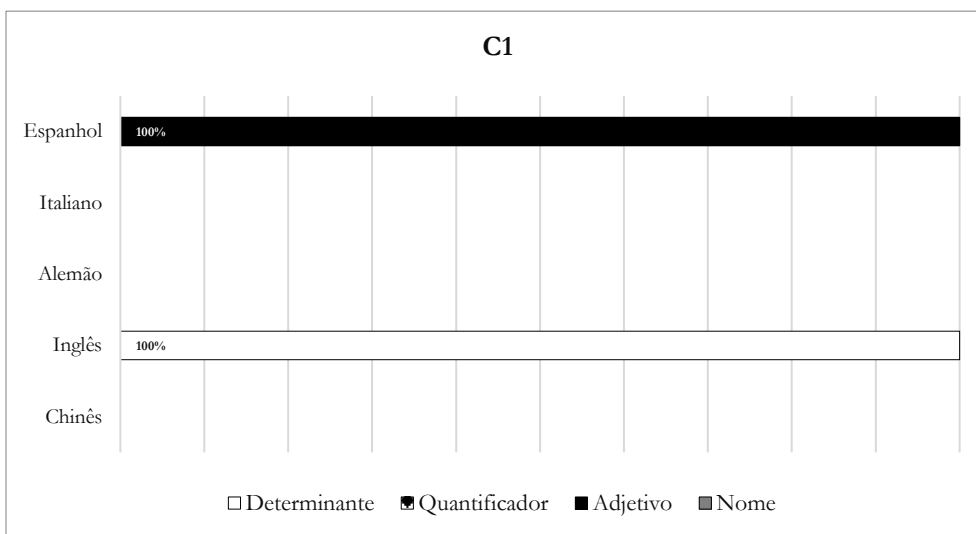
Os Gráficos 4.11, 4.12 e 4.13 registam, em valores percentuais, a proporção dos desvios de atribuição de valores de género nominal e da concordância nominal em género da categoria 1 por constituinte sintático afetado (determinante, quantificador, adjetivo e nome).



**Gráfico 4.11** – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos nos níveis A1-A2



**Gráfico 4.12** – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos nos níveis B1-B2



**Gráfico 4.13** – Percentagem de desvios da categoria 1, distinguidos por classes de palavras afetadas, produzidos no nível C1

Nos níveis A1-A2 regista-se uma maior proporção de desvios nos especificadores, nomeadamente, no determinante:

- (26) «\*Gozar muito tempo com **meu filha** e saver de suas niquietudes.»  
(Espanhol.A1-A2)
- (27) \* **O meu** melhor **amiga** chama-se XXX.» (Inglês.A1-A2)
- (28) «\* Eu conheci <muito> **as** muitos novos **amigos** e aprendi a língua portuguesa mais ou menos.» (Alemão.A1-A2)

É igualmente assinalável, nestes níveis, a proporção de desvios nos quantificadores:

- (29) «\***Algumas** dos meus **amigos** já chegaram no dia 29 e tiveram ficar num pensão por um noite.» (Alemão.A1-A2)
- (30) «\* Em Veneza eu abito com **dois filhas** de Verona e Rovigo e um filho.»  
(Italiano.A1-A2)

Nos níveis B1 e B2, continua a haver uma maior proporção de desvios nos especificadores, nomeadamente nos determinantes, sendo que o segmento de informantes, que são falantes nativos de alemão não regista qualquer desvio nestes níveis.

Observa-se ainda, nestes níveis, um predomínio de desvios de AGN nas produções escritas por falantes nativos de chinês, já que as únicas duas ocorrências desviantes da categoria 1 registadas neste segmento incidem nesta classe de palavras:

- (31) «\*Eu tive um emprego temporário com **duas amigos** na Bobadela.»  
(Chinês.B1-B2)
- (32) «\*A sua carreira é **uma enfermeiro** de plantão.» (Chinês.B1-B2)

Como referido anteriormente, a averiguação destes casos depende do contexto em que estes itens são utilizados. Assim sendo, partindo da leitura da produção textual infere-se que a forma pretendida pelos aprendentes é *amigas* e *enfermeira*.

Por fim, no nível C1, para além de haver segmentos da amostra sem quaisquer desvios de AGN e de CNG (os textos escritos por falantes nativos de italiano, de alemão e de chinês não apresentam desvios da categoria 1), regista-se um predomínio de desvios na classe dos adjetivos nos textos escritos por aprendentes que são falantes nativos de espanhol:

- (33) «\*Sei que eres uma **mau rapariga** mas depois do acontecido não sei como olhar para ti.» (Espanhol.C1)

Já os aprendentes de LM inglesa registam, à semelhança do que se verificara em outros níveis de proficiência, desvios no especificador, nomeadamente no determinante:

- (34) «\*Seu pai, XXX, vem a casa às 18h45 da noite, uma hora depois de **sua marido** quem chega à casa as 17h45.» (Inglês.C1)

No que concerne ao item lexical afetado, considere-se o Quadro 4.9 em que se registam os nomes afetados por segmentos da amostra, com a indicação do respetivo número de ocorrências no *corpus* e com a identificação, por item, do número de ocorrências de desvios de AGN e de CNG.

Nível QERCL									
LM	A1-A2			B1-B2			C1		
	Nomes	# de ocorrências no <i>corpus</i>	# de desvios	Nomes	# de ocorrências no <i>corpus</i>	# de desvios	Nomes	# de ocorrências no <i>corpus</i>	# de desvios
Espanhol	filha	2	1	amigo	9	1	rapariga	1	1
Italiano	amiga	12	2	amigo	18	2			
	cadela	4	1	rapariga	5	1			
	filha	4	1						
	gato	3	1						
	vizinho	3	1						
Alemão	amiga	19	4						
	amigo	51	3						
	madrasta	2	1						
	namorado	8	1						
	tia	2	1						
Inglês	amiga	21	4	amigo	36	2	amigo	10	1
		-			-		marido	3	1
Chinês	amigo	33	1	amiga	12	1			
	primo	2	1	enfermeira	1	1			
	professora	3	1						
$\Sigma$		169	24	$\Sigma$	81	8	$\Sigma$	14	3

**Quadro 4.9** – Distribuição do número de ocorrências de nomes com desvio da categoria 1 no *corpus* e respetivo número de desvios por segmentos da amostra (LM e Nível QERCL)

Como se depreende da leitura do Quadro 4.9, registaram-se 24 nomes da categoria 1 que deram origem a um total de 35 ocorrências de desvios de AGN e de CNG, havendo, entre este conjunto, 13 itens nominais distintos: *amigo*, *amiga*, *cadela*, *enfermeiro*, *filha*, *gato*, *madrasta*, *namorado*, *marido*, *primo*, *professora*, *rapariga*, *vizinho* e *tia* (cf. Quadro 4.9). Verifica-se que, na maioria dos segmentos da amostra, os nomes afetados não surgem recorrentemente com desvios nos diferentes níveis QERCL, à exceção do segmento de

falantes de LM inglesa no qual os nomes *amigo* e *amiga* apresentam desvios nos textos representativos dos diferentes níveis.

Para além disso, e como se depreende do Quadro 4.9, na maior parte dos nomes afetados, e salvo casos pontuais assinalados em alguns segmentos da amostra selecionada, o peso relativo dos desvios face ao número absoluto de ocorrências dos nomes afetados nos textos é baixa. Por outras palavras, verifica-se que, aquando da produção destes itens, há uma maior incidência de acertos do que de desvios. Por exemplo, no segmento da amostra de informantes de LM alemã a frequentar turmas de níveis A1-A2, o nome *amigo* registou um total de 51 ocorrências, tendo-se apenas apurado 3 desvios neste item. Contudo, como já referido anteriormente, em alguns segmentos da amostra verifica-se que a proporção de desvios face ao número absoluto de ocorrências do nome afetado é o mesmo.



Por fim, da análise global dos desvios da categoria 1, poder-se-á concluir que este tipo de itens não é particularmente problemático em todos os segmentos da amostra estudados, já que, para além de serem poucos os itens nominais afetados, a proporção de desvios em relação aos acertos é tendencialmente baixa. Quanto à classe de palavras que, no SN, apresenta desvios, verifica-se também que, entre os diferentes segmentos da amostra por LM, há uma maior incidência de comportamentos desviantes nos especificadores, nomeadamente nos determinantes. Só no nível C1 é que se observa um padrão ligeiramente distinto, já que para além de haver segmentos da amostra sem quaisquer desvios de AGN e de CNG neste nível, nas produções escritas por falantes nativos de espanhol o único desvio detetado em C1 incide sobre a forma do adjetivo ‘mau’ que integra o SN cujo núcleo é o nome *rapariga*.

#### 4.2.3.1.2. Categoria 2 [SEM +, FORM -] (*o homem, a mulher*)

A categoria em análise integra os desvios de AGN e CNG relativos a nomes com referentes extralinguísticos [+ sexuais] e cujos valores de género são corretamente dedutíveis a partir de critérios de natureza semântica [SEM +], mas não de critérios de natureza formal [FORM -]. Nestes itens observa-se uma relação entre o valor de género gramatical do item lexical, masculino e feminino, e o género natural da entidade designada, macho e fêmea, respetivamente. Por se tratar de nomes em que os indícios semânticos auxiliam na associação dos valores de género gramatical, postula-se que este tipo de itens apresente um menor grau de dificuldade em todos os segmentos da amostra de informantes,

tendência, aliás, observada na literatura disponível, mesmo no caso dos aprendentes cuja LM não possui um sistema de atribuição de gênero gramatical.

Nos exemplos (35) a (37) apresentam-se alguns dos desvios assinalados sob a categoria 2:

- (35) «\*Na minha família, só há 3 membros, o meu pai, o minha mãe e mim.»  
(Chinês.A1-A2)
- (36) «\*As rapazes e as raparigas do curso são simpáticos também.»  
(Italiano.A1-A2)
- (37) «\*Finalmente, ela tinha de fazer um funeral em sua coração para o seu irmã.» (Chinês.B1-B2)

O exemplo (35) ilustra ainda um caso típico de ambivalência do aprendente relativamente à atribuição do valor de gênero, já que, no momento de estabelecer a concordância com os demais elementos que integram o SN com o núcleo ‘*mãe*’, o informante atribui incorretamente o valor de gênero masculino ao artigo definido, utilizando a forma ‘*o*’ em vez de ‘*a*’, embora recorra à forma do possessivo adequada: ‘*minha*’.

Tal como foi feito para os casos que integram a categoria 1, para calcular o peso relativo dos desvios respeitantes à categoria 2, contabilizou-se o número absoluto de ocorrências de nomes em SN não reduzidos que, atendendo aos indícios de atribuição de valores de gênero nominal, integram esta categoria. Assim, contabilizou-se no *corpus* selecionado o número de ocorrências de nomes como *homem*, *mulher* e *rapaz*, por exemplo, tendo-se registado, igualmente, os valores da média de ocorrências de nomes da categoria 2 produzidos por texto em cada segmento da amostra (cf. Quadro 4.10).

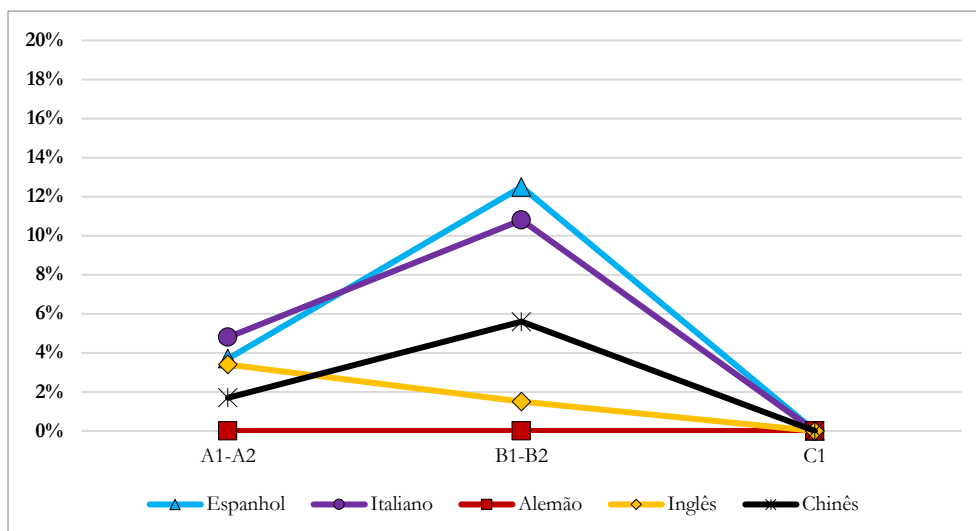


Informantes		Produções Escritas		Categoria 2				
LM	Nível QERCL	#	# de ocorrências de nomes	Nomes		Desvios		
				# de ocorrências	$\bar{X}$	# de ocorrências	$\bar{X}$	%
Espanhol	A1-A2	46	1 280	54	1,17	2	0,04	3,7
	B1-B2	26	933	32	1,23	4	0,15	12,5
	C1	11	549	15	1,36	0	0	0
Italiano	A1-A2	104	3 141	147	1,41	7	0,07	4,8
	B1-B2	40	1 385	37	0,93	4	0,10	10,8
	C1	12	1 699	114	9,50	0	0	0
Alemão	A1-A2	74	1 784	48	0,65	0	0	0
	B1-B2	60	2 407	35	0,58	0	0	0
	C1	20	928	32	1,60	0	0	0
Inglês	A1-A2	46	926	29	0,63	1	0,02	3,4
	B1-B2	54	2 783	68	1,26	1	0,02	1,5
	C1	29	1 127	17	0,59	0	0	0
Chinês	A1-A2	37	957	60	1,62	1	0,03	1,7
	B1-B2	34	1 221	54	1,59	3	0,09	5,6
	C1	10	343	5	0,50	0	0	0
$\Sigma$		603	21 463	747	1,24	23	0,04	3,08

**Quadro 4.10** – Distribuição do número absoluto e dos valores de média de ocorrências de nomes da **categoria 2** produzidos no *corpus* e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QERCL)

Assim, registou-se, em termos globais, um valor médio de ocorrências de 1 nome da categoria 2 por texto, à semelhança do que se apurara para a categoria 1 (cf. Quadro 4.8). Já o valor da média de desvios de AGN e de CNG por texto é ligeiramente inferior ao registado na primeira categoria analisada, sendo igualmente muito baixo: 0,04 desvios por texto selecionado.

Tendo em conta o número absoluto de ocorrências de nomes da categoria 2, calculou-se a percentagem relativa de desvios por cada grupo de aprendentes. Os valores percentuais apurados figuram no Gráfico 4.14.

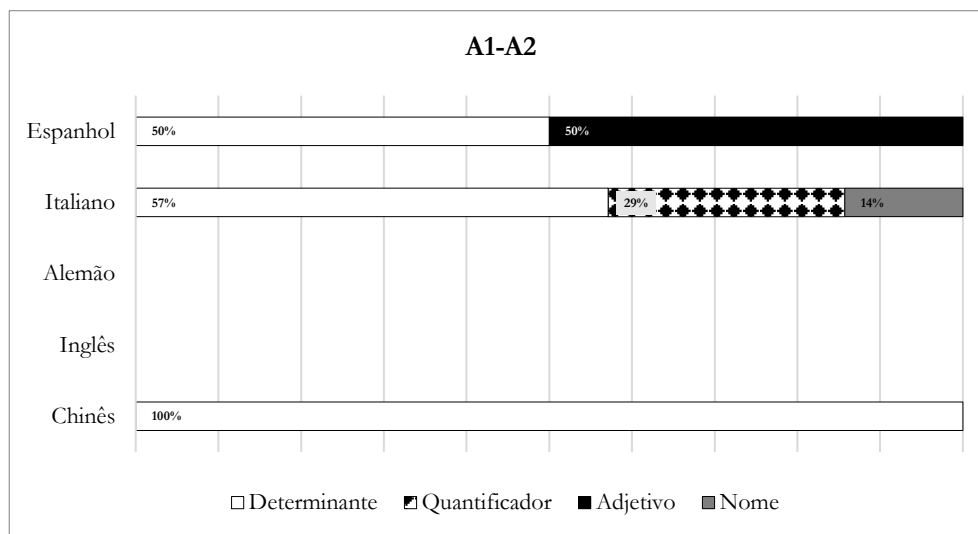


**Gráfico 4.14** – Percentagem relativa de desvios da categoria 2

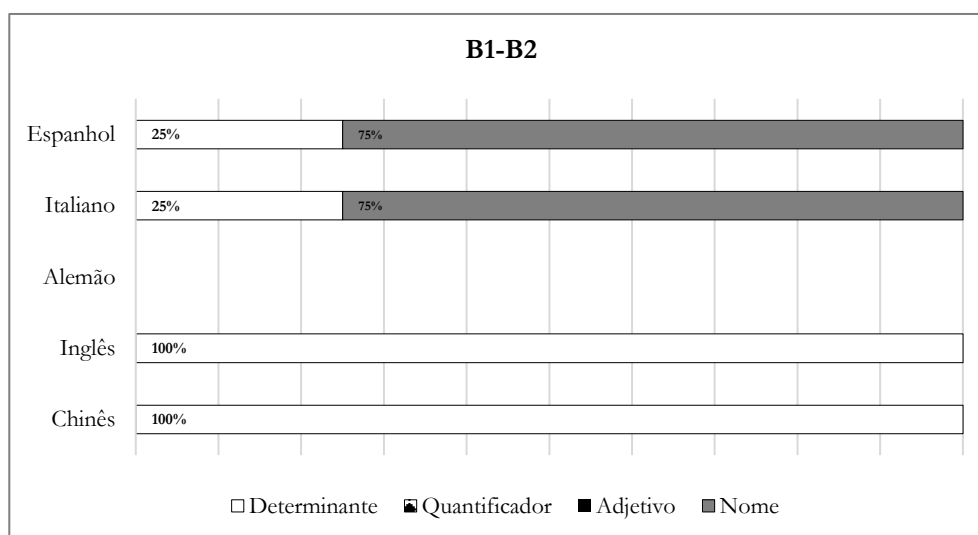
Da leitura do Quadro 4.10 e do Gráfico 4.14, verifica-se que em todos os segmentos da amostra por LM o nível C1 não apresenta quaisquer desvios da categoria 2. Para além disso, nos textos dos falantes nativos de alemão há uma ausência total de desvios de AGN e de CNG em nomes desta categoria.

Regista-se ainda, sobretudo nos segmentos da amostra B1-B2, padrões distintos, por LM, da expressão do desvio. Com efeito, segundo os dados cartografados no Gráfico 4.14, nestes níveis regista-se uma maior proporção de desvios nos textos produzidos por aprendentes que são falantes nativos de línguas românicas, relativamente aos valores apurados nos textos escritos por falantes nativos de inglês e de chinês. Ora, poder-se-ia postular que a maior proporção de desvios nos segmentos da amostra de informantes que são falantes nativos dos idiomas românicos se deve a uma transferência linguística dos valores de género da LM para os itens correspondentes na LA. Porém, atendendo aos itens nominais afetados nos segmentos da amostra de informantes de LM italiana e espanhola, verifica-se que a maior proporção de desvios registada não é atribuível à transferência dos valores de género do idioma nativo para a LA, já que os itens correspondentes da LM possuem os mesmos valores de género do português. Assim, estes resultados poder-se-ão justificar pelo desconhecimento, por parte destes aprendentes, das formas concordantes dos constituintes sintáticos que integram o SN.

Os Gráficos 4.15 e 4.16 registam a proporção, em valores percentuais, dos desvios da categoria 2 por constituinte afetado nos vários segmentos da amostra.



**Gráfico 4.15** – Percentagem de desvios da categoria 2, distinguidos por classes de palavras afetadas, nos níveis A1-A2



**Gráfico 4.16** – Percentagem de desvios da categoria 2, distinguidos por classes de palavras afetadas, nos níveis B1-B2

Tendo em conta os dados cartografados nos Gráficos 4.15 e 4.16, é visível o predomínio, quer nos níveis A1-A2, quer nos níveis B1-B2, dos desvios sobre os especificadores, nomeadamente os determinantes:

- (38) «\*Tenho **a irmão** que se chama XXXXX, **a pai**, XXXXX e a mãe e depois tenho dois cão.» (Italiano. A1-A2)
- (39) «\*Por exemplo, nos Estados Unidos e no Europa **os mulheres** têm mais directos.» (Inglês. B1-B2)

A persistência de desvios nos determinantes é ilustrativa da dificuldade, por parte da maioria dos aprendentes, na correta atribuição do valor de gênero nesta classe de palavras, tal como se observara em relação à categoria 1.

Regista-se ainda, nas produções dos falantes nativos de línguas românicas a frequentar turmas dos níveis B1-B2, uma maior proporção de desvios da categoria 2 sobre nomes:

- (40) «\*Se falares em público a fazer uma intervenção útil, a **professor** não deveria zangar-se.» (Espanhol.B1-B2)
- (41) «\*Eu sempre vivi na cidade, mas quando era pequena costumava passar as férias de verão na casa do meu **avó** no campo.» (Italiano.B1-B2)

Portanto, a partir da análise dos textos, constata-se que o item nominal pretendido no exemplo (40) é *professora* e não *professor*, e em (41), é *avô* e não *avó*.

Apresentada a quantificação dos dados, veja-se, no Quadro 4.11, a distribuição dos nomes afetados com a indicação do respetivo número de desvios a que deram origem.

LM	Nível QERCL					
	A1-A2			B1-B2		
	Nomes	# de ocorrências no corpus	# de desvios	Nomes	# de ocorrências no corpus	# de desvios
Espanhol	irmão	10	2	competidor	1	1
	-			professor	5	3
Italiano	avô	3	1	avô	4	2
	irmão	14	2	avó	1	1
	irmã	8	1	homem	13	1
	mãe	14	1	-		
	pai	25	1			
	rapaz	12	1			
Inglês	irmão	5	1	mulher	18	1
Chinês	mãe	20	1	irmã	4	3
Σ		111	11	Σ	46	12

**Quadro 4.11** – Distribuição do número de ocorrência de nomes com desvio da categoria 2 no *corpus* e respetivo número de desvios por segmentos da amostra (LM e nível QERCL)

O total de nomes afetados é muito reduzido, correspondendo a 11 itens distintos: *avô*, *avó*, *competidor*, *homem*, *irmão*, *irmã*, *mãe*, *mulher*, *pai*, *professor*, *rapaz*, que originaram, por sua vez, 23 desvios. Trata-se, na sua maioria, de itens que nomeiam relações de parentesco.

Tal como se verificou na categoria 1, em todos os segmentos da amostra a tendência é a de não haver desvio no mesmo item nominal nos diferentes níveis QECRL, o que sugere, então, que os informantes assimilaram adequadamente o valor de género associado a cada um dos itens. Para além disso, o número de ocorrências convergentes destes nomes é, na esmagadora maioria dos casos, muito superior ao dos desvios de AGN e de CNG e, por este motivo, poder-se-á considerar que estes são casos pontuais, não havendo, portanto, uma recorrência do desvio sempre que o falante aprendente produz este tipo de itens.



Em suma, os dados apurados parecem indiciar que, ao longo do processo de assimilação dos valores de género dos nomes em português, os critérios semânticos de atribuição de valores de género aos nomes possuem um peso mais significativo do que os critérios formais, em todos os segmentos da amostra seleccionada.

#### 4.2.3.1.3. Categoria 3 [SEM -, FORM +] (*o carro, a casa, o indivíduo, a pessoa*)

Consideremos, agora, os desvios relativos a itens nominais cujo valor de género não se infere a partir da aplicação de critérios de natureza semântica [SEM -], mas sim a partir de critérios formais [FORM +]. Nos nomes afetados pertencentes à categoria 3, observa-se uma relação, ainda que parcial, entre os índices temáticos *-o* e *-a* e valores de género masculino e feminino, respetivamente. Nos exemplos (42) a (45), apresentam-se alguns dos desvios de AGN e de CNG da categoria 3 registados no *corpus*:

- (42) «\*<(...)> Quase todos os dias eu vou **ao aldeia mais próximo** a minha morada para estar com minha namorada.» (Espanhol.A1-A2)
- (43) «\*Não pôde ver **o seu cara**» (Chinês.B1-B2)
- (44) «\*Eu sou muito triste porque Dortmund, **o meu equipe preferido**, perdeu contra Schalke, sabes?» (Alemão.A1-A2)
- (45) «\*Rapidamente, preparei a minha (bagga) bagagem e apanhei um taxi **na aeroporto**.» (Inglês.B1-B2)

À semelhança do modo como se procedeu no caso das outras categorias, fez-se a contagem do número de ocorrências de nomes em SN não reduzidos que integram a categoria 3 nos textos seleccionados (cf. Quadro 4.12), com vista a determinar a sua frequência de uso por produção escrita. Para além do valor absoluto, contabilizou-se também os valores da média ( $\bar{X}$ ) de ocorrências de nomes e de desvios da categoria 3 por

texto em cada segmento da amostra, tendo-se determinado no conjunto global, aproximadamente, um valor médio de ocorrências de 20 itens por produção textual, tratando-se, então, de uma categoria com alta frequência de nomes. Quanto aos valores da média de desvio, apurou-se, em termos globais, uma média de 0,72 desvios por texto, sendo este valor consideravelmente mais alto do que o registado nas categorias anteriormente analisadas (cf. Secções 4.2.3.1.1. e 4.2.3.1.2.).

Informantes		Produções Escritas		Categoria 3				
LM	Nível QERCL	#	# de ocorrências de nomes	Nomes		Desvios		
				# de ocorrências	$\bar{X}$	# de ocorrências	$\bar{X}$	%
Espanhol	A1-A2	46	1 280	771	16,76	28	0,61	3,6
	B1-B2	26	933	525	20,19	7	0,27	1,3
	C1	11	549	287	26,09	4	0,36	1,4
Italiano	A1-A2	104	3 141	1 801	17,32	56	0,54	3,1
	B1-B2	40	1 385	734	18,35	18	0,45	2,5
	C1	12	1 699	878	73,17	6	0,50	0,7
Alemão	A1-A2	74	1 784	1 012	13,68	63	0,85	6,2
	B1-B2	60	2 407	1 429	23,82	47	0,78	3,3
	C1	20	928	479	23,95	9	0,45	1,9
Inglês	A1-A2	46	926	538	11,70	45	0,98	8,4
	B1-B2	54	2 783	1 575	29,17	44	0,81	2,8
	C1	29	1 127	598	20,62	24	0,83	4
Chinês	A1-A2	37	957	498	13,46	34	0,92	6,8
	B1-B2	34	1 221	710	20,88	45	1,32	6,3
	C1	10	343	186	18,60	5	0,5	2,7
$\Sigma$		<b>603</b>	<b>21 463</b>	<b>12 021</b>	<b>19,94</b>	<b>435</b>	<b>0,72</b>	<b>3,37</b>

**Quadro 4.12** – Distribuição do número absoluto e dos valores de média de ocorrências de nomes da **categoria 3** produzidos no *corpus* e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QERCL)

Como se pode observar no Quadro 4.12, os nomes cuja atribuição de género não depende de critérios semânticos, mas apenas de critérios formais, são muito frequentes no *corpus* analisado. Efetivamente, constituem mais de metade de todas as ocorrências nominais registadas em SN nos textos seleccionados. Este elevado índice de ocorrências já seria expectável, atendendo ao facto de este tipo de itens ser muito frequente em português (cf. Capítulo 1, Quadro 1.3). Consequentemente, a um maior número de ocorrências associa-se um maior número de desvios de atribuição de género em todos os segmentos da amostra, comparativamente ao apurado nas categorias 1 e 2 (cf. Secções 4.3.1.1. e 4.3.1.2.).

No Gráfico 4.17, consta a distribuição das percentagens relativas de desvios apuradas. Os dados cartografados no Gráfico 4.17 representam os valores percentuais dos

desvios de AGN e CNG calculados em função do número total de ocorrências de itens da categoria 3 produzidos nos textos selecionados de todos os segmentos da amostra de informantes por LM e nível QECRL da turma frequentada (Quadro 4.12). Trata-se, portanto, do peso relativo dos comportamentos desviantes face ao número de ocorrências de nomes nas produções textuais elaboradas por segmento, que compõem o *corpus* do estudo.

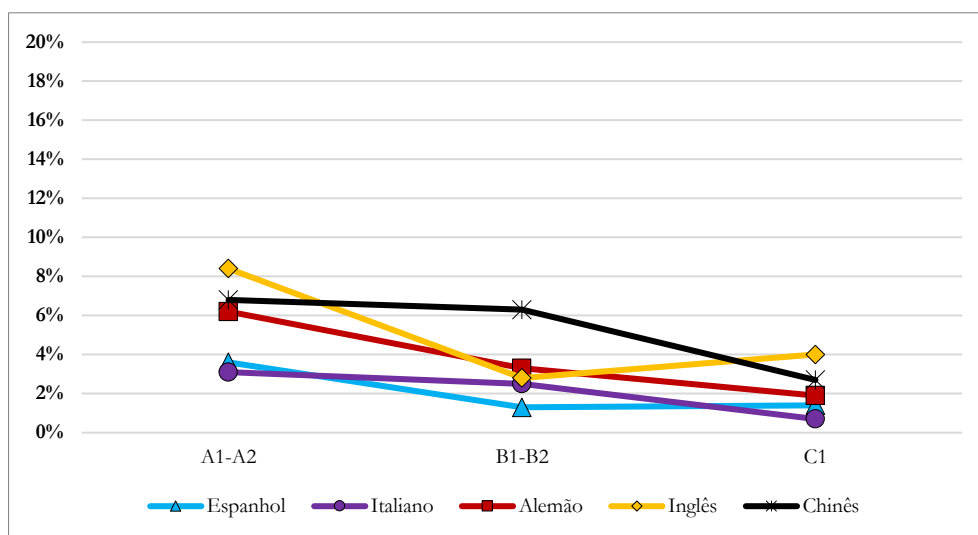


Gráfico 4.17 – Percentagem relativa de desvios da categoria 3 por segmento da amostra (LM e nível QECRL)

Atendendo ao conjunto global de dados representado no Gráfico 4.17, é possível constatar que os falantes aprendentes de LM românicas apresentam índices de desvio consideravelmente mais baixos do que os índices assinalados nos restantes segmentos da amostra por LM, registando igualmente valores da média de desvio por texto inferiores aos registados nos outros segmentos da amostra de informantes (cf. Quadro 4.12). Assinale-se ainda que nestes segmentos, apesar de haver uma diminuição gradual dos desvios dos níveis A1-A2 para o nível C1, os valores das percentagens relativas registados ao longo das diferentes fases do desenvolvimento interlinguístico dos informantes de LM espanhola e italiana são muito próximos (cf. Quadro 4.12).

Os índices de desvio relativamente baixos nos falantes nativos de espanhol e de italiano são, assim, reveladores de uma maior sensibilidade destes informantes aos indícios formais de atribuição dos valores de género nominal, relativamente à que se regista nos restantes segmentos da amostra por LM. Ou seja, parece haver, por parte destes aprendentes, um reconhecimento dos constituintes temáticos dos nomes e da sua correlação parcial com os valores de género disponíveis no português, fruto, talvez, da forte semelhança que a este

respeito se observa nos sistemas de atribuição de género gramatical das respetivas LM (cf. Capítulo 1, Secções 1.4.1. e 1.4.2.).

Para além disso, o facto de haver, nestas línguas, itens lexicais com estruturas formais similares e com o mesmo valor de género, i.e., o facto de haver um grande número de palavras cognatas, com uma mesma origem etimológica, poderá condicionar os aprendentes nas tarefas de atribuição de valor de género a estes itens. Deste modo, é possível que estes aprendentes recorram a uma transferência direta de valores de género no momento da associação do valor ao nome e aos demais constituintes que integram o SN. Por esse motivo, o tratamento destes dados depende igualmente da ponderação do peso relativo de eventuais processos de transferência direta de valores de género, análise que será empreendida na parte final desta subsecção.

Quanto aos falantes de LM alemã, verifica-se a tendência para a diminuição dos desvios da categoria 3 à medida que progridem na aprendizagem do português, a LA. Com efeito, se neste segmento da amostra dos níveis A1-A2 se regista uma percentagem relativa de desvios de, aproximadamente, 6%, nos restantes níveis QECRL as percentagens relativas de desvio vão decrescendo: nos níveis B1-B2, o peso relativo de desvios registado face ao número absoluto de ocorrências de nomes da categoria 3 é de, aproximadamente, 3%, e, no nível C1 é de 1,8%.

A par desta tendência, observa-se igualmente, neste subconjunto de dados, uma diminuição dos valores da média de desvio por texto à medida que se progride no nível QECRL. O facto de estes informantes terem a categoria gramatical de género representada na gramática da sua LM parece auxiliar a assimilação dos valores de género dos nomes em português, pese embora os critérios de atribuição de género serem muito distintos nos dois idiomas, e, portanto, apesar da impossibilidade de transferirem diretamente os valores de género do idioma nativo para a LA.

Ao atingirem o nível C1, os falantes nativos de alemão registam, como se viu, uma percentagem relativa de desvio com valores muito próximos dos apurados nos textos dos falantes nativos de espanhol e de italiano no mesmo nível, para além de se verificarem valores da média de desvio por texto muito próximos entre estes segmentos da amostra (cf. Quadro 4.12). Parece, assim, haver indícios de que a presença da categoria de género na gramática da LM contribui favoravelmente para o processo de assimilação desta estrutura no PLNM. Pode, assim, estar em causa um fenómeno de transferência profunda (cf. Sabourin *et al.* 2006). A consideração de eventuais processos correlacionados com a transferência linguística encontra-se, como já referido, na parte final desta subsecção.



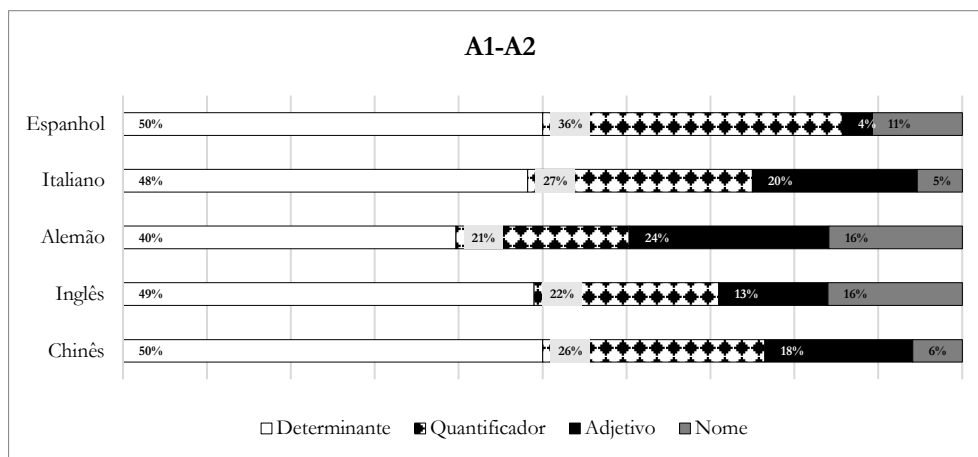
Relativamente ao desempenho dos aprendentes que são falantes nativos de inglês e de chinês, observa-se que estes apresentam, em termos globais, um maior índice de desvios e de valores da média de desvios relativamente aos registados nos restantes segmentos da amostra. Para além disso, e apesar de se registar uma diminuição considerável da percentagem relativa de desvios no nível C1, os índices de desvio da categoria 3 apurados neste nível são sensivelmente mais elevados do que os registados nos restantes segmentos da amostra de informantes por LM.

É preciso, contudo, estabelecer uma distinção entre os segmentos da amostra de informantes de LM inglesa e de LM chinesa. Com efeito, os falantes nativos de inglês registam nos níveis A1-A2 uma percentagem relativa de desvios de, aproximadamente, 8%, valor mais elevado do que o registado nos textos produzidos por falantes nativos de chinês desses níveis (6,8%). Já os valores da média de desvios de AGN e de CNG por texto nestes dois segmentos da amostra são muito próximos. Com efeito, nos níveis A1-A2, nos textos de falantes nativos de inglês, apurou-se um valor médio de desvios da categoria 3 de 0,98, e, nos textos de falantes nativos de chinês, de 0,92. Assim sendo, apurou-se uma média aproximada de 1 item com desvio da categoria 3 por texto produzido nos segmentos da amostra de informantes de LM inglesa e chinesa dos níveis A1-A2. Na transição para os níveis B1-B2, no segmento da amostra de informantes de LM inglesa regista-se uma considerável descida do peso relativo dos casos desviantes, uma vez que de 8,4% nos níveis A1-A2 passa a 2,8%, valor justificável, em parte, pelo aumento do número de ocorrências dos itens da categoria 3 em textos de aprendentes dos níveis B1-B2 (cf. Quadro 4.12).

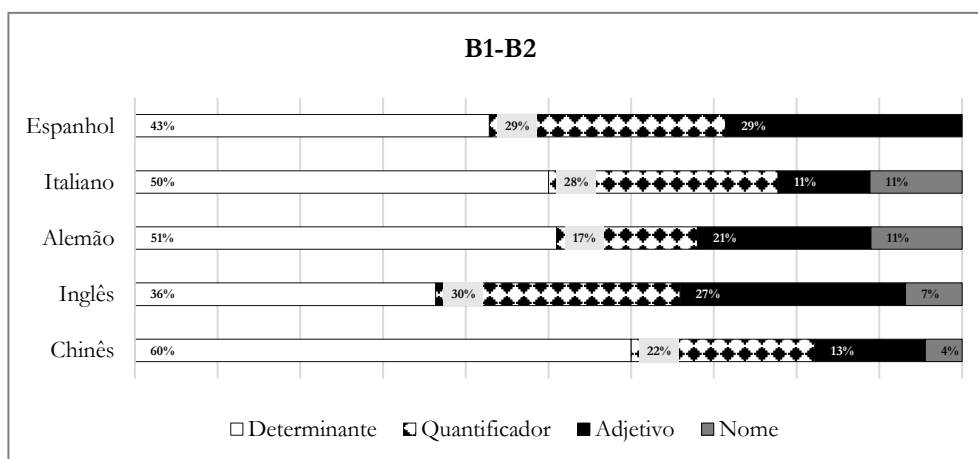
Em contrapartida, nos segmentos da amostra de falantes nativos de chinês regista-se somente uma ligeira descida do peso relativo de desvios na passagem dos níveis A1-A2 para os níveis B1-B2, de 6,8% para 6,3%, respetivamente. Em relação aos valores da média de nomes da categoria 3 com desvios por texto assinalam-se, igualmente, diferenças, visto que, nos textos escritos por informantes que são falantes nativos de inglês de B1-B2, foi apurado um valor da média de desvios de 0,82, enquanto que nos textos de falantes nativos de chinês dos mesmos níveis se apurou um valor médio de desvios ligeiramente superior, i.e., 1,32.

Por sua vez, em C1, regista-se a situação inversa: a proporção do número de desvios face às ocorrências de itens da categoria 3 é ligeiramente mais alta nos textos produzidos por informantes de LM inglesa (4%) do que a proporção registada nos textos escritos por falantes nativos de chinês (2,7%).

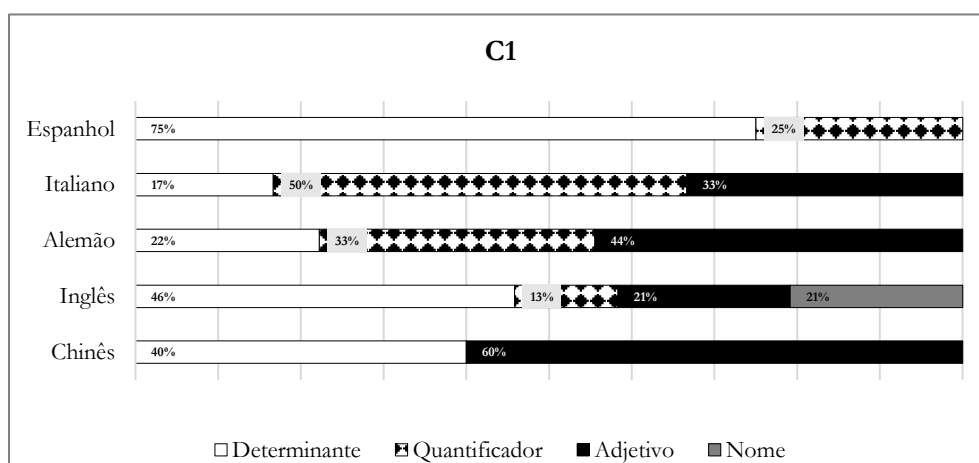
Os Gráficos 4.18, 4.19 e 4.20 registam, em valores percentuais, a proporção de desvios de AGN e de CNG da categoria 3 por constituinte afetado, atendendo à LM e ao nível QECRL da turma frequentada pelos aprendentes.



**Gráfico 4.18** – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos nos níveis A1-A2



**Gráfico 4.19** – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos nos níveis B1-B2



**Gráfico 4.20** – Percentagem de desvios da categoria 3, distinguidos por classe de palavras afetadas, produzidos no nível C1

Conforme é possível constatar nos Gráficos 4.18 a 4.20, nos níveis A1-A2 e B1-B2, a maior proporção de desvios da categoria 3 incide sobre os especificadores, nomeadamente sobre os determinantes, sendo que, na maioria dos segmentos da amostra por LM este subtipo de desvios representa, em termos proporcionais, metade dos desvios detetados. Os exemplos seguintes ilustram casos de CNG desviante no determinante:

- (46) «\*Porque **esta tempo** em Coimbra eu posso aprender mais facil (...)»  
(Alemão.A1-A2)
- (47) «\*O tempo de Macau /no Verão/ é muito húmido e <carol> calor, por isso nós precisamos ir **ao praia**, (...)» (Chinês.A1-A2)
- (48) «\*Quando era pequena (...) eu brincava com **os meus brincadeiras** (...)»  
(Italiano.B1-B2)
- (49) «\*Depois de lavar **meus roupas**, ir de compras (...)» (Espanhol.B1-B2)

Da análise dos Gráficos 4.18 a 4.20, é possível constatar ainda que entre alguns segmentos da amostra de informantes por LM, se observam padrões idênticos relativamente à proporção dos desvios da categoria 3 por constituinte afetado no SN. Saliente-se também uma tendência para o aumento da proporção dos desvios de CNG nos adjetivos da passagem dos níveis A1-A2 para os níveis B1-B2 em alguns segmentos, nomeadamente entre os falantes de LM espanhola.

Já no nível C1 (cf. Gráfico 4.20), os textos dos falantes nativos de espanhol não apresentam qualquer desvio nos adjetivos. Em contrapartida, nos restantes segmentos, observa-se um aumento da proporção de desvios nos adjetivos relativamente à proporção registada em níveis inferiores. Aliás, nos textos escritos por falantes de LM chinesa deste nível QECRL, os adjetivos apresentam uma maior proporção de desvios relativamente à registada nos determinantes.

Como já referido na Secção 4.2.3.1., é de salientar que uma parte substancial dos desvios nos adjetivos integram situações de ambivalência da atribuição do valor de género nominal, em que, por exemplo, diante de um nome feminino, o aprendente seleciona o especificador de género feminino, mas usa a forma do masculino do adjetivo:

- (50) «\*Nas nossas férias, só passeámos nas ruas limpos, visitámos o símbolo da cidade.» (Chinês.C1)
- (51) «\* (...) e foi uma **experiância** muito **novo** para mim de ver (...)»  
(Alemão.A1-A2)
- (52) «\*Era uma **ideia** muito **bom!**» (Espanhol.B1-B2)

- (53) «\* Só queria visitar minha mãe que fica doente mas porque os problemas com sua (agência) **mau empresa** não possui se você quer manter (...) eu como cliente». (Inglês.C1)

Numa segunda fase da análise dos dados, fez-se o levantamento dos nomes afetados, tendo em conta, quer o número absoluto das suas ocorrências nos textos produzidos por cada segmento da amostra, quer o número de ocorrências de desvios de AGN e de CNG. Pretende-se, assim, dar conta do léxico afetado por segmento, a fim de averiguar se há incidência de desvios sobre um conjunto de itens lexicais em particular, por um lado, e, por outro, se é possível aferir uma relação entre a proporção de desvios com a própria estrutura morfológica do nome, i.e., se a maior proporção de desvios incide sobre formas nominais simples ou complexas, sejam portadoras de afixos ou compostas.

No Quadro 4.13 apresenta-se o registo quantitativo dos desvios da categoria 3 distribuídos por LM e nível QECRL da turma frequentada pelo informante, com a indicação, por segmento da amostra:

- (i) do número de nomes da categoria 3 com desvios (**# de nomes afetados**);
- (ii) do número de ocorrências desses nomes afetados no *corpus* (**# de ocorrências dos nomes afetados no corpus**);
- (iii) do número de desvios da categoria 3 (**# de desvios**);
- (iv) do valor da média de desvios da categoria 3 por nome afetado ( **$\bar{X}$  de desvios**);
- (v) da percentagem relativa dos desvios, calculada em função do respetivo número absoluto de ocorrências dos nomes afetados (**% de desvios**).

Informantes		Categoria 3				
LM	Nível QERCL	# de nomes afetados	# de ocorrências dos nomes afetados no <i>corpus</i>	# de desvios	$\bar{X}$ de desvios	% de desvios
Espanhol	A1-A2	21	165	28	1,3	16,9
	B1-B2	7	51	7	1	13,7
	C1	3	10	4	3,3	40,0
Italiano	A1-A2	38	515	56	1,5	10,9
	B1-B2	13	161	18	1,4	11,2
	C1	6	41	6	1	14,6
Alemão	A1-A2	46	369	63	1,4	17,1
	B1-B2	35	503	47	1,3	9,3
	C1	6	58	9	1,5	15,5
Inglês	A1-A2	28	194	45	1,6	23,2
	B1-B2	33	385	44	1,3	11,4
	C1	21	56	24	1,1	42,9
Chinês	A1-A2	22	119	34	1,5	28,6
	B1-B2	29	170	45	1,6	26,5
	C1	5	21	5	1	23,8
$\Sigma$		313	2 818	435	1,4	15,4

**Quadro 4.13** – Distribuição do número de nomes afetados, do respetivo número de ocorrências no *corpus* e do número de desvios da categoria 3

Portanto, no conjunto global de textos que compõem o *corpus* selecionado, apurou-se, na categoria 3, um total de 435 ocorrências de desvios de atribuição de valor de género e de concordância nominal em género incidentes sobre 313 itens nominais<sup>167</sup>, que correspondem a um total de 2818 ocorrências no *corpus* selecionado. Entre este conjunto de nomes afetados, verificou-se ainda que os desvios incidem sobre 195 itens nominais distintos, tratando-se, assim, de valores muito superiores aos registados nas categorias anteriormente analisadas (categoria 1 e 2).

Conforme se pode verificar no Quadro 4.13, o peso relativo dos desvios de AGN e de CNG face ao número de ocorrências dos nomes afetados no *corpus* por segmento da amostra é, salvo alguns casos pontuais, baixo, i.e., há uma maior taxa de acertos do que de desvios entre os nomes assinalados. Para além disso, nem sempre se verifica nos diferentes segmentos da amostra uma descida dos valores das percentagens relativas de desvios dos níveis A1-A2 para os restantes níveis. Com efeito, no nível C1, há, por vezes, uma maior proporção de desvios relativamente ao número absoluto de ocorrências dos nomes afetados registado nos textos selecionados, como é visível nos segmentos da amostra de informantes de LM espanhola, italiana e inglesa. Tal tendência não se verifica, contudo, nos segmentos da amostra de informantes que são falantes nativos de alemão e de chinês.

Veja-se no Quadro 4.14 a distribuição dos 313 nomes com desvios da categoria 3 por segmento da amostra (LM e nível QECRL da turma frequentada). Note-se ainda que os itens foram organizados por ordem decrescente do número de desvios a que deram origem.

---

<sup>167</sup> Recorde-se que o maior número de desvios em relação ao número de nomes justifica-se pelo facto de, por vezes, num SN se terem registado desvios em diferentes constituintes, nos especificadores, modificadores e nomes. Deste modo, cada constituinte com marcas desviantes é assinalado como um desvio.

Nível QERCL												
LM	A1-A2					B1-B2					C1	
	Nomes (# de ocorrências no <i>corpus</i> / # de desvios)					Nomes (# de ocorrências no <i>corpus</i> / # de desvios)					Nomes (# de ocorrências no <i>corpus</i> / # de desvios)	
Espanhol	<p><b>peessoa (20/3)</b> sexta-feira (2/2) aldeia (4/2) casa de banho (17/2) equipa (2/2)</p>	<p>morada (6/2) arranjo (1/1) assoalhada (2/1) bancada (2/1)</p>	<p>cadeira (7/1) casa (57/1) cena (2/1) copo (4/1)</p>	<p>discoteca (2/1) estatura (1/1) hora (2/1) momento (6/1)</p>	<p>ortografia (1/1) sanita (6/1) semana (7/1) vida (14/1)</p>	<p>ano (22/1) campanha (1/1) coisa (19/1) hora (2/1)</p>	<p>ideia (3/1) jogo (3/1) roupa (1/1)</p>	<p><b>casaco (2/2)</b> carnificina (1/1) vida (7/1)</p>				
Italiano	<p><b>peessoa (70/6)</b> comida (29/4) casa de banho (14/3) desculpa (2/3) língua (8/3) rádio (3/3)</p>	<p>casa (107/2) coisa (39/2) altura (7/1) ano (91/1) atmosfera (1/1) bacia (2/1) cadeira (5/1) cama (16/1)</p>	<p>casino (1/1) cómoda (1/1) curso (15/1) energia (2/1) entrada (3/1) fogueira (2/1) fundamento (1/1) garganta (1/1)</p>	<p>janela (12/1) mala (4/1) mancira (3/1) manta (2/1) meio de transporte (23/1) mesa (16/1) mesa de cabeceira (2/1) montanha (6/1)</p>	<p>olho (16/1) peça (1/1) pergunta (1/1) pimento (1/1) prato (5/1) prédio (2/1) receita (1/1) terra (2/1)</p>	<p><b>coisa (31/3)</b> <b>loja (6/3)</b> comida (14/2) ajuda (2/1) brincadeira (2/1) campo (21/1) casa (27/1)</p>	<p>ditadura (1/1) história (11/1) limpeza (1/1) peessoa (38/1) prejuizo (1/1) semana (6/1)</p>	<p><b>peessoa (22/1)</b> bilheteira (1/1) forma (4/1) peessoa (5/1) loja (4/1) meio de comunicação (5/1)</p>				
Alemão	<p><b>peessoa (45/6)</b> cerveja (7/4) coisa (31/4) equipa (5/3) escola (8/2) experiência (5/2) vida (21/2) ano (52/1) armário (4/1)</p>	<p>bolcia (1/1) bolsa (1/1) brincadeira (1/1) cabelo (8/1) caipirinha (1/1) canto (2/1) casino (1/1) criança (7/1) data (1/1)</p>	<p>desporto (6/1) dinheiro (12/1) estudo (4/1) família (16/1) feriado (1/1) folha (5/1) grupo (8/1) história (5/1) hora (14/1)</p>	<p>horário (1/1) medicamento (2/1) método (1/1) mundo (12/1) palmeira (1/1) passoio (3/1) película (1/1) piscina (4/1) plano (2/1)</p>	<p>remédio (1/1) ressaca (1/1) trabalho (11/1) técnica (1/1) tempo (40/1) testemunha (2/1) vista (4/1) vitória (2/1) volta (4/1)</p>	<p><b>criança (60/5)</b> <b>peessoa (12/3)</b> <b>semana (25/3)</b> experiência (10/2) loja (6/2) movimento (4/2) peessoa (75/2) silêncio (3/2) alma (2/1)</p>	<p>ano (58/1) apoio (1/1) caipirinha (1/1) casa (49/1) conjunto (3/1) cópia (2/1) cultura (17/1) diferença (6/1)</p>	<p>energia (3/1) estado (8/1) estrangeiro (4/1) existência (1/1) floresta (2/1) força (3/1) indústria (3/1) lado (21/1) passatempo (3/1)</p>	<p>ponto (5/1) ponto de vista (3/1) praça (3/1) praia (21/1) raposa (5/1) tempo (66/1) trabalho (8/1) violência (10/1)</p>	<p><b>coisa (12/2)</b> <b>palavra (5/2)</b> <b>peessoa (27/2)</b> aventura (1/1) circunstância (1/1) praia (12/1)</p>		
Inglês	<p><b>desporto (3/4)</b> <b>casa (33/3)</b> ano (25/2) coisa (16/2) comida (6/2)</p>	<p>cultura (5/2) equipa (4/2) história (12/2) festa (6/2) junta de freguesia (1/2) paraíso (2/2)</p>	<p>peessoa (16/2) semana (10/2) cabelo (13/1) colégio (1/1) copa (taça) (1/1) discoteca (4/1) estado (2/1) exemplo (4/1)</p>	<p>experiência (2/1) feriado (1/1) fortaleza (1/1) hora (10/1) janela (1/1)</p>	<p>lado (2/1) loja (2/1) monumento (7/1) taça (1/1) supermercado (3/1)</p>	<p><b>coisa (42/3)</b> <b>peessoa (64/3)</b> <b>aldeia (11/2)</b> <b>aventura (3/2)</b> <b>história (22/2)</b> <b>polícia (12/2)</b> aeroporto (3/1) camada (3/1) casa (57/1)</p>	<p>cena (5/1) cerimónia (1/1) cerveja (2/1) certeza (2/1) colónia (3/1) consciência (1/1) cortejo (5/1) cozinha (2/1)</p>	<p>descendência (2/1) discoteca (7/1) família (25/1) maioria (7/1) hora (11/2) loja (11/1) perna (1/1) pré-independência (1/1)</p>	<p>primavera (1/1) semana (11/1) subúrbio (3/1) teoria (2/1) uso (4/1) vida (54/1) vila (1/1) vocabulário (1/1)</p>	<p><b>orquestra (2/3)</b> <b>diferença (11/2)</b> aparência (1/1) área (4/1) carta (5/1) confiança (2/1) culpa (1/1)</p>	<p>data (1/1) deserto (3/1) empresa (2/1) força (1/1) frio (2/1) fronteira (3/1) grupo (4/1)</p>	<p>injustiça (1/1) mirada (1/1) mundo (6/1) proposta (1/1) sobremesa (1/1) taxa (1/1) tentativa (1/1)</p>
Chinês	<p><b>língua (14/4)</b> coisa (10/3) peessoa (22/3) comida (4/2) companhia (1/2)</p>	<p>marca (2/2) praça (2/2) terra (3/2) apartamento (2/1) aula (6/1)</p>	<p>biblioteca (1/1) casa de banho (2/1) copa(taça) (1/1) garrafa (1/1)</p>	<p>loja (10/1) máscara (1/1) praia (4/1) roupa (8/1)</p>	<p>rua (6/1) salário (1/1) sexta-feira (2/1) trabalho (16/1)</p>	<p><b>roupa (8/6)</b> <b>cara (3/4)</b> <b>porta (41/3)</b> altura (4/2) coisa (20/2) comida (8/2) feriado (3/2) história (4/2)</p>	<p>promessa (1/2) amendoieira (1/1) armário (1/1) cultura (11/1) distância (1/1) dureza (1/1) empresa (8/1)</p>	<p>esquerda (1/1) lembrança (1/1) janela (1/1) massa (2/1) paraíso (1/1) peessoa (23/1) pintura (6/1)</p>	<p>regra (3/1) residência (1/1) segunda-feira (5/1) semana (5/1) serra (1/1) terra (4/1) vício (1/1)</p>	<p>coisa (7/1) discoteca (1/1) forma (2/1) peessoa (9/1) rua (2/1)</p>		

Quadro 4.14 – Distribuição de nomes afetados da categoria 3, com a indicação do número absoluto de ocorrências no *corpus* e do respetivo número absoluto de desvios

A partir do Quadro 4.14, observa-se, em linhas gerais, uma elevada diversidade lexical de itens nominais afetados, sendo que, entre os falantes nativos de línguas românicas se regista uma diminuição acentuada desta diversidade lexical na passagem dos níveis A1-A2 para os níveis B1-B2 e destes para C1. Esta tendência já não se verifica tão acentuadamente nos restantes segmentos da amostra por LM.

Para além disso, e numa primeira abordagem aos dados, constata-se que a estrutura morfológica do nome não afeta o maior ou menor índice de desvios registados nesta categoria nos vários segmentos da amostra, já que os casos assinalados incidem, sobretudo, em formas nominais simples, sendo pouco frequentes os desvios de AGN e de CNG sobre formas nominais complexas, i.e., portadoras de afixos ou compostas. Entre os compostos, assinalem-se as ocorrências de desvios nos sintáticos, como, por ex., *casa de banho*, *ponto de vista* e *meio de transporte*, já que, nestes, o facto de haver dois nomes com constituintes temáticos distintos poderá ter condicionado os aprendentes no momento da atribuição dos respetivos valores de género.

Tendo em conta a distribuição dos desvios por nível QECRL, registam-se, em todos os segmentos da amostra por LM, um maior número de desvios nos níveis A1 e A2. Nestes níveis, os itens nominais afetados correspondem, predominantemente, a formas simples, sendo que neste segmento da amostra apenas um nome surge recorrentemente com desvio em todos os grupos de informantes por LM. Trata-se do item *pessoa* que, para além disso, deu origem a um número considerável de desvios, sobretudo nas produções escritas por falantes nativos de espanhol, italiano e alemão a frequentar turmas de A1-A2. Ou seja, são recorrentes os desvios sobre este item nas produções dos informantes cuja LM possui a categoria de género gramatical. Já no trabalho desenvolvido por Leiria (2006), a autora identificara o sobrecomum *pessoa* como um dos itens mais propícios à ocorrência de desvios de atribuição de valores de género nominal por aprendentes tardios de PLNM. A incidência de desvios nos sobrecomuns leva Leiria (2006:241-242) a considerar que:

“no léxico mental de alguns aprendentes, os nomes sobrecomuns (...) estão representados como se fossem comuns de dois (*o/a pessoa*, *o/a criança*, como *o/a artista* ou *o/a turista*). No caso dos comuns de dois, sendo o masculino em português, tal como nas outras línguas românicas, o valor não-marcado, é ele que é seleccionado sempre que se pretende obter uma referência genérica”.

Com efeito, e à semelhança do que observa Leiria, também nos dados aqui apurados há, efetivamente, uma tendência para associar o valor de género masculino a estes itens,

sugerindo-se, mais uma vez, que no momento de atribuição de valores de género, os indícios semânticos predominam sobre os indícios formais:

- (52) «\*Eu moro com **outros** <(…)> três **pessoas** <(…)> no <Ba> bairro muito bonito perto da escola de Português.» (Espanhol.A1-A2)
- (53) «\* (...) é falar com **estes** **pessoas**: posso descobrir características e habituais do país que antes não sabia.» (Italiano.A1-A2)
- (54) «\*Queria falar com os Portugueses e **os** outros **pessoas** de Lisboa para saber como é a vida aqui e quais são os problemas que a população encontra todas as dias.» (Alemão.A1-A2)
- (55) «\* A veces fui agitada porque fue **muitos** **pessoas**.» (Inglês.A1-A2)
- (56) «\* Mas houve **uns** **pessoas** na restaurante...» (Chinês.A1-A2)

Assinalemos, em seguida, em cada segmento da amostra (por LM e nível QECRL), os casos em que o número de ocorrências dos nomes afetados pelos desvios e do número de desvios propriamente ditos são os mesmos. Ou seja, nas situações em que, aquando do uso de um nome por parte dos informantes, a sua utilização resultou sempre em desvio de AGN e de CNG (cf. Quadro 4.14).

Nos níveis A1 e A2, no segmento da amostra dos informantes, que são falantes nativos de espanhol, entre os 21 nomes afetados, 5 apresentam o mesmo número de ocorrências e de desvios: *sexta-feira* (2/2), *equipa* (2/2), *arranjo* (1/1), *estatura* (1/1) e *ortografia* (1/1).

Entre os falantes nativos de italiano, do conjunto total de 38 nomes afetados, 10 possuem igual número de ocorrências nos textos selecionados e de desvios: *rádio* (3/3), *atmosfera* (1/1), *casino* (1/1), *cómoda* (1/1), *fundamento* (1/1), *garganta* (1/1), *peça* (1/1), *pergunta* (1/1), *pimento* (1/1) e *receita* (1/1). Há ainda, neste segmento, 1 nome, *desculpa*, cujo número de desvios é maior do que o número de ocorrências do item. Tal advém do facto de em relação a este item se registarem desvios expressos, não só nos especificadores e modificadores, mas também na própria forma morfológica do nome<sup>168</sup>:

- (57) «\*O que custou o maior esforço para sair desta situação foi a vitória sobre a resignação que tinha dominado o Senhor Costa e que o embalava **no seu novo desculpe** (...)» (Italiano.A1-A2)

---

<sup>168</sup> Sobre o número de desvios por nome, veja-se no Capítulo 3, Secção 3.4.2. os critérios adotados na presente investigação para a contagem dos desvios assinalados nos textos.



Nas produções escritas por falantes nativos de alemão, verificou-se que, entre os 46 nomes afetados, 14 apresentam o mesmo número de desvios e de ocorrências: *boleia* (1/1), *bolsa* (1/1), *brincadeira* (1/1), *caipirinha* (1/1), *casino* (1/1), *data* (1/1), *feriado* (1/1), *horário* (1/1), *método* (1/1), *palmeira* (1/1), *película* (1/1), *remédio* (1/1), *ressaca* (1/1) e *técnica* (1/1).

Quanto aos falantes nativos de inglês, verificou-se que, em 29 itens nominais, 7 possuem igual número de ocorrências e de desvios: *paraíso* (2/2), *colégio* (1/1), *copa* (1/1), *feriado* (1/1), *fortaleza* (1/1), *janela* (1/1) e *taça* (1/1). Neste segmento da amostra também se registou um item nominal composto, *junta de freguesia*, com maior número de desvios do que de ocorrências. Novamente, trata-se de um caso de alteração morfológica do nome nuclear e de seleção incorreta do valor de género nos demais constituintes que integram o SN:

- (58) «\*Depois <a> **o junto da freguesia** mandou erigir um relógio em frente da casa.» (Inglês.A1-A2)<sup>169</sup>

Por fim, entre as produções escritas por falantes nativos de chinês dos níveis A1-A2, registam-se, entre os 22 nomes afetados, 7 que são usados sempre com valor desviante de género: *marca* (2/2), *praça* (2/2), *biblioteca* (1/1), *copa* (1/1), *garrafa* (1/1), *máscara* (1/1) e *salário* (1/1). À semelhança do que se verificara nos segmentos dos informantes que são falantes nativos de espanhol e dos informantes de LM inglesa, também aqui se registou a ocorrência de um nome, *companhia*, com um maior número de desvios do que de ocorrências, circunstância que decorre dos motivos já atrás assinalados:

- (59) «\*Acho que Pingo doce é **um bom companhia**, os vegetais e frutas são fresco e (care) caro.» (Chinês.A1-A2)

Nos níveis B1-B2, regista-se, em grande parte dos segmentos da amostra por LM, uma menor diversidade lexical dos itens afetados, sobretudo nas produções de falantes nativos das línguas românicas.

Considerando, em primeiro lugar, os aprendentes de LM espanhola, constata-se que nos níveis B1-B2, dos 7 itens afetados, apenas o nome *hora* havia surgido com desvio nos níveis A1-A2, i.e., a maioria dos nomes afetados nos níveis anteriores não reaparece com

---

<sup>169</sup> Este exemplo revela também a hesitação por parte do aprendente na atribuição do valor de género. Com efeito, num primeiro momento, o informante optou pela forma feminina do determinante 'a', apesar de, em seguida, a ter emendado erroneamente, rasurando-a e substituindo-a pela forma masculina.

desvio nos níveis seguintes, o que poderá denotar a aprendizagem do respetivo valor de género. Assinale-se ainda que o número de desvios por item nominal é inferior ao registado nos níveis A1 e A2. Para além disso, dos 7 nomes afetados, somente 2 itens, *campanha* (1/1) e *roupa* (1/1), registam igual número de ocorrências e de desvios.

Analisando, agora, os nomes afetados da categoria 3 nas produções escritas por falantes de LM italiana, observa-se que, dos 13 nomes afetados, apenas 2 surgem também com desvios nos níveis A1-A2 (*coisa* e *pessoa*). Para além disso, regista-se igualmente uma tendência para a diminuição do número de desvio por nome. Entre os 13 itens afetados, 3 registam o mesmo número de ocorrências e de desvios: *ditadura* (1/1), *limpeza* (1/1) e *prejuízo* (1/1).

Passando, agora, à análise dos desvios detetados nas produções escritas por falantes nativos de alemão a frequentar turmas de B1-B2, verifica-se que do conjunto total de 35 nomes afetados, 6 que já tinham surgido com desvio nos níveis anteriores (A1-A2), a saber: *ano*, *caipirinha*, *criança*, *experiência*, *pessoa* e *tempo*. Entre este subconjunto de nomes assinale-se ainda que, em grande parte dos casos, se regista uma diminuição do número de desvios nos níveis B1-B2 por item face ao número de desvios registado nos níveis A1-A2. Ou seja, e no que a estes nomes diz respeito, parece haver, em termos globais, uma melhoria do desempenho dos aprendentes na atribuição dos valores de género à medida que progridem na aprendizagem do português.

Convém, no entanto, assinalar que o sobrecomum *criança* regista, por sua vez, um aumento do número de desvios nos níveis B1-B2 em relação ao valor registado em A1-A2, o que poderá estar correlacionado com o facto de, como já vimos, haver uma tendência para os aprendentes representarem, no seu léxico mental, estes nomes como comuns de dois, seleccionando, conseqüentemente, a forma não-marcada do masculino como referência genérica da espécie. Para além disso, no alemão, o nome *Kind* ‘criança’ é de género neutro<sup>170</sup>, o que pode ter conduzido os aprendentes a assumir que o item correspondente em português possui um valor de género que permite uma leitura genérica, i.e., o masculino.

---

<sup>170</sup> Com efeito, e tal como se verificou no Capítulo 1, Secção 1.4.3.2., em alemão é possível encontrar um subconjunto de itens com referentes [+sexuados] de género neutro, tais como *Kind* (criança) e *Mädchen* (menina jovem). Segundo Zubin & Köpcke (1986), a associação do neutro a estes nomes poderá estar correlacionada com o facto de se assumir que as respetivas entidades designadas são assexuadas, por ainda não terem atingido a maturidade sexual. Assim, e nas palavras dos autores: “*Masc* and *fem*-gender mark the terms for the male and female adult of each species, while *neut*-gender is assigned to the nonsexspecific generic and juvenile terms” (1986:175).

Quanto ao número de desvios apurado por nome, regista-se, igualmente, uma ligeira diminuição dos casos com a progressão do nível QECRL. No total do número de nomes afetados nas produções escritas por informantes de LM alemã a frequentar turmas dos níveis B1-B2, apenas 3 apresentam o mesmo número de ocorrências e de desvios: *apoio* (1/1), *caipirinha* (1/1), e *existência* (1/1).

Entre o segmento da amostra de falantes nativos de idiomas sem sistema de classificação de nomes baseado em valores de género, regista-se, ainda que ligeiramente, uma maior diversidade dos itens lexicais afetados nos níveis B1-B2 face à registada em A1-A2, contrariando, assim, a tendência assinalada nos outros segmentos da amostra por LM.

Nos casos apurados nas produções escritas por falantes nativos de inglês, observa-se que dos 33 itens nominais afetados, 6 também apresentavam desvios nos textos dos níveis A1-A2 (*casa, coisa, história, hora, loja e pessoa*), não havendo, contudo, uma diferença acentuada do número de desvios por cada um dos itens. Relativamente às formas nominais com o mesmo número de ocorrências e de desvios, registaram-se, nos níveis B1-B2, 6 nomes nesta situação, a saber: *cerimónia* (1/1), *consciência* (1/1), *perna* (1/1), *primavera* (1/1), *vila* (1/1) e *vocabulário* (1/1).

Por fim, no que concerne aos desvios apurados no segmento da amostra de falantes nativos de chinês dos níveis B1-B2, entre os 29 itens afetados, encontram-se 4 nomes que já haviam surgido com desvio nos níveis A1-A2: *coisa, comida, pessoa e roupa*. Assinala-se ainda neste segmento da amostra uma ligeira subida do número de nomes com o mesmo número de ocorrências e de desvios nos níveis B1-B2, face ao registado em A1-A2, contrariando, assim, a tendência observada em outros segmentos da amostra por LM. Com efeito, entre o conjunto total de 29 formas nominais afetadas, 11 apresentam o mesmo número de ocorrências e de desvios, tratando-se, portanto, de um subconjunto ligeiramente superior ao registado nos níveis A1-A2 do segmento da amostra dos informantes de LM chinesa: *amendoeira* (1/1), *armário* (1/1), *distância* (1/1), *dureza* (1/1), *esquerda* (1/1), *lembrança* (1/1), *janela* (1/1), *paraíso* (1/1), *residência* (1/1), *serra* (1/1) e *vício* (1/1). Nos níveis B1-B2, assinala-se ainda que o nome *promessa* apresenta um maior número de desvios do que de ocorrências, pelo facto de se registarem desvios, quer na forma morfológica do nome, quer nos especificadores que integram o SN:

- (60) «\*Ela não deve lavar a sua cara com lágrimas (...) **pelo seu promessa** com a sua vida toda» (Chinês.B1-B2)

Por fim, no nível C1 é notória não só uma diminuição do número de desvios relativamente ao registado nos níveis precedentes, como uma menor diversidade dos itens lexicais afetados, em todos os segmentos da amostra por LM.

Entre os falantes nativos de espanhol, registam-se apenas 3 itens com desvios, sendo que, em dois desses itens, o número de ocorrências nos textos selecionados da amostra e de desvios é igual: *casaco* (2/2) e *carnificina* (1/1). Nenhuma destas formas dera origem a desvios nos textos selecionados de outros níveis QECRL.

No segmento da amostra de informantes de LM italiana, há 6 nomes afetados, sendo que somente *bilheteira* apresenta o mesmo número de desvios e de ocorrências. Acrescente-se ainda que, neste segmento da amostra, apenas *pessoa* reaparece com desvio nas produções escritas de outros níveis QECRL o que poderá indicar, assim, para os restantes casos a aprendizagem do valor de género.

Relativamente ao segmento da amostra de informantes que são falantes nativos de alemão, identificaram-se 6 nomes afetados, sendo que, desse conjunto, 3 reaparecem com desvios nos restantes níveis analisados: *coisa*, *pessoa* e *praia*. Entre o total de nomes afetados, 2 apresentam o mesmo número de ocorrências e de desvios: *aventura* (1/1) e *circunstância* (1/1), tratando-se, para além disso, de itens lexicais que, como vimos, não surgiram com desvios nos textos produzidos por este segmento da amostra em outros níveis QECRL.

Passando, em seguida, aos dados assinalados no segmento da amostra de informantes de LM inglesa a frequentar turmas de nível QECRL C1, é notória a diversidade de nomes afetados, mantendo-se, neste segmento, um padrão relativamente homogéneo em todos os níveis QECRL (cf. Quadros 4.13 e 4.14). Poder-se-á, então, inferir que, mesmo em C1, estes aprendentes apresentam manifestas dificuldades na associação dos valores de género aos nomes da categoria 3, bem como no estabelecimento das relações de concordância nominal.

No entanto, convém assinalar que o número de desvios por item é inferior ao registado nos outros níveis anteriormente analisados. Para além disso, do conjunto total de 21 nomes afetados no nível C1, nenhum apresentara desvios nos níveis A1-A2 e B1-B2. Há ainda, neste segmento da amostra, 10 itens com o mesmo número de desvios e de ocorrências: *aparência* (1/1), *culpa* (1/1), *data* (1/1), *força* (1/1), *injustiça* (1/1), *mirada* (1/1), *proposta* (1/1), *sobremesa* (1/1), *taxa* (1/1) e *tentativa* (1/1).

Por fim, entre o segmento da amostra de informantes de LM chinesa de nível C1, é assinalável, quer a diminuição da diversidade de nomes afetados quer a diminuição do

número de desvios por item nominal. Além disso, entre os 5 nomes afetados, 3 já reapareceram com desvios nos restantes níveis analisados (*coisa, pessoa e rua*) e apenas 1 item apresenta o mesmo número de desvios e de ocorrências: *discoteca* (1/1).



Num último momento da análise, procurou-se averiguar o eventual efeito da transferência linguística na assimilação dos valores de género nominal em PLNM. Para tal, consideraram-se os desvios assinalados nos segmentos da amostra de informantes de LM com a categoria de género gramatical, i.e., falantes nativos de espanhol, italiano e alemão. Como já assinalado nesta monografia, entre os idiomas selecionados com género gramatical, as línguas românicas (espanhol e italiano) apresentam um sistema de classificação dos nomes baseado em dois valores de género (masculino e feminino) com particularidades muito próximas do português, quer no que concerne à atribuição dos valores de género aos nomes quer no que diz respeito ao estabelecimento de relações de concordância sintática (veja-se no Capítulo 1, Secções 1.4.1. e 1.4.2.). Por sua vez, a língua alemã possui um sistema de género tripartido, com três valores em oposição: masculino, feminino e neutro (cf. Capítulo 1, Secção 1.4.3.), com um sistema de concordância sintática complexo, já que se estabelecem relações de concordância nominal em género, número e caso.

Assim sendo, atendendo às características dos sistemas de classificação nominal do espanhol, italiano e alemão, procurar-se-á averiguar, entre os desvios de AGN e de CNG assinalados nos vários segmentos da amostra de informantes que são falantes nativos destes idiomas, manifestações de uma eventual transferência linguística dos valores de género dos nomes da LM para os itens equivalentes na LA, o português.

Consideremos, em primeiro lugar, os comportamentos desviantes assinalados nas produções escritas por falantes nativos das línguas românicas, nos diferentes níveis QECRL. Neste conjunto de dados, averiguou-se, em cada caso, qual seria a forma lexical correspondente na LM, assinalando-se também o respetivo valor de género. Assim, em função dessa análise, sempre que se verificara a não correspondência de valores de género entre os itens, considerou-se que os desvios nesses nomes são potencialmente atribuíveis à transferência linguística. Convém, no entanto, ressaltar que nem sempre foi possível

determinar, com precisão, qual seria o item correspondente da LM e, por conseguinte, se atuaria, aí, uma possível estratégia de transferência linguística<sup>171</sup>.

Assim sendo, e tendo em conta a análise empreendida, apresenta-se no Quadro 4.15 a distribuição dos itens nominais afetados da categoria 3 nos textos produzidos pelos informantes que são falantes nativos de espanhol e italiano nos vários níveis QECRL (A1-A2, B1-B2 e C1), com a indicação dos nomes correspondentes na LM e dos respetivos valores de género.

Posteriormente, no Quadro 4.16, apresenta-se a distribuição do número de ocorrências de desvios de AGN e de CNG considerados como potencialmente atribuíveis à transferência linguística, i.e., os casos em que não se observa a correspondência de valores de género (masculino e feminino) entre os itens lexicais afetados na LA e os correspondentes da LM dos informantes<sup>172</sup>. Assim, no Quadro 4.16 regista-se o número absoluto dos desvios da categoria 3 distribuídos por segmentos da amostra selecionada e que, em função dos itens afetados, são considerados como potencialmente atribuíveis à transferência de valores de género da LM. Para além disso, no Quadro 4.16 regista-se, em valores percentuais, o peso relativo dos comportamentos desviantes potencialmente atribuíveis à transferência linguística, apurado em função do número absoluto de desvios produzido por segmento da amostra de informantes de LM espanhola e italiana.

---

<sup>171</sup> A averiguação das formas nominais correspondentes no espanhol e italiano resultou de pesquisas feitas em dicionários bilingues (espanhol/português, italiano/português) e monolingues. Assim, para o espanhol, fez-se a consulta do *Diccionario de espanhol-português e português-espanhol* (2016) e do *Diccionario de la lengua española* (Real Academia Española 2001); para o italiano, consultou-se o *Diccionario de italiano-português, português-italiano* (2016).

<sup>172</sup> Convém ter em linha de conta que não foram contabilizados como “desvios potencialmente atribuíveis à transferência da LM” os casos em que se identificou, para um nome, duas ou mais formas lexicais correspondentes na LM do falante aprendente com diferentes valores de género, coincidentes ou não com os valores da forma-alvo, já que nestes não é possível inferir um eventual efeito negativo da transferência linguística.

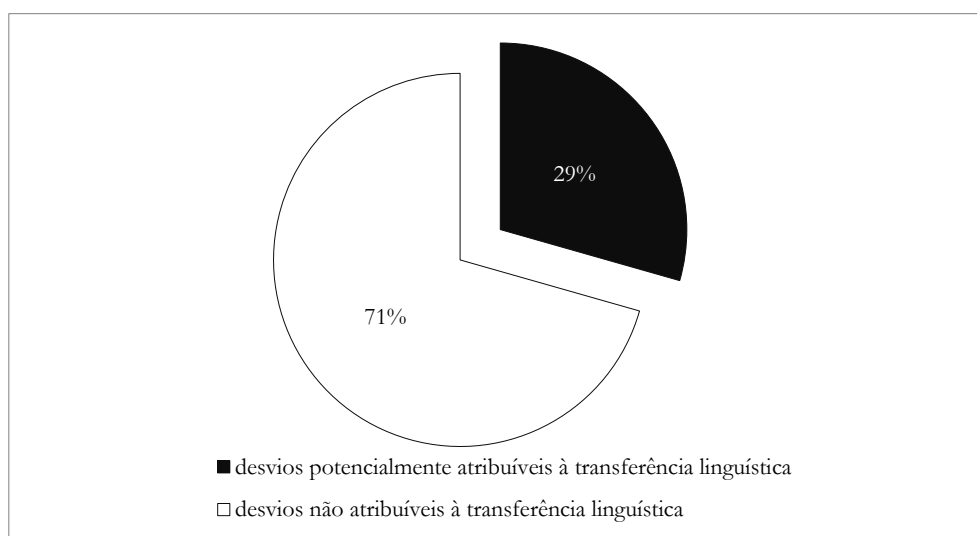
Informantes		Valor de género da palavra em português → Valor de género da palavra na LM do informante						
LM	Nível QECRL	Masc. → Masc.	Masc. → Fem.	Fem. → Fem.		Fem. → Masc.	Masc. ou Fem. → Masc. / Fem.	
Espanhol	A1-A2	•arranjo: <i>arreglo</i> •momento: <i>momento</i>	• copo: <i>copa</i>	• bancada: <i>encimera</i> • cadeira: <i>silla</i> • casa: <i>casa</i> • cena: <i>escena</i> • estatura: <i>estatura</i>	• hora: <i>hora</i> • morada: <i>dirección</i> • ortografia: <i>ortografía</i> • pessoa: <i>persona</i> • semana: <i>semana</i> • vida: <i>vida</i>	• casa de banho: <i>baño</i> • equipa: <i>equipo</i> • sanita: <i>váter, inodoro, retrete</i> • sexta-feira: <i>viernes</i>	• aldeia (Fem.): <i>aldea</i> (Fem.) / <i>pueblo</i> (Masc.) • assoalhada (Fem.): <i>cuarto</i> (Masc.) / <i>sala, habitación</i> (Fem.) • discoteca (Fem.): <i>club</i> (Masc.) / <i>discoteca</i> (Fem.)	
	B1-B2	•ano: <i>año</i> •jogo: <i>juego</i>		• campanha: <i>campaña</i> • coisa: <i>cosa</i> • hora: <i>hora</i> • ideia: <i>idea</i> • roupa: <i>ropa</i>				
	C1			• carnificina: <i>carniceria, matanza</i> • vida: <i>vida</i>			• casaco (Masc.): <i>abrigo</i> (Masc.), <i>chaqueta</i> (Fem.)	
Italiano	A1-A2	•ano: <i>anno</i> •casino: <i>casinò</i> •curso: <i>corso</i> •fundamento: <i>fondamento</i> •meio (de transporte): <i>mezzo di trasporto</i> •olho: <i>occhio</i> •pimento: <i>peperone</i> •prato: <i>piatto</i> •prédio: <i>edificio</i>	• rádio: <i>radio</i>	• atmosfera: <i>atmosfera</i> • cadeira: <i>sedia</i> • casa: <i>casa</i> • coisa: <i>cosa</i> • desculpa: <i>scusa</i> • energia: <i>energia</i> • garganta: <i>gola</i> • janela: <i>finestra</i>	• língua: <i>lingua</i> • mala: <i>valigia</i> • maneira: <i>maneira</i> • manta: <i>coperta</i> • montanha: <i>montagna</i> • pergunta: <i>domanda / interrogazione</i> • pessoa: <i>persona</i> • receita: <i>ricetta</i> • terra: <i>terra</i>	• altura: <i>momento</i> • bacia: <i>bacino</i> • cama: <i>letto</i> • casa de banho: <i>bagno</i> • comida: <i>cibo, vitto, alimento</i>	• cómoda: <i>cassettone</i> • entrada: <i>antipasto</i> • fogueira: <i>falò</i> • mesa de cabeceira: <i>comodino</i>	• mesa (Fem.): <i>tavolo</i> (Masc.), <i>tàvola</i> (Fem.) • peça (Masc.): <i>pezza</i> (Fem.), <i>pezzo</i> (Masc.)
	B1-B2	•campo: <i>campo</i> •prejuízo: <i>pregiudizio</i>		• casa: <i>casa</i> • coisa: <i>cosa</i> • ditadura: <i>dittatura</i>	• história: <i>storia</i> • limpeza: <i>pulizia</i> • pessoa: <i>persona</i> • semana: <i>settimana</i>	• ajuda: <i>aiuto</i> • brincadeira: <i>scherzo</i>	• comida: <i>cibo, vitto, alimento</i> • loja: <i>negozio</i>	
	C1	•meio (de comunicação): <i>mezzo di comunicazione</i>		• bilheteira: <i>biglietteria</i> • forma: <i>forma</i> • história: <i>storia</i> • pessoa: <i>persona</i>		• loja: <i>negozio</i>		

**Quadro 4.15** – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 3, detetados nas produções escritas por informantes de LM espanhola e italiana, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical

Informantes		Desvios da Categoria 3		
LM	Nível QECRL	Total	# de desvios potencialmente atribuíveis à transferência linguística	%
Espanhol	A1-A2	28	9	32,1
	B1-B2	7	0	0
	C1	4	0	0
Italiano	A1-A2	56	17	30,4
	B1-B2	18	7	38,9
	C1	6	1	16,7
$\Sigma$		119	34	28,5

**Quadro 4.16** – Distribuição do número de desvios da categoria 3 produzidos nos segmentos da amostra de informantes de LM espanhola e italiana potencialmente atribuíveis à transferência linguística

Tendo em conta os dados apresentados no Quadro 4.16, no Gráfico 4.21 apresenta-se, em valores percentuais, a proporção de desvios de AGN e de CNG, potencialmente atribuíveis a um efeito negativo de transferência linguística.



**Gráfico 4.21** – Percentagem de desvios da categoria 3 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM espanhola e italiana

Segundo os dados cartografados no Gráfico 4.21, do conjunto total de desvios de AGN e de CNG da categoria 3 detetados nas produções escritas por informantes que são falantes nativos de espanhol e de italiano, cerca de 29% dos desvios são potencialmente atribuíveis à transferência linguística.

Assim sendo, verifica-se que a maioria dos desvios da categoria 3 assinalados nestes segmentos da amostra não advém da transferência linguística. Com efeito, como se viu no Quadro 4.15, numa parte substancial dos nomes afetados, a forma da LM possui o mesmo valor de género (masculino ou feminino) da forma correspondente em português. Por



consequente, os desvios assinalados nestes segmentos da amostra dever-se-ão a outros fatores, nomeadamente a ambiguidade do *input* da LA no que concerne à atribuição dos valores de género.

A fim de obter uma visão mais fina destes dados, procede-se, em seguida, à sua análise por segmento da amostra.

De acordo com os dados apresentados nos Quadros 4.15 e 4.16, verifica-se que, nas produções dos aprendentes de LM espanhola, apenas se identificam desvios potencialmente atribuíveis à transferência de valores de género da LM nos níveis A1-A2, sendo que nos restantes níveis QECRL, não há registo de desvios que possam ter resultado da transferência linguística de valores de género.

Para além disso, é também muito frequente nos níveis A1-A2 neste segmento da amostra a utilização, no texto, de formas lexicais diretamente importadas da LM e também do valor de género gramatical a elas associadas. Já nos níveis subsequentes tal importação direta das formas lexicais dos nomes não é tão expressiva. São exemplos típicos desta importação direta, quer da forma lexical quer dos valores de género, os seguintes exemplos:

(61) «\*(...) estávamos tomando **umas copas**» (Espanhol.A1-A2)

(62) «\*(...) **o nosso equipo** ganhou a taça» (Espanhol.A1-A2)

(63) «\*(...) eu desejo procurar **um equipo feminino**» (Espanhol.A1-A2)

Em (61), o aprendente utiliza a forma *copas*, item diretamente importado da sua LM, em vez de *copos*, e em (62) e (63) recorre a *equipo* em vez de *equipa*, a forma-alvo.

Quanto ao segmento da amostra de informantes, que são falantes nativos de italiano, os desvios potencialmente atribuíveis à transferência de valores de género da LM registam-se em todos os níveis QECRL. Para além disso, ao contrário do que se verificara nos textos dos falantes nativos de espanhol, neste segmento não é recorrente a importação direta dos nomes da LM. Por sua vez, o que se regista mais frequentemente é a importação dos valores de género dos nomes da LM para os nomes da LA que, ao não serem coincidentes, conduzem ao desvio de AGN e de CNG (cf. Quadro 4.15). Por exemplo, ao descrever uma refeição, com as suas diferentes partes, o falante nativo de italiano utiliza o nome “*entrada*” associando-o ao valor de género masculino, podendo este comportamento desviante ser justificado pelo facto de, em italiano, o equivalente a *entrada* de uma refeição ser o nome masculino *antipasto*:

- (64) «\*A noite fizemos um jantar enorme: mexilhões com sumo de tomate, batatas fritas, massa com nata e fiambre e **muitos entradas**.»  
(Italiano.A1-A2)

Consideremos, agora, os desvios da categoria 3 detetados nas produções escritas por aprendentes de LM alemã. À semelhança do que foi feito anteriormente, apresenta-se, em primeiro lugar, no Quadro 4.17, a distribuição dos nomes afetados por valor de género na LA e das formas correspondentes em alemão com a indicação do valor de género a elas associadas<sup>173</sup>.

É sabido que o alemão apresenta um sistema tripartido de valores de género, não havendo, portanto, coincidência total do número de valores de género deste idioma para o português. Procura-se, então, averiguar se é possível estabelecer uma correspondência, direta ou indireta, entre os valores de género dos itens da LM e os valores de género dos itens da LA. Para além disso, pretende-se averiguar se existe um padrão recorrente de associação de um determinado valor de género em português (masculino ou feminino) a formas cujo correspondente em alemão é de valor neutro. Ou seja, procura-se apurar nos nomes afetados com desvio da categoria 3 cuja forma correspondente em alemão é neutra, se há uma marcação preferencial, por parte dos informantes, do valor de género masculino ou feminino.

Posteriormente, no Quadro 4.18 regista-se a distribuição do número absoluto de ocorrências desviantes consideradas como potencialmente atribuíveis à transferência direta de valores de género, i.e., os casos em que não se observa uma correspondência direta de valores de género (masculino ou feminino) entre os itens lexicais afetados na LA e os correspondentes em alemão. Assim sendo, no Quadro 4.18 é apresentado (i) o número absoluto de desvios da categoria 3 produzido neste segmento da amostra de informantes por nível QECRL; (ii) o número de desvios que, em função dos itens afetados, são potencialmente atribuíveis à transferência de valores de género (masculino ou feminino); e (iii) o peso relativo, em valores percentuais, dos desvios da categoria 3 potencialmente atribuíveis à transferência linguística, calculado a partir do número absoluto de desvios

---

<sup>173</sup> A averiguação das formas nominais correspondentes no alemão resultou de pesquisas feitas em dicionário bilingue (alemão/português), o *Dicionário de alemão-português, português-alemão* (2016).

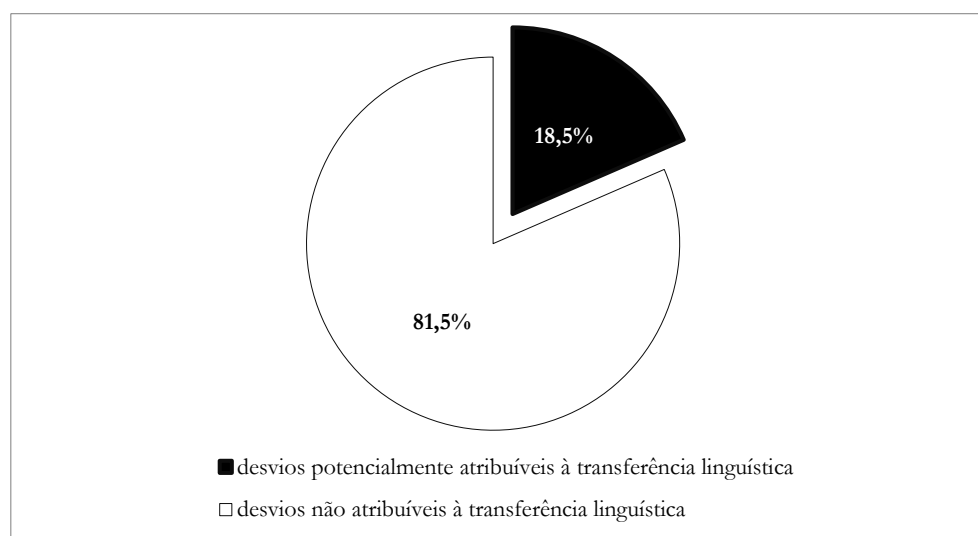
apurado por segmento da amostra de informantes de LM alemã nos diferentes níveis QECRL.

Valor de género da palavra em português → Valor de género da palavra na LM do informante							
Nível QECRL	Masc. → Masc.	Masc. → Fem.	Masc. → Neut.	Fem. → Fem.	Fem. → Masc.	Fem. → Neut.	Masc. ou Fem. → Masc. / Fem. / Neut.
A1-A2	<ul style="list-style-type: none"> <li>armário: <i>Schrank</i></li> <li>desporto: <i>Sport</i></li> <li>feriado: <i>Feiertag, Urlaub</i></li> <li>grupo: <i>Gruppe</i></li> <li>horário: <i>Stundenplan</i></li> <li>passeio: <i>Spaziergang, Ausflug</i></li> <li>plano: <i>Plan</i></li> </ul>	<ul style="list-style-type: none"> <li>canto: <i>Ecke, Kante</i></li> <li>método: <i>Methode</i></li> <li>mundo: <i>Welt</i></li> <li>tempo: <i>Zeit</i></li> <li>trabalho: <i>Arbeit</i></li> </ul>	<ul style="list-style-type: none"> <li>ano: <i>Jahr</i></li> <li>cabelo: <i>Haar</i></li> <li>casino: <i>Kasino</i></li> <li>dinheiro: <i>Geld</i></li> <li>estudo: <i>Studium</i></li> <li>medicamento: <i>Medikament</i></li> <li>remédio: <i>Heilmittel</i></li> </ul>	<ul style="list-style-type: none"> <li>boleia: <i>Ortscheit</i></li> <li>caipirinha: <i>Caipirinha</i></li> <li>discoteca: <i>Diskothek</i></li> <li>escola: <i>Schule</i></li> <li>experiência: <i>Erfahrung</i></li> <li>família: <i>Familie</i></li> <li>história: <i>Geschichte</i></li> <li>hora: <i>Stunde</i></li> <li>palmeira: <i>Palme</i></li> <li>peessoa: <i>Person</i></li> <li>ressaca: <i>Brandung</i></li> <li>técnica: <i>Technik</i></li> </ul>	<ul style="list-style-type: none"> <li>bolsa: <i>Beutel</i></li> <li>brincadeira: <i>Scherz, Spaß</i></li> <li>vista: <i>Anblick</i></li> <li>vitória: <i>Sieg</i></li> <li>volta: <i>Spaziergang</i></li> </ul>	<ul style="list-style-type: none"> <li>cerveja: <i>Bier</i></li> <li>coisa: <i>Ding</i></li> <li>criança: <i>Kind</i></li> <li>data: <i>Datum</i></li> <li>equipa: <i>Mannschaft</i></li> <li>folha: <i>Blatte</i></li> <li>película: <i>Häutchen</i></li> <li>piscina: <i>Schwimmbad</i></li> <li>vida: <i>Leben</i></li> </ul>	<ul style="list-style-type: none"> <li>testemunha (Fem.): <i>Zeuge (Masc.) / Zeugin (Fem.)</i></li> </ul>
B1-B2	<ul style="list-style-type: none"> <li>estado: <i>Stand</i></li> <li>estrangeiro: <i>Ausländer</i></li> <li>passatempo: <i>Zeitvertreib</i></li> <li>ponto: <i>Punkt</i></li> <li>ponto de vista: <i>Standpunkt</i></li> </ul>	<ul style="list-style-type: none"> <li>conjunto: <i>Gesamtheit</i></li> <li>lado: <i>Seite</i></li> <li>movimento: <i>Bewegung</i></li> <li>tempo: <i>Zeit</i></li> <li>trabalho: <i>Arbeit</i></li> </ul>	<ul style="list-style-type: none"> <li>ano: <i>Jahr</i></li> <li>silêncio: <i>Schweigen</i></li> </ul>	<ul style="list-style-type: none"> <li>cópia: <i>Kopie</i></li> <li>cultura: <i>Kultur</i></li> <li>diferença: <i>Verschiedenheit</i></li> <li>energia: <i>Energie</i></li> <li>experiência: <i>Erfahrung</i></li> <li>força: <i>Kraft</i></li> <li>indústria: <i>Industrie</i></li> <li>peessoa: <i>Person</i></li> <li>semana: <i>Woche</i></li> <li>violência: <i>Gewalt</i></li> </ul>	<ul style="list-style-type: none"> <li>floresta: <i>Wald</i></li> <li>praça: <i>Platz</i></li> <li>praia: <i>Strand</i></li> <li>raposa: <i>Fuchs</i></li> </ul>	<ul style="list-style-type: none"> <li>casa: <i>Haus</i></li> <li>criança: <i>Kind</i></li> <li>existência: <i>Dasein</i></li> <li>feira: <i>Fest</i></li> <li>loja: <i>Geschäft</i></li> </ul>	<ul style="list-style-type: none"> <li>alma (Fem.): <i>Seele (Fem.), Geist (Masc.)</i></li> <li>apoio (Masc.): <i>Stütze, Hilfe (Fem.), Halt (Masc.)</i></li> </ul>
C1				<ul style="list-style-type: none"> <li>peessoa: <i>Person</i></li> </ul>	<ul style="list-style-type: none"> <li>circunstância: <i>Umstand</i></li> <li>praia: <i>Strand</i></li> </ul>	<ul style="list-style-type: none"> <li>aventura: <i>Abenteuer</i></li> <li>coisa: <i>Ding</i></li> <li>palavra: <i>Wort</i></li> </ul>	

Quadro 4.17 – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 3, detetados nas produções escritas por informantes de LM alemã, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical

Nível QECRL	Desvios da Categoria 3		
	Total	# de desvios potencialmente atribuíveis à transferência linguística de valores de género (masculino ou feminino)	%
A1-A2	63	10	15,9
B1-B2	47	10	21,3
C1	9	2	22,2
$\Sigma$	119	22	18,5

**Quadro 4.18** – Distribuição do número de desvios da categoria 3 produzidos nos segmentos da amostra de informantes de LM alemã potencialmente atribuíveis à transferência linguística



**Gráfico 4.22** – Percentagem de desvios da categoria 3 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM alemã

Atendendo aos dados apresentados no Quadro 4.18 e também no Gráfico 4.22, é possível verificar que, em valores percentuais, o peso relativo de desvios da categoria 3 potencialmente atribuíveis à transferência linguística de valores de género masculino ou feminino, detetados nas produções escritas por informantes que são falantes nativos de alemão, a frequentar turmas de diferentes níveis QECRL (A1-A2, B1-B2 e C1), aproxima-se dos 20 pontos percentuais, não havendo, a este respeito, significativas diferenças entre os vários níveis de aprendizagem.

Importa ainda assinalar que numa grande parte dos nomes afetados, as formas lexicais que, em português, são de género feminino, tais como *coisa*, *criança*, *existência* e *vida*, têm como itens correspondentes, formas lexicais neutras, *Ding*, *Kind*, *Dasein* e *Leben*, respetivamente (cf. Quadro 4.17). Portanto, verifica-se que aquando da produção de nomes femininos cujos correspondentes em alemão são neutros, os aprendentes optam,

preferencialmente, por lhes associar o valor de género masculino, dando, assim, origem a desvios.

A estratégia de atribuição de valor de género masculino por *default* aos itens em português é, aliás, uma tendência registada igualmente em outros segmentos da amostra de informantes por LM, conforme já referido em 4.2.2.3 (veja-se, ainda, o Gráfico 4.5). Para além disso, e como se viu no Capítulo 1, Secção 1.4.3.2., em algumas categorias semânticas dos itens lexicais do alemão, o valor neutro encontra-se associado a termos lexicais que denotam uma certa ‘vagueza’ de significado, i.e., referem-se a noções genéricas da realidade. Por conseguinte, poder-se-á considerar que, em nomes como *coisa* e *existência*, por exemplo, os informantes poderão associar ao significado abrangente e vago do item lexical, o valor de género que, em português, permite obter uma leitura genérica, não-marcada, ou seja, o masculino.



Em suma, através da análise dos desvios da categoria 3, relativos a nomes cujo valor de género masculino e feminino é inferível a partir dos índices temáticos *-o* e *-a*, respetivamente, verifica-se que, ao longo do desenvolvimento da aprendizagem do português, os aprendentes contactam com uma maior diversidade de itens lexicais e, em consequência desse contacto, os dados aqui analisados sugerem que é sobre os novos conjuntos de nomes que incidem, preferencialmente, os desvios de AGN e de CNG.

A partir dos resultados apurados nesta categoria, constata-se ainda que, entre os diferentes segmentos da amostra por LM e nível QECRL, no momento de associação de valores de género aos nomes, os indícios semânticos predominam sobre os indícios formais. Mais ainda, e no caso particular dos dados extraídos dos segmentos da amostra de informantes, que são falantes nativos de espanhol, italiano e alemão, não se registou um peso significativo de eventuais efeitos da transferência linguística de valores de género das respetivas LM. Por conseguinte, os índices de desvio assinalados nesta categoria nestes segmentos da amostra poder-se-ão dever à ambiguidade do *input* da LA que se traduz numa certa atitude de ‘desconfiança’ por parte dos aprendentes, face ao poder preditivo dos indícios morfológicos no momento da produção dos nomes.

Para além disso, o facto de, nos vários segmentos da amostra, o número de ocorrências no *corpus* dos nomes afetados ser, tendencialmente, mais elevado do que o número de desvios de atribuição de género, leva-nos a concluir que, em grande medida, os aprendentes são capazes de reconhecer o valor de género do nome. Todavia, e como se

verificou, em determinados contextos de uso dos nomes, os informantes não selecionam a forma adequada quanto ao género do especificador e/ou do modificador que integra o SN. Ou seja, os informantes apresentam, em geral, dificuldades na instrumentalização do mecanismo de concordância nominal em género (CNG).

#### 4.2.3.1.4. Categoria 4 [SEM -, FORM -] (*o pente, a ponte, o coração, a viagem*)

A categoria 4 integra desvios de atribuição e de concordância nominal em género relativos a nomes cujo valor masculino ou feminino não se infere a partir de critérios de natureza semântica e/ou formal. Os exemplos (65) a (68) ilustram os desvios da categoria 4 assinalados no *corpus* em estudo:

- (65) «\*Das ~~(8:00)~~ 8:30 as 9:00 de manhã, tomava-me **uma douche** e depois fazia a cama» (Espanhol.A1-A2)
- (66) «\*O filme começa com **uns imagens** de mortos em África (...)» (Alemão.C1)
- (67) «\***Na fim de semana passada**» (Inglês.A1-A2)
- (68) «\*No <Res> res-do-chão há **uma restaurante**.» (Chinês.A1-A2)

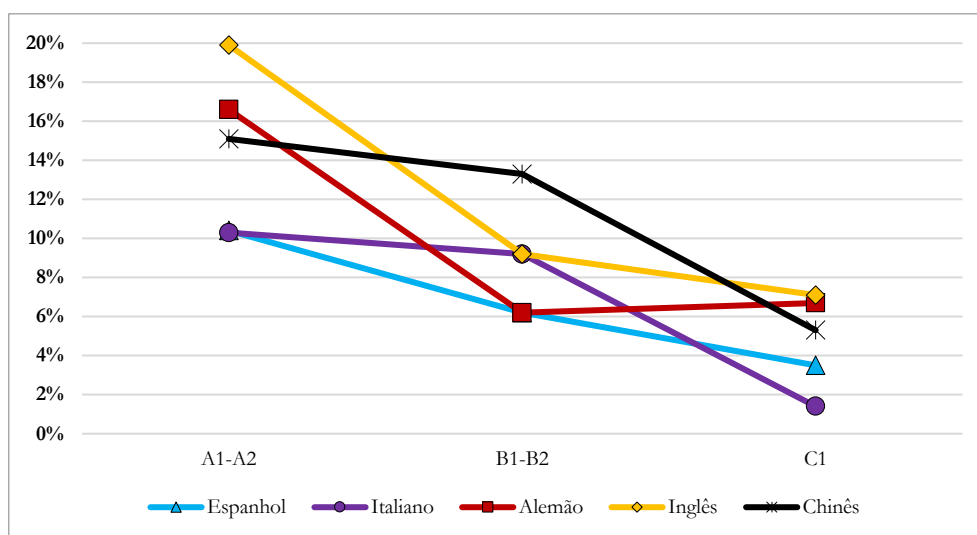
Tal como foi feito para as demais categorias de desvio, fez-se a contagem do número de ocorrências de nomes da categoria 4 em SN não reduzidos nos textos selecionados (cf. Quadro 4.19), de modo a determinar a sua frequência de uso por produção escrita. Além do valor absoluto, calcularam-se também os valores da média ( $\bar{X}$ ) de ocorrências de nomes e de desvios por texto, apurados em cada segmento da amostra (por LM e nível QECRL), tendo-se, então, determinado, no conjunto total, um valor médio de 13 ocorrências de nomes da categoria 4 por produção textual.

Deste modo, e a par do que se viu para a categoria 3 (cf. Quadro 4.12), a categoria 4 apresenta uma frequência relativamente alta de nomes por texto. Quanto aos valores da média de desvios, apurou-se, em termos globais, uma média de 2 desvios por texto, sendo o valor mais elevado relativamente aos valores registados nas restantes categorias já analisadas (categorias 1, 2 e 3).

Informantes		Produções Escritas		Categoria 4				
				Nomes em SN não reduzidos		Desvios		
LM	Nível QERCL	#	# de ocorrências de nomes	# de ocorrências	$\bar{X}$	# de desvios	$\bar{X}$	%
Espanhol	A1-A2	46	1 280	415	9,02	43	1,48	10,4
	B1-B2	26	933	353	13,58	22	1,29	6,2
	C1	11	549	227	20,64	8	1	3,5
Italiano	A1-A2	104	3 141	1087	10,45	112	1,53	10,3
	B1-B2	40	1 385	573	14,325	53	2,04	9,2
	C1	12	1 699	633	52,75	9	1,2	1,4
Alemão	A1-A2	74	1 784	615	8,31	102	1,89	16,6
	B1-B2	60	2 407	900	15,00	56	1,44	6,2
	C1	20	928	386	19,30	26	1,73	6,7
Inglês	A1-A2	46	926	316	6,87	63	1,85	19,9
	B1-B2	54	2 783	1088	20,15	100	2,27	9,2
	C1	29	1 127	493	17,00	35	1,84	7,1
Chinês	A1-A2	37	957	370	10,00	56	1,87	15,1
	B1-B2	34	1 221	412	12,12	55	2,29	13,3
	C1	10	343	150	15,00	8	1,33	5,3
$\Sigma$		603	21 463	8 018	13,30	748	1,76	9,33

**Quadro 4.19** – Distribuição do número absoluto e dos valores média de nomes da categoria 4 produzidos no corpus, e dos desvios de AGN e de CNG por segmento da amostra (LM e nível QERCL)

Tendo em conta o número absoluto de ocorrências de nomes em SN não reduzidos da categoria 4 produzidos no *corpus*, calculou-se as percentagens relativas de desvio por segmentos da amostra (cf. Gráfico 4.23).



**Gráfico 4.23** – Percentagem relativa de desvios da categoria 4



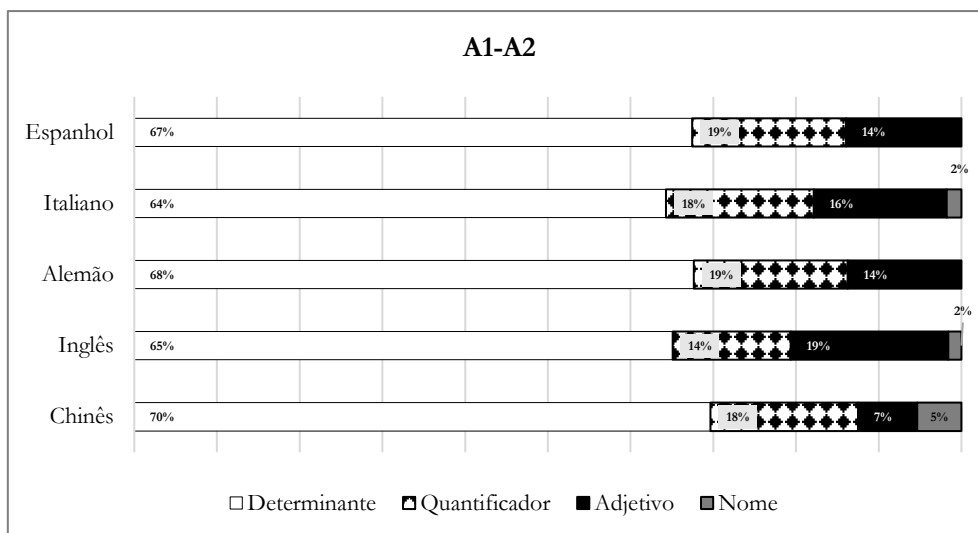
Segundo os dados representados no Gráfico 4.23, no segmento da amostra de falantes nativos das línguas românicas dos níveis A1-A2 registam-se as percentagens relativas de desvios mais baixas relativamente às verificadas nos restantes segmentos de informantes destes níveis. De facto, nestes dois segmentos da amostra, os desvios de AGN e de CNG afetam cerca de 10% do conjunto total de ocorrências itens produzidos pertencentes à categoria 4 (veja-se, ainda, o Quadro 4.19).

Em contrapartida, nos textos escritos por falantes nativos de idiomas sem a categoria de género gramatical, registam-se percentagens relativas de desvios elevadas, sendo o segmento da amostra de informantes de LM inglesa o que apresenta um pior desempenho nos níveis A1-A2, visto que, aproximadamente 20% das ocorrências de nomes da categoria 4 registam, neste segmento da amostra, desvios de AGN e de CNG (cf. Quadro 4.19 e Gráfico 4.23).

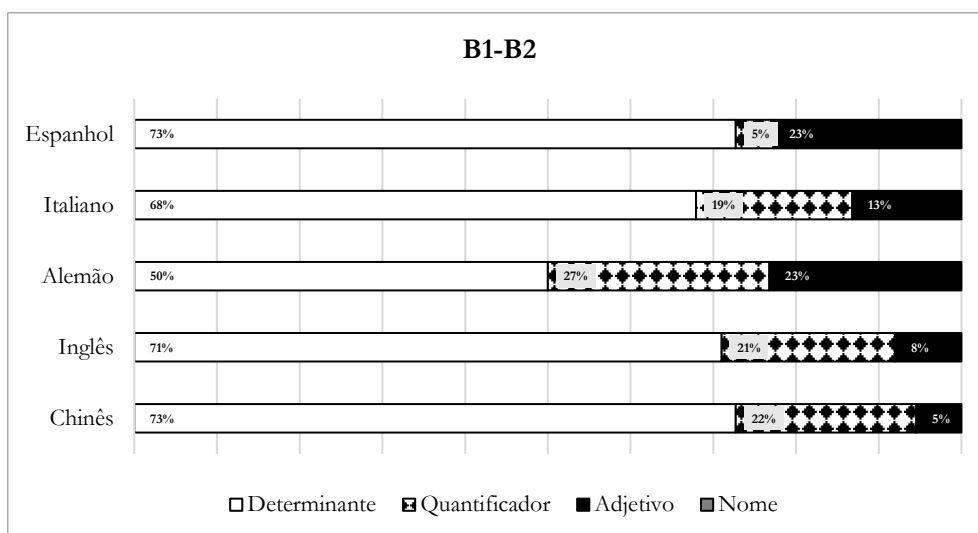
Entre os falantes de LM alemã, os valores apurados nos níveis A1-A2 revelam igualmente, neste estágio da aprendizagem, algumas dificuldades por parte destes aprendentes na associação dos valores de género aos nomes da categoria 4 e no consequente estabelecimento das relações de concordância nominal em género.

Já na passagem para os níveis B1-B2, é notória a progressiva descida das percentagens relativas de desvios nos diferentes segmentos da amostra por LM, sobretudo nas produções dos falantes nativos de alemão e de inglês. Portanto, é observável, nestes níveis, um desenvolvimento da assimilação dos valores de género nominal nos vários segmentos da amostra por LM, mantendo-se essa tendência, embora não tão acentuada, na passagem para o nível C1. Por conseguinte, poder-se-á assumir que, ao longo das diferentes fases do desenvolvimento interlinguístico dos aprendentes, há uma progressão gradual relativamente à assimilação dos valores de género dos nomes nos quais a atribuição desses valores não depende nem de critérios formais nem de critérios semânticos.

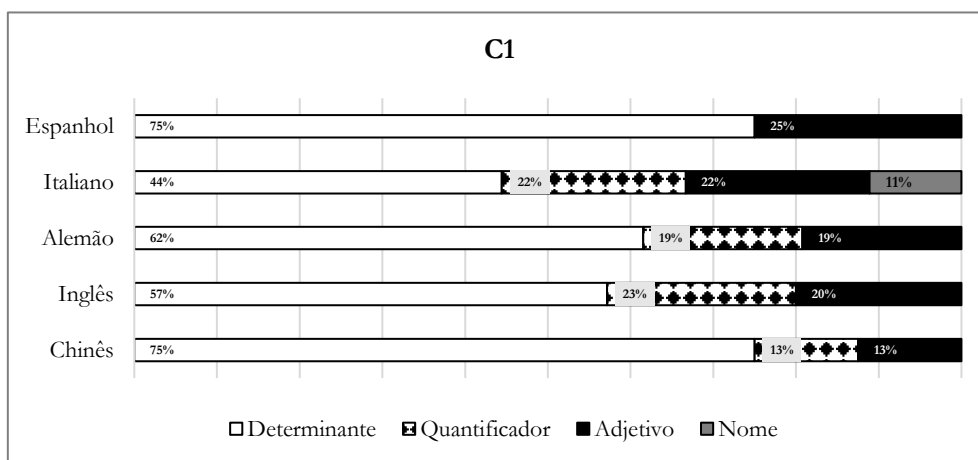
Os Gráficos 4.24, 4.25 e 4.26 registam, em valores percentuais, a proporção dos desvios da categoria 4 por constituinte afetado (determinante, quantificador, adjetivo e nome).



**Gráfico 4.24** – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, nos níveis A1-A2



**Gráfico 4.25** – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, nos níveis B1-B2



**Gráfico 4.26** – Percentagem de desvios da categoria 4, distinguidos por classes de palavras afetadas, no nível C1

Como é possível constatar nos Gráficos 4.24 a 4.26, em todos os níveis QECRL a maior proporção de desvios da categoria 4 incide nos especificadores, nomeadamente nos determinantes, sendo muito pouco frequentes os desvios que se registam na forma morfológica do nome. Não há, por conseguinte, diferenças significativas relativamente ao constituinte mais afetado nos segmentos da amostra de LM dos informantes e também nos diferentes níveis QECRL.

Veja-se, em seguida, alguns exemplos de desvios extraídos do *corpus* em análise:

- (69) «\*Durante **o noite** há muitos fogos artificio e também grandes fogueiras.»  
(Inglês.B1-B2)
- (70) «\*Tenho **muito saudade** para ti» (Alemão.A1-A2)
- (71) «\*A casa tem 4 quartos, dois casas de banho, uma sala de jantar, uma cozinha grande e **um garagem**.» (Chinês.A1-A2)
- (72) «\*As mães sempre querem que ajudemos **às labores** da (~~coisa~~) casa, mas nunca estão satisfeitas, se nós fizermos muitas coisas exemplo: lavar a loiça, arrumar a (~~casa~~) cozinha etc. (...)» (Espanhol.C1)

São ainda frequentes, nesta categoria, os casos de ambivalência da associação dos valores de género, i.e., observa-se, nas produções dos aprendentes, a opção por uma estratégia de ‘seleção mista’ dos valores de género nos diferentes elementos que compõem o SN:

- (73) «\*Na escola si temos uma **colificação** [qualificação] **bom** os nossos pais não estão satisfeitos porque pudesse ser melhor.» (Espanhol.A1-A2)
- (74) «\*Nos comprámos e cozinhámos peixe no **última dia** do ano e depois o jantar fomos para praça de commercio junto.» (Alemão.A1-A2)
- (75) «\*A Suíça é um **país montanhosa** pero não em todas partes.»  
(Inglês.B1-B2)
- (76) «\*Aproveitamos assim, e no mesmo tempo somos vítimas, do **sistema económica** global.» (Inglês.C1)

Portanto, estes dados são reveladores das dificuldades que a maioria dos aprendentes, independentemente da sua LM, tem no domínio dos mecanismos de concordância sintática nominal em género.

Após a indicação, em termos quantitativos, do peso relativo de desvios apurado na categoria 4, procede-se, em seguida, à análise dos dados em função do item nominal afetado. Deste modo, e à semelhança do que foi feito nas categorias de desvio anteriormente analisadas, fez-se o levantamento dos nomes afetados, tendo em conta, quer o número absoluto das suas ocorrências nos textos produzidos por cada segmento da amostra, quer o número de ocorrências de desvios de AGN e de CNG. Pretende-se, assim, dar conta do léxico afetado por segmento, a fim de averiguar se há incidência de desvios sobre um conjunto de itens lexicais em particular, por um lado, e, por outro, se é possível aferir uma relação entre a proporção de desvios com a própria estrutura morfológica do nome, i.e., se há uma maior proporção de desvios sobre formas nominais simples ou complexas, sejam portadoras de afixos ou compostas.

No Quadro 4.20 apresenta-se o registo quantitativo dos desvios da categoria 4 distribuídos por LM e nível QECRL da turma frequentada pelo informante, com a indicação:

- (i) do número de nomes da categoria 4 com desvios (**# de nomes afetados**);
- (ii) do número de ocorrências desses nomes afetados no *corpus* (**# de ocorrências dos nomes afetados no corpus**);
- (iii) do número de desvios da categoria 4 (**# de desvios**);
- (iv) do valor da média de desvios da categoria 4 por nome afetado ( **$\bar{X}$  de desvios**);
- (v) da percentagem relativa dos desvios, calculada em função do respetivo número de ocorrências dos nomes afetados (**% de desvios**).

LM	Nível QECRL	# de nomes afetados	# de ocorrências dos nomes afetados no <i>corpus</i>	# de desvios	$\bar{X}$ de desvios	% de desvios
Espanhol	A1-A2	20	89	43	2,2	48
	B1-B2	15	70	22	1,5	31
	C1	7	13	8	1,1	62
Italiano	A1-A2	46	415	112	2,3	27
	B1-B2	27	208	53	1,9	25
	C1	9	16	9	1	56
Alemão	A1-A2	57	379	102	1,8	27
	B1-B2	39	363	56	1,5	16
	C1	20	51	26	1,3	51
Inglês	A1-A2	37	173	63	1,7	36
	B1-B2	55	525	100	1,8	19
	C1	30	156	35	1,2	22
Chinês	A1-A2	37	230	56	1,5	24
	B1-B2	28	164	55	2	34
	C1	8	16	8	1	50
$\Sigma$		435	2 868	748	1,7	26

**Quadro 4.20** – Distribuição do número de nomes afetados, do respetivo número de ocorrências no *corpus* e do número de desvios na categoria 4

De acordo com o exposto, na categoria 4, apurou-se um total de 748 desvios de AGN e de CNG que dizem respeito a 435 itens nominais, sendo que estes nomes totalizam 2868 ocorrências no *corpus* selecionado. O valor da média do número de desvios por nome é de, aproximadamente, 2 desvios por item.

Entre o conjunto de nomes afetados, constata-se que os desvios afetam 233 nomes distintos. Trata-se, portanto, da categoria não só com um maior número de desvios, como a que apresenta um maior número de nomes afetados. Todavia, e à semelhança das restantes categorias anteriormente analisadas, tendo em conta o número de ocorrências dos nomes afetados no *corpus*, constata-se que, no momento de produção desses nomes, os aprendentes dos vários segmentos da amostra selecionada acertam mais do que erram.

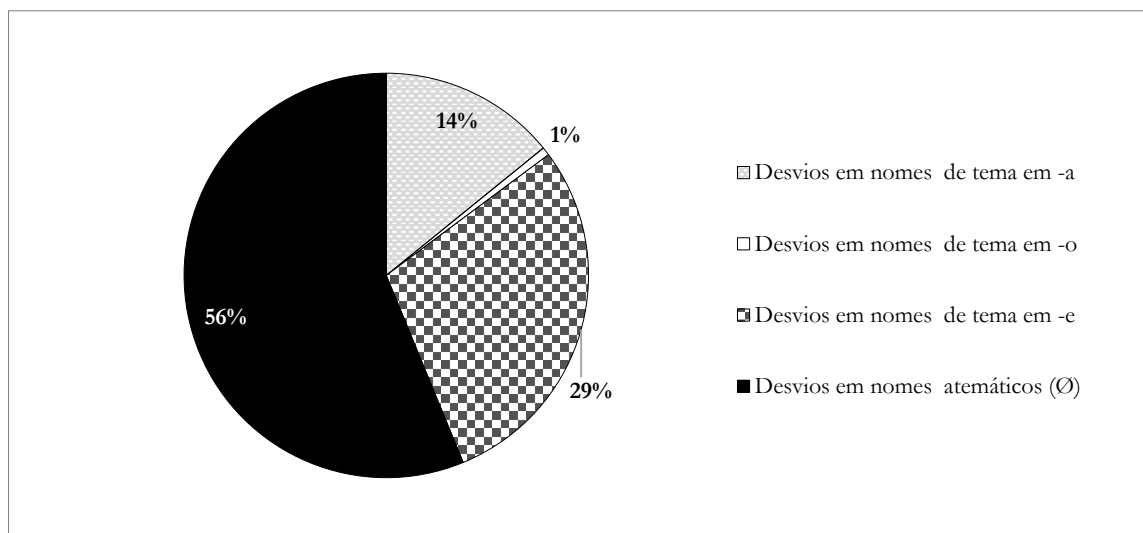
A categoria 4 integra itens nominais muito variados, quer no que diz respeito à sua estrutura formal, quer no que diz respeito ao seu conteúdo semântico. Com efeito, sob esta categoria registaram-se desvios de nomes com referentes sexuais e não sexuais, cujo valor de género não se infere nem a partir de critérios de natureza semântica nem de tipo formal.

Dada a variedade de itens pertencentes a esta categoria de desvio, para o tratamento dos dados empíricos, considerou-se relevante distinguir os comportamentos desviantes atendendo à estrutura formal do nome. Mais especificamente, distinguiram-se os desvios, por segmentos da amostra (LM e nível QECRL), tendo em conta o constituinte temático do nome afetado, com o objetivo de averiguar se há uma maior incidência de desvios da categoria 4 em: (i) nomes nos quais não se verifica a correlação mais frequente entre índice temático *-o* e género masculino e índice temático *-a* e género feminino, como *a tribo*, *o dia* e *o/a jornalista*; (ii) nomes de tema em *-e* (*o/a viajante*, *o pente*, *a ponte*); e (iii) nomes atemáticos  $\emptyset$ , como *o leão*, *a mansão*, *o pontapé*, *a viagem* (cf. Quadro 4.21).

Informantes		Categoria 4				
Língua Materna	Nível QECRL	# de desvios	-o	-a	-e	atemáticos (Ø)
Espanhol	A1-A2	43	0	4	9	30
	B1-B2	22	0	1	10	11
	C1	8	0	0	3	5
Italiano	A1-A2	112	0	11	21	80
	B1-B2	53	0	1	13	39
	C1	9	0	0	1	8
Alemão	A1-A2	102	0	23	31	48
	B1-B2	56	0	11	14	31
	C1	26	0	4	5	17
Inglês	A1-A2	63	1	3	17	42
	B1-B2	100	0	14	44	42
	C1	35	0	7	8	20
Chinês	A1-A2	56	1	8	26	21
	B1-B2	55	1	18	13	23
	C1	8	2	0	2	4
$\Sigma$		748	5	105	217	421

**Quadro 4.21** – Distribuição dos desvios da categoria 4 produzidos por segmento da amostra (LM e nível QECRL) em função do constituinte temático do item lexical afetado (-o, -a, -e, atemáticos (Ø))

A partir do número absoluto de desvios da categoria 4, calculou-se o peso relativo de desvios atendendo ao constituinte temático dos nomes afetados. A distribuição dos valores apurados encontra-se representada no Gráfico 4.27.



**Gráfico 4.27** – Percentagem de desvios da categoria 4, calculada em função do constituinte temático do item lexical afetado (-o, -a, -e, e atemáticos (Ø))

Segundo os dados apresentados no Quadro 4.21 e no Gráfico 4.27, há uma incidência preferencial de desvios nos nomes atemáticos, seguindo-se os nomes de tema em -e.

Por sua vez, os itens femininos de tema em *-o* são os menos afetados. Ora, o número reduzido de desvios nestes nomes, registado em todos os segmentos da amostra, poder-se-á justificar pelo facto de, no léxico do português, serem residuais e de baixa frequência os nomes femininos de tema em *-o*, como, por exemplo, *a tribo* e *a libido* (veja-se, a este respeito, o Capítulo 1, Quadro 1.3).

Com efeito, no *corpus* selecionado para este estudo, registaram-se apenas 5 desvios de AGN e de CNG em nomes femininos de tema em *-o*. Estes desvios incidem sobre 3 nomes distintos, dois dos quais correspondem a formas truncadas: *foto* (fotografia), *moto* (motorizada) e *tribo*. Em nenhum dos segmentos da amostra de informantes que são falantes nativos de idiomas com categoria de género gramatical – espanhol, italiano e alemão – se registaram desvios nestes itens. Somente se registaram desvios em nomes femininos de tema em *-o*, nas produções escritas por informantes de LM inglesa e chinesa, sendo que, nestes últimos, foram assinalados desvios em todos segmentos da amostra por nível QECRL, tal como se ilustra nos exemplos (81) a (83):

- (81) «\*No tempo livre, eu sempre dei uma volta das cidade de Lisboa e tirei **muitos fotos**.» (Chinês.A1-A2)
- (82) «\*E também vi uma pintura famosa, que é “auto-retrato” por “düerer” e eu fiz **um foto** com esta pintura.» (Chinês.B1-B2)
- (83) «\*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 (...)» (Chinês.C1)

Portanto, a forma do item lexical parece ser relevante para este segmento da amostra, já que estes dados indiciam que, no momento da atribuição do valor de género aos nomes, os informantes de LM chinesa recorrem à pseudorregra de associação de valor de género nominal ‘se nome termina em *-o* é masculino, se termina em *-a* é feminino’.

Entre os nomes da categoria 4 de tema em *-a*, registou-se um total de 105 desvios, sendo este um valor muito superior ao registado em nomes femininos de tema em *-o* (cf. Quadro 4.21). Apresenta-se no Quadro 4.22 a distribuição, por segmentos da amostra (LM e nível QECRL), dos nomes afetados de tema em *-a*, com a indicação, por item, do número de ocorrências no *corpus* e do número de desvios de AGN e de CNG a que deram origem.

LM	Nível QECRL		
	A1/A2	B1/B2	C1
	Nomes (# de ocorrências / # de desvios)	Nomes (# de ocorrências / # de desvios)	Nomes (# de ocorrências / # de desvios)
Espanhol	cinema (4/3) colega (6/1)	problema (4/1)	
Italiano	dia (101/5) problema (14/3) mapa (3/1) sistema (1/1)	dia (25/1)	
Alemão	dia (67/11) turista (11/5) chapa (3/2) automobilista (1/1) cinema (3/1) colega (4/1) problema (7/1)	dia (58/4) tema (4/3) mapa (1/1) (os) <i>media</i> (2/1) problema (13/1) turista (2/1)	fantasma (2/3) problema (7/1)
Inglês	dia (13/1) tema (2/1) meio-dia (5/1)	problema (53/4) dia (35/3) turista (5/3) clima (1/1) jornalista (1/1) tema (3/1) terrorista (5/1)	dia (23/1) drama (1/1) (os) <i>media</i> (1/1) poema (1/1) problema (11/1)
Chinês	dia (43/2) programa (1/2) cinema (3/1) colega (7/1) guarda-chuva (4/1) problema (3/1)	guarda-chuva (7/8) dia (66/7) cinema (5/3)	

**Quadro 4.22** – Distribuição de nomes afetados da categoria 4 de tema em *-a*, por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no *corpus* e de desvios por item

O conjunto total de 105 desvios apurado afeta 21 nomes distintos, a saber: *automobilista*, *chapa*<sup>174</sup>, *cinema*, *clima*, *colega*, *dia*, *drama*, *fantasma*, *guarda-chuva*, *jornalista*, *mapa*, *meio-dia*, *media*, *poema*, *problema*, *programa*, *samba*, *sistema*, *tema*, *terrorista* e *turista*. Relativamente à estrutura morfológica dos itens, conforme se pode verificar, os nomes afetados correspondem, maioritariamente, a formas simples, havendo ainda 2 compostos sintáticos (*meio-dia* e *guarda-chuva*) e 3 nomes derivados em *-ist(a)* que são comuns de dois (*o/a automobilista*, *o/a jornalista* e *o/a terrorista*).

É, sobretudo, nos níveis A1-A2, na maior parte dos segmentos da amostra por LM, que se regista o maior número de desvios de AGN e de CNG por item lexical, sendo que na passagem para os níveis B1-B2 e destes para C1, o número de desvios por item diminui consideravelmente em alguns segmentos da amostra de LM. Com efeito, nos textos de nível C1 produzidos por falantes nativos de espanhol, italiano e chinês não há registo de ocorrências de desvios nos nomes de tema em *-a* pertencentes à categoria 4, ao passo que

<sup>174</sup> O nome *chapa* corresponde, no contexto da produção escrita, a um moçambicanismo. Segundo informação disponível no endereço da *Cátedra Português Língua Segunda e Língua Estrangeira*, “o *chapa*” corresponde a um transporte coletivo, semiformal; por extensão, qualquer automóvel que transporte pessoas a troco de algum dinheiro. Cf. <http://www.catedraportugues.uem.mz/> e <http://mocambicanismos.blogspot.com/2009/01/s.html>.



nos textos produzidos por falantes nativos de alemão e de inglês no nível C1 registam-se ainda desvios sobre este tipo de itens.

Assinale-se ainda que, em grande parte dos segmentos da amostra por LM nos vários níveis QECRL, os desvios incidem preferencialmente sobre os nomes *dia* e *problema*, facto justificável, talvez, pelo número elevado de ocorrências destes itens nos textos selecionados:

- (84) «\* Com 17 anos (1974) decidi que não queria mais continuar porque tinha **imensas problemas** com matemática e além disso estava cheia de vontade de ir viajar.» (Alemão.A1-A2)
- (85) «\*O filme, Zona J considera **varias problemas** sociais em Lisboa.» (Inglês.B1-B2)
- (86) «\***Nestas dias** eu fiz tantas coisas.» (Inglês.C1)
- (87) «\*Na minha cidade, Sasari, eu vou **todas as dia** em Faculdade.» (Italiano.A1-A2)

Note-se ainda que o número de ocorrências dos nomes *problema* e *dia*, no *corpus*, é elevado em relação ao número de desvios apurado (cf. Quadro 4.22). Ou seja, no momento de produção destes itens, os aprendentes acertam mais do que erram.

Consideremos, agora, no Quadro 4.23 a distribuição do número de ocorrências dos nomes afetados da categoria 4 de tema em *-e*, e do respetivo número de desvios de AGN e de CNG por segmentos da amostra (LM e nível QECRL).

LM	Nível QECRL					
	A1/A2		B1/B2			C1
	Nomes (# de ocorrências / # de desvios)		Nomes (# de ocorrências / # de desvios)			Nomes (# de ocorrências / # de desvios)
Espanhol	<b>árvore (3/3)</b> autocolante (2/1) cidade (11/1) costume (1/1)	duche (2/1) noite (11/1) sangue (1/1)	<b>costume (4/2)</b> <b>legume (3/2)</b> chatice (1/1)	leite (1/1) noite (13/1) parte (4/1)	<b>costume (2/2)</b> análise (1/1)	
Italiano	<b>duche (3/3)</b> <b>lebre (6/2)</b> <b>noite (43/2)</b> <b>restaurante (6/2)</b> carpete (1/1) espécie (1/1) estante (3/1)	exame (5/1) hipótese (1/1) novidade (1/1) oportunidade (1/1) parte (10/1) tarde (12/1) universidade (22/1)	<b>cidade (51/3)</b> <b>árvore (2/2)</b> atitude (1/1) atividade (7/1) ave (5/1)	faculdade (1/1) horizonte (1/1) possibilidade (3/1) restaurante (17/1)	possibilidade (1/1)	
Alemão	<b>parte (11/8)</b> <b>aguardente (2/2)</b> <b>clube (1/2)</b> <b>noite (28/2)</b> <b>morte (3/2)</b> <b>parque (6/2)</b> duche (2/1) exame (2/1) faculdade (6/1) frente (7/1)	gente (1/2) greve (2/1) oportunidade (1/1) personalidade (1/1) ponte (1/1) saudade (1/1) tarde (8/1)	<b>saudade (7/3)</b> <b>árvore (1/2)</b> amizade (1/1) arte (6/1) filme (5/1) noite (29/1)	possibilidade (16/1) sede (1/1) sorte (3/1) teste (1/1) universidade (13/1)	<b>corrente (1/2)</b> nome (1/1) saudade (3/1) variedade (1/1)	
Inglês	<b>cidade (13/4)</b> <b>parte (4/3)</b> <b>restaurante (3/3)</b> bigode (1/1) carne (1/1) classe (1/1)	duche (3/1) estante (1/1) nacionalidade (1/1) noite (16/1) pele (1/1) saudade (1/1)	<b>ganguê (8/10)</b> <b>noite (19/5)</b> <b>atitude (9/3)</b> <b>cidade (96/3)</b> <b>parte (19/3)</b> <b>torre (8/3)</b> <b>saudade (14/3)</b> restaurante (8/2)	análise (1/1) árvore (3/1) ataque (9/1) autoridade (2/1) chave (2/1) estudante (18/1) hipótese (1/1)	nacionalidade (3/1) ponte (2/1) realidade (9/1) semente (3/1) sorte (3/1) vale (1/1)	<b>sanduíche (1/2)</b> atividade (1/1) ponte (2/1) chocolate (4/1) detalhe (1/1) hipótese (1/1) parte (7/1) torre (1/1)
Chinês	<b>atividade (6/4)</b> <b>restaurante (11/4)</b> <b>parque (4/3)</b> ponte (4/2) árvore (4/1) cidade (24/3)	cliente (3/1) faculdade (2/1) filme (13/2) karaoke (1/1) neve (2/1) qualidade (1/1)	saudade (4/1) saúde (2/1) série (3/1) universidade (27/1) viagante (1/1)	<b>torre (7/3)</b> carne (1/1) chave (1/1) filme (8/1) fome (1/1)	greve (1/1) gripe (2/1) liberdade (1/1) neve (1/1) nordeste (1/1) restaurante (3/1)	torre (1/1) velhice (1/1)

Quadro 4.23 – Distribuição de nomes afetados da categoria 4 de tema em -e, por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no *corpus* e de desvios por item

Entre os nomes afetados de tema em *-e*, apurou-se um total de 217 desvios que incidem sobre 77 nomes distintos. A tendência assinalada nos segmentos da amostra, em geral, é a de haver uma maior diversidade de itens afetados nos níveis A1-A2 e B1-B2 relativamente ao nível C1, tendência, aliás, já registada noutras categorias de desvio previamente analisadas.

No que diz respeito à estrutura morfológica do nome, constata-se que a maior parte dos desvios afeta nomes simples e são pouco frequentes os desvios em formas nominais complexas, sejam compostas (*aguardente*), portadoras de afixos (*chatice*, *estudante*, *liberdade*, *variedade*, *velhice* e *viajante*) ou resultantes da recategorização do radical verbal (*detalhe*). Assim sendo, não é possível estabelecer uma relação entre a forma morfológica do nome e a incidência de desvios de AGN e de CNG nos vários segmentos da amostra por LM, já que, a este respeito, não há padrões substancialmente divergentes.

O número de desvios por item é tipicamente mais alto nos níveis de A1-A2, tendo em conta os valores da média de desvios por nome apurados nos diferentes segmentos da amostra de informantes por LM na categoria 4 (cf. Quadro 4.20).

Já no nível C1, para além de o número de nomes afetado ser mais reduzido do que o registado nos restantes níveis, o número de desvios por item tende a ser igual ao número de ocorrências desse item. Portanto, embora se registre uma diminuição do número de nomes afetados, aquando da produção de um item nominal de tema em *-e* em textos produzidos por informantes a frequentar turmas do nível C1, verifica-se que, de cada vez que se produz o nome, esta ocorrência dá origem ao desvio. Assim, este resultado sugere que o valor de género não é uma das primeiras propriedades do nome a serem adquiridas pelos aprendentes tardios.

Veja-se, em seguida, no Quadro 4.24, a distribuição dos nomes atemáticos ( $\emptyset$ ) afetados por segmentos da amostra, com a indicação do respetivo número de ocorrências e de desvios por item lexical afetado.

		Nível QECRL																																														
LM	A1/A2				B1/B2				C1																																							
	Nomes (# de ocorrências / # de desvios)				Nomes (# de ocorrências / # de desvios)				Nomes (# de ocorrências / # de desvios)																																							
Espanhol	<b>viagem (16/15)</b> cor (8/3) passagem (3/3) sinal (2/2) dor (1/1) férias (12/1)				garagem (1/1) lençol (2/1) mestiçagem (1/1) qualificação (1/1) <i>t-shirt</i> (1/1)				<b>dor (2/3)</b> <b>viagem (1/2)</b> competição (3/1) desordem (1/1)				lavor (1/1) origem (1/1) país (23/1) par (2/1)				dor (1/1) lavor (1/1) instalação (2/1)				rodagem (1/1) vantagem (5/1)																											
	Italiano	<b>viagem (33/26)</b> <b>passagem (11/9)</b> <b>fim (17/5)</b> <b>paisagem (3/5)</b> <b>reciclagem (9/4)</b> <b>cor (7/4)</b> ar (1/2) capital (1/2) liquidação (1/2)				origem (4/2) vez (42/2) aplicação (1/1) civilização (1/1) formação (1/1) habitação (2/1) coragem (1/1) divisão (2/1) faláfel (1/1) <i>internet</i> (4/1)				investigação (1/1) motivação (1/1) patinagem (1/1) pastel (3/1) residencial (2/1) sofá (5/1) televisão (12/1) versão (2/1) visão (1/1)				<b>dor (6/6)</b> <b>internet (11/6)</b> <b>vez (47/5)</b> <b>vantagem (5/4)</b> <b>viagem (4/4)</b> <b>região (4/3)</b>				atenção (1/1) canção (2/1) cor (5/1) <i>email</i> (1/1) emoção (2/1) ligação (1/1)				moral (1/1) ocasião (1/1) paisagem (2/1) pensão (1/1) <i>t-shirt</i> (1/1)				computador (1/1) cor (2/1) maquilhagem (1/1)				matriz (1/1) mensagem (1/1) origem (1/1)				sandes (1/1) vez (7/1)														
Alemão		<b>televisão (11/5)</b> <b>viagem (12/5)</b> <b>razão (8/3)</b> bar (9/2) fim de semana (6/2) impressão (5/2) vez (41/2) voz (1/2)				andar (10/1) catedral (1/1) cor (2/1) emoção (1/1) férias (24/1) habitação (1/1) <i>internet</i> (1/1) mão (2/1)				mar (7/1) maré (1/1) nível (6/1) observação (1/1) país (26/1) passagem (2/1) paz (1/1) pensão (1/1)				profissão (3/1) sandes (1/1) situação (4/1) sofá (3/1) <i>tour</i> (2/1) tradição (3/1)				<b>luz (2/3)</b> <b>verão (12/3)</b> <b>viagem (29/3)</b> contradição (4/2) desvantagem (3/2) informação (1/2) mão (4/2) animal (2/1)				calor (2/1) capital (4/1) casal (1/1) emoção (1/1) férias (15/1) <i>internet</i> (3/1) lugar (20/1)				modificação (1/1) paisagem (7/1) pombal (2/1) razão (4/1) televisão (18/1) tradição (9/1) vez (56/1)				<b>cor (4/4)</b> discussão (1/1) imagem (3/1) lei (1/1) margem (1/1)				país (12/1) porção (1/1) profissão (1/1) <i>reality-show</i> (2/1) refeição (2/1)				solidão (3/1) televisão (3/1) televisor (1/1) voz (1/1)										
	Inglês	<b>fim de semana (6/5)</b> <b>vez (15/5)</b> <b>viagem (20/4)</b> <b>opinião (3/3)</b> café (5/2) flor (3/2)				lugar (11/2) amor (1/1) capital (3/1) casal (9/1) computador (2/1)				cor (1/1) férias (6/3) manhã (5/1) mensagem (1/1) óculos (1/1)				paisagem (1/1) relação (1/1) relaxação (1/1) sul (5/1) televisão (6/2)				<b>vez (32/4)</b> <b>café (10/3)</b> <b>função (3/3)</b> cor (4/2) dor (2/2) refeição (2/2) razão (8/2)				tradição (6/2) vantagem (8/2) jardim (6/2) bar (9/1) camião (1/1) catedral (3/1) chaminé (1/1)				colonização (1/1) fim de semana (3/1) imagem (5/1) <i>internet</i> (3/1) intenção (1/1) mês (6/1) nação (7/1)				natal (10/1) órgão (1/1) país (38/1) população (2/1) questão (8/1) raiz (4/1)				<b>viagem (5/3)</b> <b>situação (5/2)</b> alimentação (1/1) casal (11/1) comparação (1/1)				imagem (2/1) informação (13/1) jantar (1/1) lugar (4/1)				ordem (2/1) país (38/1) paisagem (1/1) poder (5/1)				razão (7/1) revolução (3/1) solução (1/1) vantagem (2/1)		
Chinês		<b>par (4/3)</b> <b>bar (2/2)</b> <b>paisagem (4/2)</b> <b>vez (21/2)</b>				computador (5/1) diversão (1/1) estação (2/1)				exposição (1/1) férias (10/1) garagem (1/1)				margem (1/1) razão (1/1) situação (2/1)				<b>férias (20/4)</b> <b>coração (2/3)</b> <b>cartão (4/3)</b> <b>edredão (4/3)</b>				viagem (6/2) exposição (2/1) final (5/1)				frustração (1/1) imagem (1/1) mês (5/1)				natal (4/1) paisagem (2/1) postal (2/1)				capital (1/1) férias (8/1) material (2/1) população (1/1)														

Quadro 4.24 – Distribuição de nomes afetados da categoria 4 atemáticos (Ø), por segmentos da amostra (LM e nível QECRL), com a indicação do número de ocorrências no *corpus* e de desvios por item

A partir dos dados apresentados no Quadro 4.24, é possível verificar o elevado número de nomes atemáticos ( $\emptyset$ ) com desvios de AGN e de CNG. De facto, apurou-se no *corpus* um total de 421 desvios que incidem sobre 136 nomes atemáticos distintos.

Atendendo à estrutura morfológica, o número de desvios mais elevado por item concentra-se em nomes deverbais derivados com *-agem*. Com efeito, e à exceção dos dados recolhidos de textos produzidos por falantes nativos de chinês, regista-se, nos restantes segmentos da amostra, uma maior proporção de desvios por item sobre os nomes femininos *passagem* e *reciclagem*, tal como atestam os seguintes exemplos:

- (85) «**O reciclagem** é <uma coi> algo que temos que fazer por respeito do ambiente e dos gerações futuras.» (Italiano.A1-A2)
- (86) «Para **o passagem** do fino de ano, foi a uma festa na casa de um amigo erasmus, mais (p) exactamente no suo te«lhado. (Espanhol.A1-A2)

É igualmente assinalável o número de desvios no nome *viagem* que, não sendo uma forma derivada, pode ter sido reanalisada pelos aprendentes como tal. Assim, e à semelhança do que se viu em *passagem* e *reciclagem*, ao item *viagem* é atribuído, tipicamente, o valor de género masculino:

- (88) «Durante os primeiros dias **do meu viagem** a bordo do submarino vi muitas coisas» (Italiano.A1-A2);
- (89) «Eu espero que gostaste <tambes> tambem dos últimos dias **do viagem**.» (Alemão.B1-B2)
- (90) «Por o fim de semana, Ana planeio **um viagem muito bom**.» (Inglês.C1)

Quanto ao número de ocorrências de nomes afetados no *corpus* e respetivo número de desvios, entre os falantes nativos de espanhol a frequentar turmas de níveis A1-A2 regista-se um maior número de desvios por item nominal relativamente ao verificado nos restantes níveis, sendo que o número de nomes afetado não diminui substancialmente na *passagem* dos níveis A1-A2 para B1-B2 e destes para C1.

Note-se ainda que entre os informantes de LM espanhola, há 2 nomes que surgem com desvios de AGN e de CNG nos vários níveis QECRL (*viagem* e *dor*) ainda que o número de desvios que sobre eles incide vá diminuindo à medida que os aprendentes progredem na aprendizagem do português. A presença constante de desvios nestes itens poderá justificar-se atendendo ao facto de a forma correspondente de *viagem* no espanhol ser masculina (*el viaje*) e da forma correspondente de *dor* ser igualmente masculina (*el dolor*) e, portanto, não coincidentes com o valor de género selecionado em português para

as formas correspondentes. Na parte final da presente subsecção, retomaremos esta questão, com a ponderação dos eventuais efeitos da transferência linguística para compreender os desvios de AGN e de CNG apurados na categoria 4 no segmento de informantes de LM espanhola.

Nas produções escritas por falantes de LM italiana, regista-se uma maior diversidade de nomes afetados nos níveis A1-A2, relativamente à registada no segmento da amostra de informantes que são falantes nativos de espanhol. À medida que vão progredindo na aprendizagem do português, quer o número de ocorrências de nomes afetados, quer o número de ocorrências de desvios diminui consideravelmente, sendo que a maioria dos nomes afetados em A1-A2 não reaparece com desvios nos níveis subsequentes.

Para além disso, também no segmento da amostra dos informantes que são falantes nativos de italiano, dos níveis A1-A2, o nome *viagem* apresenta índices de desvio consideravelmente altos relativamente aos registados nos restantes itens afetados neste segmento, justificável, talvez, pela não coincidência de valor de género com a forma correspondente no italiano (*il viaggio*).

Todavia, nem sempre os comportamentos desviantes assinalados nas produções escritas por informantes de LM italiana se deverão a uma transferência direta de valores de género. Com efeito, neste segmento da amostra nos níveis A1-A2 e B1-B2, se atendermos a um subconjunto de desvios sobre os itens femininos, que correspondem a nomes deverbais derivados com *-ção* (*liquidação, aplicação, civilização, formação, habitação, investigação e motivação*), constatamos que, no italiano, as formas correspondentes são igualmente femininas. Assim, a transferência dos valores de género não poderá estar na origem dos índices de desvios registados nestas formas (veja-se, mais adiante, o Quadro 4.25). O que poderá, então, estar em causa é a assunção, por parte destes aprendentes nos níveis A1-A2 e B1-B2, de que estas formas terminam em *-o* e, por conseguinte, no momento de atribuir o valor de género ao nomes, aplicam a pseudorregra de atribuição de valores de género “se o nome termina em *-o* é de género masculino; se termina em *-a* é feminino”:

- (91) «Personalmente, por **estes motivações** eu não estou ao computador mais de uma hora por vez.» (Italiano.A1-A2)
- (92) «\***O primeiro liquidacao** fosse de "Nuragici" (...)» (Italiano.A1-A2)

- (93) «\*Alí podemos produzir o que nós precisamos sem ter /quase/ **algum ligação** com a inflação e a crise que os habitantes das cidades têm.»  
(Italiano.B1-B2)

A mesma situação poderá estar em causa para justificar o número de desvios assinalado, por item, nos femininos terminados em *-ão* noutros segmentos da amostra, nomeadamente nos dados extraídos de textos produzidos, nos diferentes níveis QECRL, de A1 a C1, por informantes que são falantes nativos de alemão, inglês e chinês:

- (94) «\*(...) é um lugar tão especial devido à cultura (incluído a comida magnífica) e **todos os tradições** que os japoneses e as turistas também, respetam e continuam ser influenciados.» (Inglês.B1-B2)
- (95) «\***O situação** esteve muito bem e porque ele fechou a porta com um suspiro alívio.» (Alemão.A1-A2)
- (96) «\*Para mim elas são **o razão** que Barcelona é mercida ser chamada uma cidade de arte.» (Chinês.A1-A2)
- (97) «(...) não havia heróis **neste revolução** como o filme mostrava, mas dava bem para a drama.» (Inglês.C1)
- (98) «No caso de José Saramago **este solidão** pode vir do facto (...)»  
(Alemão.C1)

Também Pinto (2015) assinalara este fenómeno no seu estudo. Para o investigador, o número elevado de desvios sobre formas femininas atemáticas terminadas em *-ão* é justificável pelo forte impacto que a forma escrita da palavra tem na aquisição/aprendizagem tardia de uma LNM.



Numa fase final da análise dos desvios da categoria 4, procurou-se aferir o eventual efeito direto da LM na assimilação dos valores de género dos nomes. Assim, consideraram-se os dados recolhidos dos segmentos da amostra de informantes que são falantes nativos de espanhol, italiano e alemão, já que estes idiomas possuem a categoria de género gramatical.

Em primeiro lugar, considerem-se os desvios assinalados nos textos escritos pelos falantes nativos das línguas românicas, espanhol e italiano. Assim, em cada nome afetado, verificou-se qual seria a forma lexical correspondente na LM do informante, tendo-se

igualmente assinalado o valor de género associado<sup>175</sup>. Deste modo, sempre que se verificara a não correspondência de valores de género entre os itens, na LM do aprendiz e na LA de aprendizagem, considerou-se que os desvios são potencialmente atribuíveis à transferência linguística de valores de género do idioma nativo.

É preciso, porém, ressaltar que nem sempre foi possível determinar, com precisão, qual seria o item correspondente da LM e, em consequência disso, se atuaria, nesses casos, uma possível estratégia de transferência linguística.

Tendo em conta os pressupostos acima referidos, apresenta-se no Quadro 4.25 a distribuição dos itens nominais afetados da categoria 4 nos textos produzidos pelos aprendizes que são falantes de LM espanhola e italiana, a frequentar turmas dos níveis QECRL A1-A2, B1-B2 e C1, com a indicação dos nomes correspondentes na LM e dos valores de género a eles associados.

Em seguida, no Quadro 4.26, regista-se a distribuição do número absoluto de desvios considerados como potencialmente atribuíveis à transferência linguística e do seu peso relativo nos vários segmentos da amostra, ou seja, apresenta-se o número de casos nos quais não se observa a correspondência de valores de género, masculino ou feminino, entre os itens lexicais afetados da LA de aprendizagem e os correspondentes da LM dos informantes<sup>176</sup>.

---

<sup>175</sup> Cf. nota 171.

<sup>176</sup> Convém ressaltar que não foram contabilizados como “desvios potencialmente atribuíveis à transferência da LM” os casos em que se identificou, para um nome, duas ou mais formas lexicais correspondentes na LM do falante aprendiz com diferentes valores de género, coincidentes ou não com os valores da forma-alvo, já que nestes não é possível inferir um eventual efeito negativo da transferência linguística.



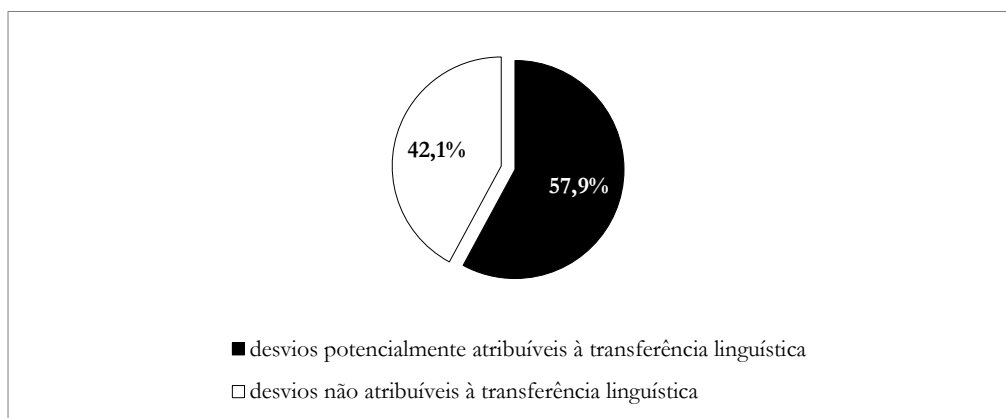
Informantes		Valor de género da palavra em português → Valor de género da palavra na LM do informante										
LM	Nível QECRL	Masc. → Masc.		Masc. → Fem.		Fem. → Fem.		Fem. → Masc.		Masc. ou Fem. → Masc. / Fem.		
Espanhol	A1-A2	* cinema: <i>cinema</i>		* costume: <i>costumbre</i> * duche: <i>ducha</i> * lençol: <i>sábana / manta</i>	* sangue: <i>sangre</i> * sinal: <i>señal</i>	* cidade: <i>ciudad</i> * noite: <i>noche</i> * passagem (de ano): <i>Nochevieja</i>	* qualificação: <i>cualifición</i> * t-shirt: <i>camiseta</i>	* árvore: <i>árbol</i> * cor: <i>color</i> * dor: <i>dolor</i>	* férias: <i>vacaciones (vacación)</i> * garagem: <i>garaje</i>	* mestiçagem: <i>mestizaje</i> * viagem: <i>viaje</i>	* autocolante: <i>pegatina (Fem.) / adhesivo (Masc.)</i>	
	B1-B2	* país: <i>país</i> * par: <i>par</i> * problema: <i>problema</i>	* costume: <i>costumbre</i> * labor: <i>labor</i>	* legume: <i>legumbre</i> * leite: <i>leche</i>	* chatice: <i>moléstia</i> * competição: <i>competición</i>	* noite: <i>noche</i> * parte: <i>parte</i>	* desordem: <i>desorden</i> * dor: <i>dolor</i>	* origem: <i>origen</i> * viagem: <i>viaje</i>				
	C1		* costume: <i>costumbre</i> * labor: <i>labor</i>		* instalação: <i>instalación</i> * vantagem: <i>vantaja</i>		* análise: <i>análisis</i> * dor: <i>dolor</i> * rodagem: <i>rodaje</i>					
Italiano	A1-A2	* dia: <i>giorno</i> * exame: <i>esame</i> * pastel: <i>pastello</i> * problema: <i>problema</i>	* restaurante: <i>ristorante</i> * sistema: <i>sistema</i> sofá: <i>sofà</i>	* ar: <i>aria</i> * duche: <i>doccia</i> * fim: <i>fine</i>	* mapa: <i>mappa</i> * verão: <i>estate</i>	* aplicação: <i>applicazione</i> * capital: <i>capitale</i> * civilização: <i>civiltà</i> * espécie: <i>specie</i> * formação: <i>formazione</i> * habitação: <i>abitazione</i> * hipótese: <i>ipotesi</i>	* internet: <i>internet</i> * investigação: <i>investigazione</i> * lebre: <i>lepre</i> liquidação: <i>liquidazione</i> * motivação: <i>motivazione</i> * noite: <i>notte</i> * novidade: <i>novità</i> * oportunidade: <i>opportunità</i>	* origem: <i>origine</i> * parte: <i>parte</i> * televisão: <i>televisione</i> * universidade: <i>università</i> * versão: <i>versione</i> * vez: <i>volta</i> * visão: <i>visione</i>	* tapete: <i>tappeto</i> * cor: <i>colore</i> * coragem: <i>coraggio</i> * divisão: <i>vano</i>	* estante: <i>scalfale</i> * paisagem: <i>paesaggio</i> * passagem: <i>passaggio</i> * patinagem: <i>patinaggio</i>	* reciclagem: <i>riciclaggio</i> * tarde: <i>pomeriggio</i> * viagem: <i>viaggio</i>	
	B1-B2	* dia: <i>giorno</i> * horizonte: <i>orizzonte</i> restaurante: <i>ristorante</i>		* ar: <i>aria</i>		* atenção: <i>attenzione</i> * atividade: <i>attività</i> * canção: <i>canzone</i> * cidade: <i>città</i>	* emoção: <i>emozione</i> * faculdade: <i>facoltà</i> * internet: <i>internet</i> * ocasião: <i>occasione</i>	* pensão: <i>pensione</i> * possibilidade: <i>possibilità</i> * região: <i>regione</i> * t-shirt: <i>maglietta / camicia</i> * vez: <i>volta</i>	* árvore: <i>albero</i> * atitude: <i>atteggiamento</i> * ave: <i>uccello</i> * cor: <i>colore</i> * dor: <i>dolore</i>	* ligação: <i>legame</i> * paisagem: <i>paesaggio</i> * vantagem: <i>vantaggio</i> * viagem: <i>viaggio</i>	moral (Fem.) – <i>moral</i> (Masc. / Fem.)	
	C1	computador: <i>computer</i>			* matriz: <i>matrice</i> * origem: <i>origine</i>	* probabilidade: <i>probabilità</i> * vez: <i>volta</i>	* cor: <i>colore</i> * maquilhagem: <i>trucco</i>	* mensagem: <i>messaggio</i> * sandes: <i>sandwich</i>				

Quadro 4.25 – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 4, detetados nas produções escritas por informantes de LM espanhola e italiana, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical

Informantes		Desvios da Categoria 4		
LM	Nível QECRL	Total	# de desvios potencialmente atribuíveis à transferência linguística	%
Espanhol	A1-A2	43	31	72,1
	B1-B2	22	13	59,1
	C1	8	6	75
Italiano	A1-A2	112	67	59,8
	B1-B2	53	22	41,5
	C1	9	4	44,4
$\Sigma$		247	143	57,9

**Quadro 4.26** – Distribuição do número de desvios da categoria 4 produzidos nos segmentos da amostra de informantes de LM espanhola e italiana potencialmente atribuíveis à transferência da LM

Tendo em conta os dados apresentados no Quadro 4.26, o Gráfico 4.28 apresenta, em valores percentuais, a proporção de desvios de AGN e de CNG detetados nas produções escritas por informantes que são falantes nativos de espanhol e de italiano, potencialmente atribuíveis a um efeito de transferência linguística.



**Gráfico 4.28** – Percentagem de desvios da categoria 4 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM espanhola e italiana

Os desvios da categoria 4 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes que são falantes nativos de espanhol e de italiano representam, proporcionalmente, cerca de 60% do total de casos assinalados. Ou seja, mais de metade dos desvios de AGN e de CNG da categoria 4 produzidos pelos segmentos da amostra de falantes nativos de espanhol e de italiano nos vários níveis QECRL devem-se, potencialmente, a um efeito da transferência direta dos valores de género da LM para a LA de aprendizagem.

Parece que face à fraca robustez do *input* da LA relativamente à atribuição dos valores de género nominal, os aprendentes tendem a associar o valor de género ao nome em função do valor que o item lexical correspondente tem na sua LM. Tal como referido

anteriormente nesta subsecção, são especialmente expressivos, nestes segmentos da amostra, os desvios sobre as formas femininas *viagem* e *dor*, itens lexicais cujos correspondentes no espanhol e no italiano são de género masculino (cf. Quadro 4.26).

Veja-se, a este respeito, os seguintes exemplos recolhidos do *corpus*:

- (99) «\***O viagem** de ela foi muito comprido (...)» (Espanhol.A1-A2)
- (100) «\*Durante **o viagem** ela começou a <(...)> imaginar a sua permanência em este novo país» (Italiano.A1-A2)
- (101) «\*(...) o dia seguinte pode acontecer que exista **um pequeno dor** de cabeça...» (Espanhol.B1-B2)
- (102) «\*Certamente temos de compreender **o dor** das pessoas» (Italiano.B1-B2)

Para além disso, ao contrário do que se verificara no caso dos nomes pertencentes à categoria 3, não é recorrente a importação direta dos vocábulos da LM para os textos, havendo, apenas, a transferência de valores de género que, ao não serem coincidentes, conduzem o informante ao desvio.

Veja-se, agora, os desvios de AGN e de CNG da categoria 4 assinalados nas produções escritas pelos informantes que são falantes nativos de alemão. À semelhança do que foi feito anteriormente, apresenta-se, em primeiro lugar, no Quadro 4.27, a distribuição dos nomes afetados por valor de género na LA e dos itens lexicais correspondentes em alemão com a indicação do valor de género a eles associados<sup>177</sup>. Como o alemão apresenta um sistema tripartido de valores de género, não há, por conseguinte, coincidência total do número de valores de género deste idioma para o português. Por este motivo, pretende-se averiguar se é possível estabelecer uma correspondência, direta ou indireta, entre os valores de género dos itens da LM e os valores de género dos itens da LA. Para além disso, procura-se determinar se existe um padrão recorrente de associação de um determinado valor em português (masculino ou feminino) a nomes cujo item lexical correspondente em alemão é de valor neutro. Ou seja, procura-se apurar se há, nestes casos, uma marcação preferencial, por parte dos informantes, do valor de género masculino ou feminino.

---

<sup>177</sup> Cf. nota 173.

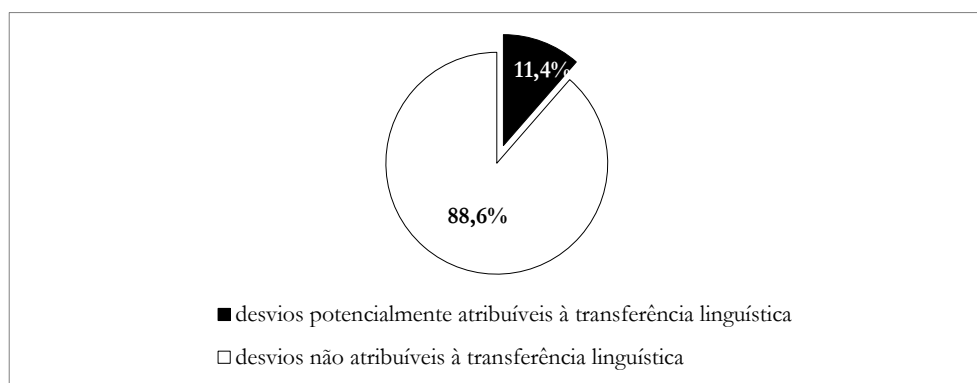
Posteriormente, no Quadro 4.28, regista-se a distribuição do número de comportamentos desviantes considerados como potencialmente atribuíveis à transferência direta de valores de género, i.e., os casos em que não se observa uma correspondência direta de valores (masculino ou feminino) entre os itens lexicais afetados e os itens correspondentes em alemão. Assim sendo, no Quadro 4.28 é apresentado (i) o número absoluto de desvios da categoria 4 produzido por nível QECRL no segmento da amostra de informantes de LM alemã; (ii) o número de desvios que, em função dos itens afetados, são potencialmente atribuíveis à transferência de valores de género (masculino ou feminino); e (iii) o peso relativo, em valores percentuais, dos desvios da categoria 4 potencialmente atribuíveis à transferência linguística, calculado a partir do número absoluto de ocorrências de desvios por nível QECRL.

Valor de género da palavra em português → Valor de género da palavra na LM do informante							
Nível QECRL	Masc. → Masc.	Masc. → Fem.	Masc. → Neut.	Fem. → Fem.	Fem. → Masc.	Fem → Neut.	Masc. ou Fem. → Masc. / Fem. / Neut.
A1-A2	<ul style="list-style-type: none"> <li>• andar: <i>Stock</i></li> <li>• tour: <i>Rundgang</i></li> </ul>	<ul style="list-style-type: none"> <li>• bar: <i>Bar</i></li> </ul>	<ul style="list-style-type: none"> <li>• fim de semana: <i>Wochenende</i></li> <li>• nível: <i>Niveau</i></li> <li>• país: <i>Land</i></li> <li>• sofá: <i>Sofa</i></li> </ul>	<ul style="list-style-type: none"> <li>• catedral: <i>Kathedrale</i></li> <li>• conversação: <i>Konversation</i></li> <li>• cor: <i>Farbe</i></li> <li>• emoção: <i>Emotion</i></li> <li>• habitação: <i>Habitation</i></li> <li>• informação: <i>Auskunft</i></li> <li>• mão: <i>Hand</i></li> <li>• maré: <i>Flut</i></li> <li>• observação: <i>Beobachtung</i></li> <li>• pensão: <i>Pension</i></li> <li>• situação: <i>Lage</i></li> <li>• tradição: <i>Tradition</i></li> <li>• viagem: <i>Reise</i></li> <li>• voz: <i>Stimme</i></li> </ul>	<ul style="list-style-type: none"> <li>• férias: <i>Urlaub</i></li> <li>• impressão: <i>Eindruck</i></li> <li>• passagem: <i>Durchgang</i></li> <li>• paz: <i>Friede</i></li> <li>• profissão: <i>Beruf</i></li> <li>• razão: <i>Grund</i></li> </ul>	<ul style="list-style-type: none"> <li>• coabitação: <i>Zusammenleben</i></li> <li>• compaixão: <i>Mitleid</i></li> <li>• internet: <i>Internet</i></li> <li>• televisão: <i>Fernsehen</i></li> <li>• vez: <i>Mal</i></li> </ul>	<ul style="list-style-type: none"> <li>• mar: <i>Meer</i> (Neut.) / <i>See</i> (Fem.)</li> <li>• sandes: <i>Sandwich</i> (Masc. / Neut.)</li> </ul>
B1-B2	<ul style="list-style-type: none"> <li>• lugar: <i>Ort</i></li> <li>• pombal: <i>Taubenschlag</i></li> <li>• verão: <i>Sommer</i></li> </ul>	<ul style="list-style-type: none"> <li>• calor: <i>Wärme, Hitze</i></li> </ul>	<ul style="list-style-type: none"> <li>• animal: <i>Tier</i></li> <li>• casal: <i>Paar, Männchen</i></li> </ul>	<ul style="list-style-type: none"> <li>• emoção: <i>Emotion</i></li> <li>• informação: <i>Auskunft</i></li> <li>• mão: <i>Hand</i></li> <li>• modificação: <i>Änderung</i></li> <li>• paisagem: <i>Landschaft</i></li> <li>• tradição: <i>Tradition</i></li> <li>• viagem: <i>Reise</i></li> <li>•</li> </ul>	<ul style="list-style-type: none"> <li>• contradição: <i>Widerspruch</i></li> <li>• desvantagem: <i>Nachteil</i></li> <li>• férias: <i>Urlaub</i></li> <li>• razão: <i>Grund</i></li> </ul>	<ul style="list-style-type: none"> <li>• capital: <i>Kapital</i></li> <li>• internet: <i>Internet</i></li> <li>• luz: <i>Licht</i></li> <li>• televisão: <i>Fernsehen</i></li> <li>• vez: <i>Mal</i></li> </ul>	
C1	<ul style="list-style-type: none"> <li>• televisor: <i>Fernseher</i></li> </ul>	<ul style="list-style-type: none"> <li>• reality-show (o programa de): <i>Reality-Show</i></li> </ul>	<ul style="list-style-type: none"> <li>• país: <i>Land</i></li> </ul>	<ul style="list-style-type: none"> <li>• cor: <i>Farbe</i></li> <li>• discussão: <i>Diskution</i></li> <li>• refeição: <i>Mahlzeit</i></li> <li>• solidão: <i>Einsamkeit</i></li> <li>• voz: <i>Stimme</i></li> </ul>	<ul style="list-style-type: none"> <li>• margem: <i>Rand</i></li> <li>• profissão: <i>Beruf</i></li> </ul>	<ul style="list-style-type: none"> <li>• imagem: <i>Bild</i></li> <li>• lei: <i>Gesetz</i></li> <li>• porção: <i>Stück</i></li> <li>• televisão: <i>Fernsehen</i></li> </ul>	

Quadro 4.27 – Distribuição, em função do valor de género gramatical, dos nomes afetados na categoria 4, detetados nas produções escritas por informantes de LM alemã, com a indicação da forma lexical correspondente na LM dos aprendentes e dos respetivos valores de género gramatical

Nível QECRL	Desvios da Categoria 4		
	Total	# de desvios potencialmente atribuíveis à transferência linguística de valores de género (masculino ou feminino)	%
A1-A2	102	11	10,7
B1-B2	56	7	12,5
C1	26	3	11,5
$\Sigma$	184	21	11,4

**Quadro 4.28** – Distribuição do número de desvios da categoria 4 produzidos nos segmentos da amostra de informantes de LM alemã potencialmente atribuíveis à transferência da LM



**Gráfico 4.29** – Percentagem de desvios da categoria 4 potencialmente atribuíveis à transferência linguística detetados nas produções escritas por informantes de LM alemã

De acordo com os dados apresentados no Quadro 4.28 e também no Gráfico 4.29, é possível verificar que, em valores percentuais, o peso relativo de desvios de AGN e da CNG da categoria 4 potencialmente atribuíveis à transferência linguística de valores de género masculino ou feminino, detetados nas produções escritas por informantes que são falantes nativos de alemão a frequentar turmas de diferentes níveis QECRL (A1-A2, B1-B2 e C1), atinge, sensivelmente, os 11 % por segmentos da amostra de nível QECRL, não havendo, a este respeito, significativas diferenças entre os vários níveis. Assim sendo, não se verifica, no segmento da amostra de informantes de LM alemã um efeito considerável da transferência linguística direta dos valores de género nos desvios de AGN assinalados, já que, em grande parte dos itens afetados, se regista uma correspondência de valores das formas da LM e das formas eventualmente correspondentes da LA.

Quanto aos nomes correspondentes aos itens lexicais afetados e que são de género neutro em alemão, não se regista, na categoria 4, uma incidência, por parte dos aprendentes, em marcar preferencialmente as formas correspondentes do português com um determinado valor de género, masculino ou feminino (cf. Quadro 4.27).

Em suma, a categoria 4 apresenta, por nome, um número de desvios consideravelmente alto em comparação com o número de desvios registados nas demais categorias analisadas. É também assinalável, nesta categoria, os valores médios de desvios mais altos por texto relativamente aos registados em anteriores categorias. A incidência de desvios nesta categoria verifica-se em todos os segmentos da amostra, independentemente da LM do falante aprendiz. Há, contudo, a assinalar a diminuição dos comportamentos desviantes à medida que os aprendentes progridem na aprendizagem do português, para além de, e salvo casos pontuais, os nomes com desvios de AGN nos níveis A1-A2 não reaparecerem com desvios no nível C1.

Nota-se ainda, entre os comportamentos desviantes assinalados nos falantes de LM espanhola e italiana, um efeito de transferência linguística direta da LM, visível, sobretudo, na incidência recorrente de desvios de atribuição de valor de género e de concordância nominal em género num subconjunto de itens nos quais não se observa a correspondência de valores de género entre as formas da LA e as correspondentes da LM do informante. Assim, face à ambiguidade do *input*, estes aprendentes não são capazes de reconhecer o valor de género do nome aplicando critérios, quer formais, quer semânticos, e, em consequência disso, recorrem aos valores de género correspondentes da sua LM, que, ao não serem coincidentes com os das formas do português, conduzem ao desvio.

Por fim, é de salientar também o número de desvios em nomes femininos atemáticos terminados em *-ã*, assinalável em grande parte dos segmentos da amostra. Como se viu, tais manifestações desviantes poderão estar correlacionadas com a tendência para os aprendentes tardios valorizarem o facto de estas formas terminarem em *-o*, o que, por conseguinte, os conduz à atribuição do valor de género masculino que, tipicamente, se associa a este constituinte temático. Assim sendo, estes dados são reveladores do peso significativo que a forma escrita da palavra tem ao longo do processo de aquisição/aprendizagem lexical do PLN.





## Considerações Finais

---

A investigação que agora se finaliza centra-se na aquisição/aprendizagem do sistema de atribuição dos valores de género aos nomes e respetiva concordância em PLN<sub>M</sub>, através da análise de desvios de atribuição de género nominal (AGN) e de concordância nominal em género (CNG), dentro do sintagma nominal (SN), detetados num conjunto de produções escritas por aprendentes tardios. Pretendeu-se, em particular, averiguar o papel (ou papéis) do conhecimento linguístico prévio proveniente da LM do aprendente para o processo de assimilação da categoria de género gramatical em PLN<sub>M</sub>, tendo-se, por isso, selecionado textos escritos por aprendentes que são falantes nativos de espanhol, italiano, alemão, inglês e chinês, a frequentar turmas de diferentes níveis QECL, de A1 a C1.

Como se viu no **Capítulo 1**, o género gramatical constitui uma área crítica para a aquisição/aprendizagem tardia de uma LNM (Oliphant 1997; Franceschina 2005; Ferreira 2011; Martins 2015; Pinto 2015), não o sendo, em contrapartida, na aquisição de um idioma nativo (Clark 1998; Costa *et al.* 2015; Corrêa & Augusto 2017) (cf. Secção 1.5.). Com efeito, no contexto da assimilação tardia das estruturas de uma LNM, os desvios relativos a esta estrutura nunca chegam a ser erradicados por completo, mesmo nos enunciados produzidos por aprendentes tardios de níveis QECL mais avançados. Tendo em conta esta evidência, diferentes fatores têm vindo a ser aduzidos na literatura especializada de forma a justificar a relativa ‘resistência’ desta categoria gramatical ao longo das diferentes fases do processo de aquisição/aprendizagem tardia das estruturas de uma LNM (Hawkins & Chan 1997; Franceschina 2003, 2005; Tsimpli & Dimitrakopoulou 2007).

Por um lado, a dificuldade observada pode ser perspectivada em função das particularidades do *input* linguístico ao qual o falante aprendente está exposto. No caso específico do português, a LA de aprendizagem, o valor de género de um nome não é inequivocamente dedutível a partir dos marcadores morfológicos, i.e., dentro das fronteiras do item nominal. Com efeito, e apesar de se observar uma correlação, ainda que parcial, entre valores de género masculino e feminino e índices temáticos *-o* e *-a*, respetivamente, nem sempre é possível estabelecer essa associação, havendo itens masculinos terminados

em *-a* como, por exemplo, *o dia*, e femininos terminados em *-o* como *a tribo*, embora estes últimos sejam muito pouco frequentes no léxico do português (cf. Capítulo 1, Quadro 1.3). Na ausência, igualmente, de critérios semânticos capazes de orientar os aprendentes tardios de PLNM na seleção dos valores adequados de género nominal, dados empíricos relativos à aquisição/aprendizagem de PLNM revelam uma certa ‘desconfiança’ por parte dos aprendentes tardios relativamente ao poder preditivo dos índices morfológicos para este efeito (Ferreira 2011; Martins 2015; Pinto 2015) que poderá estar, então, correlacionada com a perceção, por parte do aprendente, da fraca robustez do *input* a que está exposto.

Para além disso, a categoria de género gramatical encontra-se intrinsecamente relacionada com o fenómeno da concordância sintática, sendo que os vários elementos que integram o sintagma nominal devem concordar em género com o nome (Corbett 1991, 2006a; Choupina *et al.* 2016; Mota 2016b) (cf. Secção 1.1.). Ora, no contexto da assimilação linguística tardia, dados empíricos atestam precisamente a dificuldade, por parte dos aprendentes, em dominar, na sua plenitude, o mecanismo de concordância sintática (Leiria 2006; Martins 2015).

Por outro lado, os comportamentos desviantes detetados na atribuição de valores de género aos nomes e da concordância nominal em género também podem ser analisados em função do perfil do aprendente tardio, nomeadamente no que concerne à LM e, em particular, às características dos respetivos sistemas de classificação nominal. Tendo em conta este pressuposto, foram selecionadas para este estudo, produções escritas por aprendentes tardios que são falantes nativos de espanhol, italiano, alemão, inglês e chinês.

No **Capítulo 1**, na Secção 1.4., procedeu-se à descrição e análise dos sistemas de classificação nominal dos idiomas selecionados. Neste conjunto, distinguem-se os idiomas com categoria de género, como o espanhol (c. Secção 1.4.1.), o italiano (cf. Secção 1.4.2.) e o alemão (cf. Secção 1.4.3.), dos que não possuem esta categoria gramatical, como o inglês (cf. Secção 1.4.4.) e o chinês (cf. Secção 1.4.5). Entre as línguas com sistemas de classificação nominal baseados em valores de género, o espanhol e o italiano têm sistemas de dois valores em oposição (masculino e feminino), com particularidades de marcação desses valores muito idênticas às do português. Em contrapartida, o alemão apresenta um sistema de classificação dos nomes baseado em três valores, masculino, feminino e neutro, para além de exibir um sistema de concordância nominal complexo, admitindo a marcação dos valores de género, número e de caso. O inglês não apresenta sistema de género nos nomes, especificadores e adjetivos e o chinês não possui a categoria de género

morfologicamente expressa, nem admite a realização do mecanismo de concordância sintática.

Deste modo, parte-se do pressuposto de que o aprendente tardio cuja LM possui um sistema de classificação dos nomes baseado em valores de género gramatical com propriedades de atribuição e de concordância nominal muito próximas das do sistema de classificação nominal do português apresentará, no que à marcação de valores de género diz respeito, um desempenho global satisfatoriamente melhor do que aquele que é registado no aprendente tardio, falante nativo de um idioma sem sistema de classificação nominal baseado em valores de género.

De facto, numa parte substancial da investigação produzida no âmbito da Aquisição de Línguas Não Maternas (ALNM), considerada no **Capítulo 2**, é possível encontrar diferentes trabalhos empíricos nos quais se avalia, precisamente, o peso da configuração do conhecimento linguístico prévio proveniente da LM para a assimilação dos valores de género nominal em contexto de aquisição/aprendizagem tardia de uma LNM (cf. Capítulo 2, Secção 2.3.3.). Mais precisamente, em alguns estudos, pondera-se a relevância do conhecimento linguístico prévio da LM para o desenvolvimento do conhecimento gramatical da LA e de como este interage com os mecanismos inatos da aquisição linguística para a assimilação plena da categoria gramatical de género e da concordância sintática, por aprendentes tardios.

Pesem embora os vários contributos, não há, contudo, uma clara compreensão relativamente ao modo como a configuração do conhecimento linguístico prévio proveniente da LM atua, direta e indiretamente, ao longo das diferentes fases de desenvolvimento da interlíngua do aprendente tardio, e no modo como o aprendente vai mobilizando estratégias e vai processando as distintas particularidades subjacentes ao processo de atribuição dos valores de género nominal aos itens lexicais (cf. Bruhn de Garavito & White 2002; Franceschina 2003, 2005; White *et al.* 2004; Sabourin *et al.* 2006).

Atendendo ao que é postulado na literatura especializada, procedeu-se, no **Capítulo 3**, à descrição detalhada do método de seleção e de recolha das produções escritas que compõem a base empírica do estudo e, mais precisamente, dos procedimentos adotados para a seleção dos desvios de AGN e de CNG que foram objeto de análise. Indicou-se ainda, neste Capítulo referente à metodologia, as perguntas de investigação e as variáveis de análise selecionadas para este estudo, relacionadas com o perfil dos informantes (LM e nível QECRL) e com as características do sistema de atribuição de género nominal do

português. Para além disso, fez-se a referência ao perfil dos informantes, autores dos textos escritos que compõem o *corpus*, com indicação da sua distribuição por LM, outras LNM, nível QECRL em português, idade, período de exposição ao português e experiência de imersão (cf. Secção 3.3.).

Ainda neste capítulo, na parte final, apresentou-se a tipologia de desvios. Esta tipologia, elaborada em função das propriedades de associação dos valores de género aos nomes do português, serviu, posteriormente, de base para a análise dos dados empíricos. Para este estudo, registaram-se os desvios observáveis, quer na forma morfológica do item lexical afetado, quer nos vários elementos que coocorrem no sintagma nominal (especificadores e modificadores com função atributiva). A cada categoria de desvio foi atribuído um código numérico, de 1 a 4. A categoria 1 identifica os desvios que afetam nomes cujo valor de género é, simultaneamente, dedutível a partir de critérios semânticos e formais (*o menino, a menina*). A categoria 2 assinala os desvios que afetam os nomes cujo valor de género é inferível a partir de critérios semânticos, mas não de critérios formais (*o homem, a mulher*). A categoria 3 congrega os desvios que incidem sobre nomes cuja associação de valor de género depende de critérios formais, mas não de critérios semânticos (*o carro, a casa, o indivíduo, a pessoa*). Por fim, a categoria 4 assinala os desvios em itens cujo valor de género não depende nem de critérios semânticos nem de critérios formais (*o pente, a ponte, o coração, a viagem*).

Dentro de cada categoria, identificaram-se ainda os constituintes nos quais incidem as marcas de desvio de atribuição de valores de género: determinantes, quantificadores, nomes e adjetivos com função atributiva.

Estipulada a tipologia, procedeu-se, no **Capítulo 4** à apresentação dos resultados obtidos pela análise dos dados empíricos selecionados. Por conseguinte, ao longo deste capítulo, procedeu-se à averiguação, quer do ponto de vista quantitativo, quer do ponto de vista qualitativo, dos desvios de AGN e de CNG detetados nos segmentos da amostra por LM e nível QECRL da turma frequentada pelos informantes, para, a partir desses dados, identificar os (possíveis) fatores atuantes ao longo das diferentes fases do processo de assimilação dos valores de género nominal em PLNM, dos quais se destaca o conhecimento linguístico prévio representado pela LM do aprendente.

Através da análise dos resultados apresentados no Capítulo 4, procurou-se, então, responder às quatro perguntas formuladas na parte introdutória do presente estudo, evocando-se também, e sempre que pertinente, os contributos teóricos considerados previamente neste trabalho.

**P1:** O desempenho linguístico do aprendente tardio relativamente à atribuição dos valores de género aos nomes e à concordância nominal em género no português beneficia do facto de ter representada na gramática da sua LM esta categoria gramatical?

Tomando em linha de conta os resultados globais apurados, verificou-se que quanto mais distante tipologicamente é a LM do falante aprendente relativamente à LA de aprendizagem, maior é o número de desvios de AGN e de CNG que ocorrem. Ou seja, em termos gerais, verificou-se uma maior incidência de desvios nos textos produzidos por informantes que são falantes nativos de idiomas sem a categoria de género gramatical (inglês e chinês), face à proporção de desvios registada nos restantes segmentos da amostra de informantes cuja LM possui um sistema de classificação nominal baseado em valores de género (alemão, italiano e espanhol). Entre estes segmentos da amostra, verificou-se ainda que os falantes nativos de línguas românicas (espanhol e italiano) apresentam índices de desvios mais baixos do que os registados nos textos produzidos pelos falantes nativos de alemão. Portanto, uma maior proximidade tipológica dos idiomas, LM do aprendente e LA aprendizagem, beneficia a aquisição/aprendizagem do sistema de atribuição de género nominal do PLNM.

Contudo, pese embora o melhor desempenho registado entre os informantes cuja LM tem a categoria de género gramatical, os desvios nunca chegam a ser erradicados por completo das suas produções escritas, já que mesmo nos níveis QECRL mais avançados se registam desvios de AGN e de CNG em todos os segmentos da amostra, confirmando-se, assim, que o género é uma área crítica na aquisição/aprendizagem tardia de uma LNM.

Os resultados apurados suscitam algumas considerações quando observados à luz das hipóteses que, na literatura, têm sido evocadas acerca do processo de aquisição/aprendizagem tardia da categoria de género nominal numa LNM, o que agora faremos sem, contudo, nos filiarmos num posicionamento teórico específico.

Em primeiro lugar, os dados apurados não confirmam, a não ser parcialmente, os pressupostos da *Failed Functional Features Hypothesis* (FFFH) (Hawkins & Chan 1997; Franceschina 2003, 2005), segundo a qual a aquisição plena das propriedades parametrizadas, i.e., que não são universais, como o género, somente é possível em fases tardias do desenvolvimento humano quando estas se encontram representadas na gramática do idioma nativo do aprendente. Com efeito, apesar de os aprendentes beneficiarem do facto de terem na sua LM a categoria gramatical de género, os desvios de AGN e de CNG

não chegam a ser erradicados por completo das produções escritas pelos aprendentes de LM espanhola, italiana e alemã.

Para além disso, atendendo a que os aprendentes de LM inglesa e chinesa apresentam, no nível mais avançado da aprendizagem, i.e., no C1, índices de desvio consideravelmente mais baixos dos registados nos níveis iniciais, é revelador que, mesmo não tendo representada na sua LM a categoria gramatical de género, é possível assimilar, com razoável grau de sucesso, as propriedades desta categoria na LA de aprendizagem. Aliás, no segmento de informantes de LM chinesa, os valores de desvio apurados são muito próximos dos registados nos segmentos da amostra de informantes de LM espanhola e italiana em C1, sendo idênticos aos apresentados pelos aprendentes cuja LM é o alemão, uma língua com sistema de género nominal. Partindo da hipótese de que a GU atua no processo de aquisição/aprendizagem de uma LNM, parece, então, haver indícios de que, pelo menos num estado avançado de proficiência, os aprendentes poderão aceder a componentes que configuram este conhecimento linguístico inato.

Como vimos na parte inicial deste trabalho, o facto de os aprendentes proficientes cuja LM não possui um sistema de classificação nominal baseado em valores de género, apresentarem poucas ocorrências desviantes conduz White *et al.* (2004) à conclusão de que, mesmo em fases tardias do desenvolvimento ontogénico do falante, a GU continua acessível. Portanto, e à luz destas premissas, dados desta natureza são compatíveis com a hipótese de *Full Transfer / Full Access* (FTFA) (Schwartz & Sprouse 1996; White *et al.* 2004), já que, quando os aprendentes tardios se deparam com estruturas não convergentes com as da sua LM, acedem à GU de modo a suprir essa ausência e a superar os conflitos de hipóteses que dela decorrem.

Se atendermos novamente ao desempenho linguístico dos aprendentes que são falantes nativos de espanhol, verifica-se ainda que, apesar de os índices de desvio serem globalmente baixos em todos os níveis, a verdade é que, no nível C1, não há uma melhoria significativa do desempenho relativamente ao registado nos níveis anteriores. Portanto, regista-se, neste segmento da amostra, uma estabilização, desde cedo, da aprendizagem dos valores de género nominal do português. Assim, como estabilizam cedo, ficarão impermeáveis aos efeitos do *input* que, assim, não geram nestes aprendentes conflitos de hipóteses sobre a forma que deverá assumir a gramática da LA. Por conseguinte, não havendo conflitos de hipóteses, não há necessidade, por parte destes aprendentes, de os resolver com recurso à informação disponível na GU.

Em suma, para além da LM infere-se, dos resultados desta investigação, que fatores associados ao próprio desenvolvimento do processo de assimilação das propriedades estruturais do *input* ao qual o aprendiz está exposto terão igualmente um papel preponderante para a aquisição/aprendizagem do sistema de atribuição de género nominal em português como LNM. Entre esses fatores, e como Franceschina (2005:197-198) afirma, estará a quantidade do *input* ao qual os aprendizes estão expostos.

**P2:** O valor de género gramatical é uma propriedade transferível da LM para a LNM?

Durante a análise dos desvios de AGN e de CNG, procuraram-se evidências de uma possível transferência de valores de género da LM para os itens correspondentes na LA, o português. Ora, tal averiguação só foi possível nos dados apurados entre os aprendizes cuja LM possui a categoria de género gramatical. Assim sendo, partindo da averiguação dos desvios assinalados nas produções escritas por aprendizes que são falantes nativos de espanhol, italiano e alemão, verificou-se que a transferência linguística de valores de género é mais proeminente nos níveis A1-A2, do que nos níveis subsequentes, B1-B2 e C1.

É preciso, contudo, estabelecer uma distinção entre os dados apurados nos diferentes segmentos da amostra. Mais especificamente, verificou-se, entre os informantes de LM espanhola e italiana, que o peso da transferência linguística dos valores de género é mais significativo nos desvios que afetam nomes cujo valor de género não depende de critérios semânticos nem formais, i.e., na categoria 4. Com efeito, é assinalável o número de desvios sobre os nomes femininos *viagem* e *dor*, cujos correspondentes nas respetivas LM são de género masculino. Parece, então, que perante a ambiguidade do *input* da LA relativamente à associação de valores de género, estes aprendizes tendem a associar o valor ao nome em função do valor do item lexical correspondente na sua LM, um processo que caracteriza o fenómeno da transferência de superfície (Sabourin *et al.* 2006):

- (1) «\***O dor** que está nesta história não se va ver e aparece que todas os personagens sejam cansados mais por dover que por verdadeiro sentido.» (Italiano.B1-B2);
- (2) «\*Muitas coisa aconteceram **neste viagem** que sempre vou lembrar.» (Espanhol.A1-A2);
- (3) «\*Istos anos foram os melhores da minha vida e agora é **o maior dor** tambem.» (Espanhol.C1).

Já entre os informantes que são falantes nativos de alemão, uma vez que este idioma exibe um sistema tripartido de valores de género, i.e., não totalmente coincidente com o número de valores disponíveis em português, não há, por conseguinte, um peso significativo da transferência direta de valores de género. Em contrapartida, poder-se-á inferir, a partir dos dados, um efeito indireto de transferência linguística. O facto de a LM destes aprendentes apresentar a categoria de género gramatical contribui favoravelmente para o seu desenvolvimento linguístico, já que é notório, neste segmento da amostra, uma progressão significativa da aprendizagem do sistema de atribuição de género nominal do português, visível pela diminuição acentuada de desvios dos níveis A1-A2 para B1-B2, e destes para o C1. Este desempenho poderá, então, estar correlacionado com a transferência profunda (Sabourin *et al.* 2006) da categoria de género gramatical. Ou seja, a assimilação da categoria de género gramatical, por parte dos aprendentes de LM alemã, poderá ser favorecida pela possibilidade de transferirem, da sua LM e pelo menos parcialmente, propriedades subjacentes ao funcionamento da atribuição de valores de género aos nomes e da concordância nominal em género.

**P3:** Qual o grau de sensibilidade dos diferentes grupos de aprendentes relativamente aos indícios semânticos e formais de atribuição de valores de género aos nomes em português como LNM?

Da observação dos desvios apurados por categorias elaboradas para a presente investigação, constatou-se que, ao longo do processo de assimilação dos valores de género gramatical aos nomes em português, o papel dos indícios semânticos é mais significativo do que o dos indícios formais, em todos os segmentos da amostra seleccionada.

Entre os desvios assinalados sobre nomes cujo valor de género se infere a partir de indícios formais, mas não de indícios semânticos (categoria 3), verificou-se que os aprendentes que são falantes nativos de línguas românicas apresentam índices de desvio mais baixos nos diferentes níveis QECRL relativamente aos apurados nos restantes segmentos. Assim sendo, poder-se-á considerar que os informantes de LM espanhola e italiana reconhecem, mais facilmente, os índices temáticos do nome *-o* e *-a*, bem como a sua relação parcial com os valores de género disponíveis no português. Provavelmente, esta situação decorre da forte semelhança que, a este respeito, se regista nos sistemas de atribuição de género gramatical das suas LM.



O facto de se associarem as terminações *-o* e *-a* aos valores de género masculino e feminino, respetivamente, poderão igualmente justificar o número de desvios consideravelmente alto que incidem sobre formas nominais atemáticas terminadas em *-ão*. Com efeito, e como se viu na apresentação dos resultados, nas produções escritas por aprendentes de LM italiana, uma parte substancial dos desvios sobre nomes atemáticos incide nos nomes femininos que são deverbais derivados com *-ção*, sendo que, curiosamente, as formas correspondentes no italiano são igualmente femininas. Por conseguinte, parece haver indícios de que os informantes assumem estas formas como tendo o índice temático *-o* que tipicamente está associado ao valor de género masculino. Portanto, a concetualização dos aprendentes relativamente à estrutura formal dos nomes e à correlação entre as propriedades formais e os valores de género nominal disponíveis no português conduzem, assim, ao desvio.

Para além disso, registou-se, na categoria 3, e na maior parte dos segmentos da amostra, um considerável número de desvios nos nomes sobrecomuns *criança* e *pessoa*. Ora, a associação do masculino a estes sobrecomuns poderá, por sua vez, estar correlacionada com o facto de os aprendentes assumirem que se trata de nomes comuns de dois, como *turista* e *jornalista* e, por conseguinte, atribuem-lhes o valor de género tipicamente associado à referência genérica da espécie, i.e., o masculino. Assim sendo, e mais uma vez, estes dados parecem indiciar que, na altura de atribuir um valor de género ao nome e de estabelecer a concordância nominal em género, os critérios semânticos predominam sobre os formais.

**P4:** Haverá, nos diferentes grupos de aprendentes, uma maior incidência de comportamentos desviantes de atribuição de valores de género aos nomes ou de concordância nominal em género nos especificadores (determinantes e quantificadores) ou nos modificadores (adjetivos) com função atributiva?

Em termos globais, os dados apurados neste estudo são reveladores de uma maior incidência de desvios de concordância nominal em género nos especificadores, em particular sobre os determinantes. A maior incidência de desvios nesta classe de palavras poderá estar relacionada com o facto de se tratar de itens funcionais que, do ponto de vista comunicativo, não detêm um papel crucial. Ora, nas teorias sobre o processamento de *input* de uma LNM e da sua transformação em *intake* (VanPatten 2008; Pienemann 2008a e 2008b), fica subjacente a ideia de que no processamento da informação linguística, o foco

do aprendente incidirá, primeiramente, na função do item linguístico e só depois na sua forma, sendo que o valor comunicativo do item é essencial para que seja processado e integrado na gramática da interlíngua do aprendente. Decorrente disso, os nomes são a classe de palavras menos afetada por desvios, bem como a classe dos adjetivos.

É de salientar ainda que são frequentes, sobretudo nas produções dos informantes a frequentar turmas dos níveis QECRL mais avançados, as situações de ambivalência de valores de género. Ou seja, detetaram-se no *corpus* selecionado, um conjunto de casos em que, diante de um nome, o aprendente marca indiferenciadamente como masculinas e femininas as formas dos elementos constituintes do SN:

- (4) «\*<Mas> No entanto, para ir em Inglaterra, o que devo fazer bastante frequentemente, existe só o Sud-Express, o comboio do norte que termina na **fronteira franco-espanhol**.» (Inglês.C1)
- (5) «\*A Suíça é um **país montanhosa** pero não em todas partes.» (Inglês.B1-B2)
- (6) «\*Este foi uma **viagem agitado** mas que eu não esquecerei.» (Espanhol.A1-A2)
- (7) «\* **Todos essas circunstâncias** eram a razão ~~(porque)~~ por que a Maria esteve muito contente no dia da partida dos pais.» (Alemão.C1)

Portanto, os casos que resultam de uma estratégia de “seleção mista” do valor de género por parte dos informantes indiciam que os aprendentes não dominam, na sua plenitude, o sistema de concordância nominal em género. Consequentemente, torna-se necessário, numa investigação futura, proceder a uma análise detalhada destes dados, para averiguar que fatores contribuem para a seleção indiferenciada dos valores de género gramatical por aprendentes tardios de PLNM.



Por fim, o facto de, tendencialmente, nos níveis avançados, o número de desvios registado por item ser igual ao número de ocorrências do nome afetado sugere que, ao longo do processo de aquisição/aprendizagem lexical, o valor de género não é uma das primeiras propriedades do nome a serem assimiladas pelos aprendentes tardios, a não ser que haja um correferencial semântico. Por conseguinte, os resultados apurados neste estudo permitem-nos concluir que, no momento de aquisição dos valores de género dos nomes, atuam, primeiramente, os indícios semânticos.

Nos casos dos nomes cujo valor de género não é atribuível com base em critérios semânticos, é necessário o domínio do mecanismo da concordância sintática para que este se manifeste. O desenvolvimento da concordância nominal em género revela-se, por seu turno, gradual e progressivo, melhorando ao longo do processo de aquisição/aprendizagem do PLNM, sendo certo, no entanto, que os desvios de CNG persistem mesmo nas produções dos aprendentes a frequentar turmas do nível C1.

Assinale-se, por fim, que a presente investigação pretende, sobretudo, contribuir para um maior conhecimento sobre a aquisição/aprendizagem dos valores de género nominal por aprendentes tardios do PLNM. Para além disso, e ao terem sido assinalados conjuntos de itens nominais que parecem ser mais problemáticos em diferentes grupos de informantes, distinguidos por LM e em diferentes estágios de aprendizagem, este estudo deverá também servir de referência para o ensino do português como LNM.

Em conclusão, reconhecemos que a categoria gramatical de género apresenta, indubitavelmente, um estatuto particular no contexto da aquisição/aprendizagem de uma LNM. Neste âmbito, há, de facto, um longo caminho a percorrer na investigação para se compreender, na sua plenitude, as especificidades envolvidas não só no processo de assimilação dos valores de género dos nomes, mas também relativamente aos processos subjacentes à instrumentalização, por parte dos aprendentes tardios, dos mecanismos de concordância sintática nominal em género.







# Bibliografia

---

- Acosta, J. L. & Leiria, I. (1997).** O papel dos conhecimentos linguísticos prévios na aquisição de uma língua não-materna. *Polifonia*, (1), 57-80.
- Acquaviva, P. (2009).** The Structure of the Italian Declension System. In F. Montermini, G. Boyé & J. Tseng (Eds.), *Selected Proceedings of the 6th Décebrettes* (pp. 50-62), Somerville, MA: Cascadilla Proceedings Project.
- Aikhenvald, A. Y. (2000).** *Classifiers. A Typology of Noun Categorization Devices*. Oxford Studies in Typology and Linguistic Theory. Oxford: University Press.
- Aikhenvald, A. Y. (2004).** Gender and noun class. In G. Booji, C. Lehmann, J. Mugdan, S. Skopeteas (Eds.), *Morphology: an International Handbook on Inflection and Word-formation* (pp. 1031-1045). Berlin: Walter de Gruyter.
- Alarcón, I. (2004).** The sequential acquisition of L2 Spanish gender marking: Assignment and agreement. *Indiana University Linguistics Club Working Papers Online*, 4(7), 1-23. Disponível em <https://www.indiana.edu/~iulcwp/> (consultado em 06.01.2019).
- Alonso, R. A. (1999).** A relación interlingua-transferencia. *Revista Galega do Ensino*, (25), 137-147.
- Alonso, R. A. (2002).** Current issues in language transfer. In I.M. Duarte; J. Barbosa; S. Matos; T. Hüsgen (Orgs.), *Encontro Comemorativo dos 25 anos do Centro de Linguística da Universidade do Porto*, Volume 2 (pp. 232-236). Porto: Centro de Linguística da Universidade do Porto.
- Ambadiang, T. (1994).** *La morfología flexiva*. Madrid: Taurus.
- Ambadiang, T. (1999).** La flexión nominal. Género y número. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española*, Vol.3, (pp. 4843-4913). Madrid: Real Academia Española (R.A.E.), Espasa.

- Augusto, M. R. A. & Corrêa, L. M. S. (2005).** Marcação de gênero, opcionalidade e genericidade: processamento de concordância de gênero no DP aos dois anos de idade. *Revista Lingüística* 1, 207–234. Disponível em <https://revistas.ufrj.br/index.php/rl/article/view/4373/3145> (consultado em 20.10.2018).
- Baddeley, A. (1999).** Memory. In R.A. Wilson & F.C. Keil (Orgs.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass: MIT Press.
- Baptista, A.; Choupina, C.; Costa, J.A.; Querido, J. & Oliveira, I. (2013).** Conhecimentos implícitos e explícitos de gênero linguístico e suas implicações no ensino. In M. Teixeira, L. Santos, I. Silva & E. Mesquita (Orgs.), *Ensinar e Aprender Português num mundo plural* (pp. 17-51). Escola Superior de Educação e Universidade Federal Uberlândia.
- Barbosa, J. S. (1822).** *Grammatica philosophica da lingua portugueza ou principios de grammatica geral applicados à nossa linguagem*. Lisboa: Academia Real das Sciencias.
- Barcroft, J. & Wong, W. (2013).** Input, input processing and focus on form. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 627-647). Cambridge, New York [New York]: Cambridge University Press.
- Bechara, E. (1999).** *Moderna Gramática Portuguesa*. Rio de Janeiro: Editora Lucerna.
- Bialystok, E. & Hakuta, K. (1999).** Confounded Age: Linguistic and Cognitive Factors in Age Differences for Second Language Acquisition. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 161-181). New Jersey: Lawrence Erlbaum Associate Publishers.
- Bialystok, E. & Miller, B. (1999).** The problem of age in second language acquisition: influences from language, structure, and task. *Bilingualism: Language and Cognition*, 2 (2), 127-145.
- Birdsong, D. (1999).** Introduction: Whys and Why Nots of the Critical Period Hypothesis for Second Language Acquisition. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 1-22). New Jersey: Lawrence Erlbaum Associate Publishers.
- Birdsong, D., & Molis, M. (2001).** On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language*, (44), 235-249.



- Bley-Vroman, R. (1990).** The logical problem of Foreign Language Learning. *Linguistic Analysis*, 20 (1-2),3-49.
- Brito, A. M. (2003).** Categorias Sintáticas. In Mateus, Maria H.M. *et al.* (2003) *Gramática da Língua Portuguesa* (5ª ed., revista e aumentada) (pp. 323-432). Lisboa: Caminho.
- Brugmann, K. (1897).** *The nature and the origin of noun genders in the indo-european languages.* Tradução do alemão *Das Nominalgeschlecht in den indogermanischen Sprachen* (pp. 1-32). Cambridge, U.S.A.: University Press, John Wilson and Son.
- Bruhn de Garavito, J. & White, L. (2002).** L2 acquisition of Spanish DPs: the status of grammatical features. In A. T. Pérez-Leroux & J. Liceras (Eds.), *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153-178). Dordrecht: Kluwer.
- Bußmann, H. & Hellinger, M. (2003).** Engendering female visibility in German. In M. Hellinger & H. Bußmann (Eds.), *Gender Across Languages*, Vol. 3 (pp. 14-174). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Caffarra, S.; Siyanova-Chanturia, A.; Pesciarelli, F.; Vespignani, F. & Cacciari, C. (2015).** Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, 52,1019-1030.
- Câmara Jr., J. M. (1966).** Considerações sobre o gênero em português. *Revista Brasileira de Linguística Teórica e Aplicada*, I (2), 19, São Paulo: Centro de Linguística Aplicada.
- Câmara Jr., J. M. (1994).** *Estrutura da língua portuguesa*. 22ª edição. Petrópolis: Vozes.
- Carroll, S. E. (2001).** *Input and Evidence. The raw material of Second Language Acquisition.* Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Carstens, V. (2000).** Concord in Minimalist Theory. *Linguistic Inquiry*, 31 (2), 319-355.
- Carvalho, J. G. H. de (1979).** *Teoria da Linguagem*. vol. 2. Coimbra: Atlântida Editora.
- Carvalho, J. G. H. de (2000).** Género. *Enciclopédia Verbo Luso-Brasileira da Cultura*, Vol. 13, 183-186, Lisboa/São Paulo: Verbo.
- Cavele, S. (1999).** *Cancelamento da marca de género feminino no português oral de Maputo.* Tese de licenciatura. Universidade Eduardo Mondlane, Maputo. Disponível em <http://www.saber.ac.mz/bitstream/10857/547/1/Lt-011.pdf> (consultado em 01.11.2018).
- Chao, Y. R. (1968).** *A grammar of Spoken Chinese.* Berkeley and Los Angeles: University of California Press.

- Chomsky, N. (1975 [1995]).** *Aspectos da Teoria da Sintaxe*. Tradução do original inglês *Aspects of the theory of syntax*. Coimbra: Arménio Amado.
- Chomsky, N. (1995 [1999]).** *O Programa Minimalista*. Tradução do original inglês *The Minimalist Program*. Lisboa: Editorial Caminho.
- Choupina, C.; Baptista, M.A.; Costa, J. A. (2014).** A gramática intuitiva, o conhecimento linguístico e o ensino-aprendizagem do género em PE, Trabalho apresentado em IV SIELP - Simpósio Internacional de Ensino da Língua Portuguesa. In *Anais do IV SIELP - Simpósio Internacional de Ensino da Língua Portuguesa*. Uberlândia.
- Choupina, C. M.; Baptista, M. A.; Costa, J. A.; Oliveira, I. & Querido, J. (2016).** Conhecimentos e regras explícitos e implícitos sobre o género linguístico nos alunos dos 1º e 2º ciclos do Ensino Básico. *Revista da Associação Portuguesa de Linguística*, 1 (1), 201-231. Disponível em <http://ojs.letras.up.pt/index.php/APL/article/view/1590/1413> (consultado em 20.10.2018).
- Clark, Eve V. (1998).** Morphology in Language Acquisition. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 374-389). Oxford: Blackwell Publishers.
- Comrie, B. (1999).** Grammatical gender systems: a linguist's assessment. *Journal of Psycholinguistic Research*, 28, (5), 457-466.
- Conselho da Europa (2001).** *Quadro Europeu Comum de Referência para as Línguas. Aprendizagem, ensino, avaliação*. Porto: Edições Asa.
- Contini-Morava, E. & Kilarski, M. (2013).** Functions of Nominal Classification. *Language Sciences*, 40, 263-299.
- Cook, V. J. (1985).** Chomsky's Universal Grammar and Second Language Learning. *Applied Linguistics*, 6 (1), 1-18.
- Cook, V. J. (1994).** Universal Grammar and the learning and teaching of second language. In T. Odlin (Ed.), *Perspectives on Pedagogical Grammar* (pp. 25-48). Cambridge: Cambridge University Press.
- Corbett, G. G. (1991).** *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. G. (1998).** Morphology and Agreement. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 191-205), Oxford: Blackwell Publishers.

- Corbett, G. G. (2003).** Agreement: terms and boundaries. In W.E. Griffin (Ed.), *The Role of Agreement in Natural Language: TLS 5 Proceedings*, Texas Linguistic Forum, 53, 109-122.
- Corbett, G. G. (2005).** Systems of nominal classification I: Gender oppositions. In D. A. Cruse, F. Hundsnurscher, M. Job & P. R. Lutzeier (Eds.), *Lexicology: An International Handbook. on the Nature and Structure of Words and Vocabularies II* (pp. 986-994). Berlin: Mouton de Gruyter,.
- Corbett, G. G. (2006a).** *Agreement*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2006b).** Grammatical Gender. In J. Holmes & M. Meyerhoff (Eds.), *The Handbook of Language and Gender* (pp. 749-756). Oxford: Blackwell. Disponível em <http://www.surrey.ac.uk/LIS/SMG/Gender%20grammatical.pdf> (consultado em 02.03.2018).
- Corbett, G. G. (2013a).** Number of Genders. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Disponível em <http://wals.info/chapter/30> (consultado em 02.03.2018).
- Corbett, G. G. (2013b).** Sex-based and Non-sex-based Gender Systems. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Disponível em <http://wals.info/chapter/31> (consultado em 02.03.2018).
- Corbett, G. G. (2013c).** Systems of Gender Assignment. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Disponível em <http://wals.info/chapter/32> (consultado em 02.03.2018).
- Corder, S. P. (1967 [1992]).** La importancia de los errores del que aprende una lengua segunda. In J.M. Liceras, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 31-40), Madrid: Visor.
- Corder, S. P. (1971 [1992]).** Dialectos idiosincrásicos y análisis de errores. In J. M. Liceras, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 63-77), Madrid: Visor.

- Corder, S.P. (1993).** A role for the mother tongue. In S. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning (revised edition)* (pp. 18-31). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Corrêa, L.M.S. (2002).** Explorando a relação entre língua e cognição na interface: o conceito de interpretabilidade e suas aplicações para teorias do processamento e da aquisição de linguagem. *Veredas*, 6 (1), 113-129.
- Corrêa, L.M.S.; Name, M. C. L. & Ferrari-Neto, J. (2004).** O processamento de informação de interface na aquisição de gênero e de número no Português Brasileiro. *Letras de Hoje*, 39 (3), 123-137. Disponível em <http://revistaseletronicas.pucrs.br/fzva/ojs/index.php/fale/article/view/13909/9223> (consultado em 25.11.2018).
- Corrêa, L.M.S. & Augusto, M. R. A. (2017).** Primeiros passos na aquisição da sintaxe: o sintagma nominal. In M.J. Freitas & A. L. Santos (Eds.), *Aquisição de língua materna e não materna: questões gerais e dados do português* (pp. 121-154), Berlin: Language Science Press.
- Corrêa, L. M. S.; Name, M. C. L. & Ferrari-Neto, J. (2004).** O processamento de informação de interface na aquisição de gênero e de número no Português Brasileiro. *Letras de Hoje*, 39 (3), 123-137. Disponível em <http://revistaseletronicas.pucrs.br/fzva/ojs/index.php/fale/article/view/13909/9223> (consultado em 25.11.2018).
- Corrêa, L. M. S. & Augusto, M. R. A. (2017).** Primeiros passos na aquisição da sintaxe: o sintagma nominal. In M. J. Freitas & A. L. Santos (Eds.), *Aquisição de língua materna e não materna: questões gerais e dados do português* (pp. 121-154). Berlin: Language Science Press.
- Costa, J. A.; Choupina, C. M.; Baptista, A.; Oliveira, I. & Querido, J. (2015).** Gênero gramatical: a complexidade do conteúdo e a sua abordagem nos documentos reguladores do ensino do português no 1º ciclo EB. *Exedra*, 322-352.
- Croft, W. (1990).** *Typology and universals*. Cambridge: Cambridge University Press.
- Crossley, S.; Salsbury, T.; Titak, A. & McNamara, D. (2014).** Frequency effects and second language acquisition. Word types, word tokens, and word production. *International Journal of Corpus Linguistics*, 19 (3), 301-332.

- Cruse, A. (2000).** *Meaning in Language. An introduction to Semantics and Pragmatics.* Oxford: University Press.
- Cunha, C. & Cintra, L. F. (2005).** *Nova Gramática do Português Contemporâneo.* 18ª edição. Lisboa: Editora Sá da Costa.
- Dahl, O. (2000).** Animacy and the notion of semantic gender. In B. Unterbeck (Ed.), *Gender in Grammar and Cognition I: Approaches to Gender* (pp. 99-116). Berlin / New York: Mouton De Gruyter.
- Dardano, M. & Trifone, P. (1997).** *La Nuova Grammatica della Lingua Italiana.* Bologna: Zanichelli.
- DeAngelis, G. (2005).** Interlanguage Transfer of Function Words. *Language Learning*, 55 (3), 379-414.
- Dewaele, J. & Veronique, D. (2000).** Relating gender errors to morphosyntax and lexicon in advanced French interlanguage. *Studia Linguistica*, 54 (2), 212-224.
- Dewaele, J. & Veronique, D. (2001).** Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Bilingualism: Language and Cognition*, 4 (3), 275-297.
- Dicionário de alemão-português, português-alemão (1ª ed., 2ª reimp.). (2016). Porto: Porto Editora.
- Dicionário de espanhol-português, português-espanhol (1ª ed., 5ª reimp.). (2016). Porto: Porto Editora.
- Dicionário de italiano-português, português-italiano (2ª ed., 3ª reimp.). (2016). Porto: Porto Editora.
- Dixon, R. M. W. (1986).** Noun Classes and Noun Classification: in typological perspective. In C. Craig (Ed.), *Noun Classes and categorization. Proceedings of a Symposium on Categorization and Noun Classification* (pp. 105-112). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Donaldson, B. (2007).** *German. An Essential Grammar.* London: Routledge.
- Dressler, W. U. & Thornton, A. (1996).** Italian Nominal Inflection. *Wiener Linguistische Gazette*, 55-57, 1-24. Disponível em: [http://www.annathornton.net/joomla/images/Dressler\\_Thornton\\_1996.pdf](http://www.annathornton.net/joomla/images/Dressler_Thornton_1996.pdf) (consultado em 02.03.2018).

- Dubert, F. P. (2010).** Sobre o lugar do xénero na gramática do galego. *Estudos de Lingüística Galega*, 2, 57-74. Disponível em: <http://www.redalyc.org/articulo.oa?id=305626377011> (consultado em 02.03.2018).
- Ellis, N. (2002).** Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24 (2), 143-188.
- Ellis, R. (1986).** *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2000).** *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2003).** *Second language acquisition*. 9th Edition. Oxford: Oxford University Press.
- Ernaut, A. (1914).** *Morphologie historique du Latin*. Paris: Klincksieck. Disponível em <https://archive.org/details/morphologiehisto00erno> (consultado em 02.03.2018).
- Etnner, C. (2002).** Is Chinese, men and women are equal -or- women and men are equal? In M. Hellinger & H. Bußmann (Eds.), *Gender Across Languages*, Vol. 2, (pp. 29-41). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Eubank, L. & Gregg, K. R. (1999).** Critical Periods and (Second) Language Acquisition: Divide et Impera. In D. Birdsong (Ed.), *Second Language Acquisition and the Critical Period Hypothesis* (pp. 65-100). New Jersey: Lawrence Erlbaum Associate Publishers.
- Ferreira, T. S. (2011).** *Padrões na aquisição/aprendizagem da marcação de género nominal em português como L2*. Dissertação de mestrado em Português Língua Estrangeira/ Língua Segunda, apresentada à Faculdade de Letras da Universidade de Coimbra.
- Figueira, R. A. (2001).** Marcas insólitas na aquisição do gênero gramatical. *Letras de Hoje*, 36 (3), 313-320.
- Figueiredo, C. F. G. (2009).** A Configuração do SN Plural do Português Reestruturado de Almojarife – S. Tomé. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1 (1), 28-55.
- Flores, C. M. M. (2013).** Português Língua Não Materna. Discutindo conceitos de uma perspetiva linguística. In: R. Bizarro, M. Moreira & C. Flores (orgs.) *Português língua não materna: investigação e ensino* (pp. 35-46). Lisboa: Lidel.

- Flynn, S. (1996).** A Parameter-Setting Approach to Second Language Acquisition. In W.C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 121-158), San Diego, California: Academic Press.
- Foucart, A. (2008).** *Grammatical gender processing in French as a first and second language*. Université Aix-Marseille I / University of Edinburgh. Tese de Doutorado.
- Foucart, A. & Frenck-Mestre, C. (2011).** Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, 14 (3), 379-399.
- Foundalis, H. E. (2002).** Evolution of Gender in Indo-European Languages. *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society*. Fairfax: Virginia.
- Franceschina, F. (2003).** Parameterized Functional Features and SLA. In J. M. Liceras *et al.* (Eds.), *Proceedings of the 6<sup>th</sup> Generative Approaches to Second Language Acquisition, Somerville* (pp. 87-105), MA: Cascadilla Proceedings Project.
- Franceschina, F. (2005).** *Fossilized Second Language Grammars – the Acquisition of Grammatical Gender*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gass, S. (1979).** Language Transfer and Universal Grammar Relations. *Language Learning*, 29 (2) 327-344.
- Gass, S. (1996).** Second Language Acquisition and Linguistic Theory: the role of Language Transfer. In W.C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 317-345), San Diego, California: Academic Press.
- Gass, S. & Selinker, L. (1993).** Introduction. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning (revised edition)* (pp. 1-17). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Gass, S. & Selinker, L. (2008).** *Second Language Acquisition: an introductory course* (3<sup>th</sup> edition). New York: Taylor and Francis Group Routledge.
- Gil, D. (2013).** Numeral Classifiers. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Disponível em <http://wals.info/chapter/55> (consultado em 02.03.2018).
- Godinho, A. P. (2010).** A aquisição da concordância de número e a sua relação com a aquisição da concordância de gênero: um estudo realizado com aprendentes chineses de português L2. In: M. J. Marçalo, M. Lima-Hernandes, E. Esteves, M.



- C. Fonseca, O. Gonçalves, A.L., Vilela, A. A. Silva (Eds.), *Língua portuguesa: ultrapassar fronteiras juntar culturas*. Universidade de Évora.
- Gonçalves, P. (1997).** Tipologia de “erros” do português oral de Maputo: um primeiro diagnóstico. In C. Stroud & P. Gonçalves (Orgs), *Panorama do Português Oral de Maputo - Vol. II: A construção de um banco de “erros”* (pp. 37-67). Maputo: Instituto Nacional do Desenvolvimento da Educação.
- Gonçalves, P. (2015).** Aspetos Morfossintáticos da gramática do português de moçambique: a concordância nominal e verbal. *Cuadernos de la ALFAL*, 7, 9-16.
- Gouveia, M. C. F. (2004).** Considerações sobre a categoria gramatical de género: sua evolução do latim ao português arcaico. *BIBLOS*, II, 443-475.
- Greenberg, J. H. (1978).** How does a language acquire gender markers? In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language Vol. 4: Syntax* (pp. 49-81). Stanford, California: Stanford University Press.
- Grinevald, C. (2000).** A morphosyntactic typology of classifiers. In G. Senft (Ed.), *Systems of Nominal Classification* (pp. 50-92). Cambridge: Cambridge University Press.
- Grinevald, C. (2004).** Classifiers. In G. Booij, C. Lehmann, J. Mugdan & S. Skopeteas (Eds.) *Morphology: an international handbook on inflection and word-formation* (pp. 1016-1031). Berlin: Walter de Gruyter.
- Håkansson, G. (2001).** Against Full Transfer – evidence from Swedish learners of German. *Lund University, Dept. of Linguistics Working Papers*, 48, 67-86. Disponível em [http://portal.research.lu.se/portal/en/publications/against-full-transfer--evidence-from-swedish-learners-of-german\(161071e6-7469-49e0-aac5-08576a8adbe5\).html](http://portal.research.lu.se/portal/en/publications/against-full-transfer--evidence-from-swedish-learners-of-german(161071e6-7469-49e0-aac5-08576a8adbe5).html) (consultado em 10.03.2018).
- Hamers, J.F. & Blanc, M.H.A. (1990).** *Bilinguality and Bilingualism*. Cambridge: Cambridge University Press.
- Han, Z. (2004).** *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters Lda.
- Harris, J. W. (1991).** The Exponence of Gender in Spanish. *Linguistic Inquiry*, 22 (1), 27-62. Disponível em: [http://linguistics.fas.harvard.edu/files/linguistics/files/james\\_harris\\_on\\_sp\\_gender.pdf?m=1394463289](http://linguistics.fas.harvard.edu/files/linguistics/files/james_harris_on_sp_gender.pdf?m=1394463289) (consultado em 02.03.2016).



- Harris, J. W. (1992).** The Form Classes of Spanish Substantives. In G. Booji & J. Marle (Eds.), *Yearbook of Morphology* (pp. 65-88). Kluwer Academic Publisher.
- Haugen, E. (1950).** The Analysis of Linguistic Borrowing. *Language*, 26 (2), 210-231.
- Hawkins, R. & Chan, C. (1997).** The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research*, 13 (3), 187-226.
- Hellinger, M. (2001).** English - Gender in a global language. In M. Hellinger & H. Bußmann (Eds.), *Gender Across Languages*, Vol. 1 (pp. 105-113). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Herschensohn, J. (2013)** Age related effects. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 317-337). Cambridge, New York [New York]: Cambridge University Press.
- Hockett, C. F. (1958).** *A Course in modern linguistics*. New York: The Macmillan Company.
- Ibrahim, M. H. (1973).** *Grammatical Gender*. Paris: Mouton.
- Ilari, R. (2013).** O Português no contexto das línguas românicas. In E. B. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura & A. Mendes (Eds.), *Gramática do Português*, Vol. I, (pp. 49-66). Lisboa: Fundação Calouste Gulbenkian.
- Inverno, L. (2009).** *Contact-induced restructuring of Portuguese morphosyntax in interior Angola. Evidence from Dundo (Lunda Norte)*. Tese de Doutoramento em Linguística Portuguesa. Universidade de Coimbra.
- Ioup, G. (2005).** Age in Second Language Development. In E. Hinker (Ed.), *Handbook of Research in Second Language Acquisition Teaching and Learning* (pp. 419-453). New York: Taylor and Francis Group, Routledge.
- Isabelli-Garcia, C. (2010).** Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals*, 43 (2), 289-303.
- Jarvis, S. (2013).** Conceptual Transfer. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 115-117). London & New York: Routledge.
- Johnson, J. S. & Newport, E. L. (1989).** Critical Period Effects in Second Language Learning: The influence of maturational state on the acquisition of English as second language. *Cognitive Psychology*, 21, 60-99.

- Kasper, G. & Blum-Kulka, S. (1993).** Interlanguage Pragmatics: an Introduction. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage Pragmatics* (pp. 3-13). New York: Oxford University Press.
- Kellerman, E. (1977).** Towards a Characterisation of the Strategy of Transfer in Second Language Acquisition. *Interlanguage Studies Bulletin*, 2 (1), 58-145.
- Kellerman, E. (1995).** Crosslinguistic influence: transfer to nowhere?. *Annual Review Applied Linguistics*, Cambridge University Press, 125-150.
- Kibort, A. & Corbett, G. G. (2008).** Gender. *Grammatical Features*, 7. Disponível em <http://www.grammaticalfeatures.net/features/gender.html> (consultado em 15.03.2016).
- Kilarski, M. (2007).** On grammatical gender as an arbitrary and redundant category. In D. A. Kibbee (Ed.), *History of Linguistics 2005: Selected Papers from the Tenth International Conference on the History of Language Sciences (ICHOLS X)* (pp. 24-36). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Kilarski, M. (2013).** *Nominal Classification: A history of its study from the classical period to present*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Köpcke, K. & Zubin, D. A. (1996).** Prinzipien für die Genuszuweisung im Deutschen. In E. Lang & G. Zifonun (Eds.), *Deutsch typologisch. Jahrbuch des Instituts für Deutsche Sprache 1995* (pp. 473–491). Berlin: de Gruyter, Disponível em: [https://www.uni-muenster.de/imperia/md/content/germanistik/lehrende/koepcke\\_k/1996\\_k\\_pcke\\_zubin\\_prinzipien\\_genuszuweisung.pdf](https://www.uni-muenster.de/imperia/md/content/germanistik/lehrende/koepcke_k/1996_k_pcke_zubin_prinzipien_genuszuweisung.pdf) (consultado em 02.03.2017).
- Krashen, S. D. (1981).** *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press. Edição online de 2002, disponível em [http://www.sdkrashen.com/content/books/sl\\_acquisition\\_and\\_learning.pdf](http://www.sdkrashen.com/content/books/sl_acquisition_and_learning.pdf) (consultado em 02.03.2015).
- Krashen, S. D. (1982).** *Principles and Practice in Second Language Acquisition*. Edição online de 2002, disponível em [http://www.sdkrashen.com/content/books/principles\\_and\\_practice.pdf](http://www.sdkrashen.com/content/books/principles_and_practice.pdf) (consultado em 02.03.2015).
- Kurzová, H. (1993).** *From Indo-European to Latin. The evolution of a morphosyntactic type*. Amsterdam, Philadelphia: John Benjamin Publishing Company.

- Lacsán, V. (2015).** *The acquisition of gender agreement in L2 Portuguese by adult Hungarian speakers.* Dissertação de Mestrado apresentada à Faculdade de Letras da Universidade de Lisboa.
- Lado, R. (1957).** *Linguistic across cultures: applied linguistics for Language teachers.* Ann Harbor: University of Michigan Press.
- Larsen-Freeman, D. & Long, M. H. (1991 [1994]).** *Introducción al estudio de la adquisición de segundas lenguas.* Tradução do original inglês *An Introduction to Second Language Acquisition Research*, Madrid: Gredos.
- Lausberg, H. (1974).** *Linguística Românica.* Lisboa: Fundação Calouste Gulbenkian.
- Leiria, I. (2004).** Português Língua Segunda e Língua Estrangeira: Investigação e Ensino. *Idiomático, Revista Digital de Didáctica de PLNLM*, 3, Centro Virtual de Camões, 1-11. Disponível em <http://cvc.instituto-camoes.pt/idiomatico/03/portuguesLSeLE.pdf> (consultado em 10.01.2018).
- Leiria, I. (2006).** *Léxico, Aquisição e Ensino do Português Europeu língua não materna.* Lisboa: Fundação Calouste Gulbenkian, Fundação para a Ciência e Tecnologia.
- Lenneberg, E. H. (1967 [1985]).** *Fundamentos biológicos del lenguaje.* Madrid: Alianza Editorial. Tradução do original inglês *The Foundations of Language*. [s.1.]: John Wiley and Sons.
- Long, M. H. (2003).** Stabilization and Fossilization in Interlanguage Development. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 487-535), Malden: Blackwell.
- Lopes, A. C. M; Martins, C. (2017).** Para o ensino dos usos dos artigos em PLE: por onde começar? *Portuguese Language Journal* (pp. 165-189), 11.
- Lopes, R. E. V. (2006).** Traços Semânticos na Aquisição da Linguagem. *Letras de Hoje*, 41 (1), 161-178.
- Li, C. N. & Thompson, S. A. (1981).** *Mandarin Chinese. A Functional Reference Grammar.* London, England: University of California Press.
- Li, X. (2013).** *Numeral classifiers in Chinese. The syntax-semantics interface.* Berlin: Walter de Gruyter.
- Lucchesi, D. (2009).** A concordância de gênero. In D. Lucchesi, A. Baxter & I. Ribeiro (Eds.), *O português afro-brasileiro*, (pp. 295-318). Salvador: EDUFBA. Disponível em <http://books.scielo.org> (consultado em 12.11.2016).

- Marcato, G. & Thüne, E. (2002).** Gender and female visibility in Italian. In M. Hellinger & H. Bußmann (Eds.), *Gender Across Languages*, Vol. 2, (pp. 187-218). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Madeira, A. (2017).** Aquisição de língua não materna. In M.J. Freitas & A. L. Santos (Eds.), *Aquisição de língua materna e não materna: questões gerais e dados do português* (pp. 306-330), Berlin: Language Science Press.
- Mariotto, E. M. C. (2014).** *Processamento da concordância de género por aprendentes de português como língua estrangeira: evidências de um estudo de leitura automonitorada*. Dissertação de Mestrado em Língua e Cultura Portuguesas (Português Língua Estrangeira / Língua Segunda) apresentada à Faculdade de Letras da Universidade de Lisboa.
- Mariotto, E. M. C. & Lourenço-Gomes, M. C. (2013).** Análise de erros na escrita relacionados à aprendizagem da concordância de género por falantes nativos do inglês, aprendentes de português europeu como língua estrangeira. *Anais do IV Simpósio Mundial de Estudos de Língua Portuguesa (SIMELP). Língua Portuguesa: ultrapassando fronteiras, unindo culturas* (pp. 1278-1285). Faculdade de Letras/UFG, Goiânia, Goiás, Brasil. Disponível em [http://www.simelp.letras.ufg.br/anais/simposio\\_26.pdf2013](http://www.simelp.letras.ufg.br/anais/simposio_26.pdf2013) (consultado em 01.11.2017).
- Martins, C. S. P. (1997).** Bilinguismo e manifestações verbais bilingues. Uma breve sinopse teórica. *Revista Portuguesa de Filologia*, XXI. Coimbra: Faculdade de Letras da Universidade de Coimbra, 63-125.
- Martins, C. S. P. (2008).** *Línguas em contacto; “saber sobre” o que as distingue. Análise de competências metalinguísticas de crianças mirandesas em idade escolar*. Coimbra: Imprensa da Universidade de Coimbra.
- Martins, C. S. P. (2013).** O Corpus de Produções Escritas de Aprendentes de PL2 (PEAPL2/CELGA). Caracterização e desenvolvimento de uma infra-estrutura de investigação. In R. Bizarro, M. Moreira & C. Flores (orgs.), *Português Língua Não Materna: Investigação e Ensino* (pp. 69-80). Lisboa: Lidel.
- Martins, C. S. P. (2015).** Número e género nominais no desenvolvimento das interlínguas de aprendentes de português europeu como língua estrangeira. *Revista Científica da Universidade Eduardo Mondlane, Série Letras e Ciências Sociais*, 1 (1), 26-51.
- Martins, C. S. P. (2016).** O papel diferenciado de subsistemas de memória de longo prazo nos processos de aquisição e de aprendizagem de uma L2. In J. Corrêa-Cardoso

& M. C. Fialho (Coord.), *A Linguagem na Pólis* (pp. 99-120). Coimbra, Imprensa da Universidade de Coimbra.

- McLaughlin, B. (1978 [1992]).** Algunas consideraciones metodológicas sobre el modelo del monitor. In J. M. Liceras, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 153-176), Madrid: Visor.
- McLaughlin, B. (1984).** *Second-Language Acquisition in childhood*. New Jersey: New Jersey: Lawrence Erlbaum Associate Publishers.
- Meillet, A. (1921).** *Linguistique Historique et Linguistic Générale*. Librairie Ancienne Honoré Champion: Paris.
- Mendes, A. (2013).** Processos de Gramaticalização. In E. B. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura & A. Mendes (Eds.), *Gramática do Português*, Vol. I, (pp. 249-292). Lisboa: Fundação Calouste Gulbenkian.
- Moravcsik, E. A. (1978).** Agreement. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language Vol. 4: Syntax* (pp. 333-370). Stanford, California: Stanford University Press.
- Morgan-Short, K.; Sanz, C.; Steinhauer, K. & Ullman, M. T. (2010).** Second Language Acquisition of Gender Agreement in Explicit and Implicit training conditions: an event-related potencial study. *Language Learning*, 60 (1), 154-193.
- Mota, M. A. (2016a).** Morfologia nas interfaces. In Martins, A. M. e Carrilho, E. (Eds.), *Manual de Linguística Portuguesa* (pp. 156-177). Berlin / Boston: De Gruyter
- Mota, M. A. (2016b).** A categoria gramatical género, nos nomes e adjetivos do Português: algumas reflexões. *Diadorim, Especial* (pp. 150-164). Rio de Janeiro. Disponível em <https://revistas.ufrj.br/index.php/diadorim/article/view/4053/3031> (consultado em 12.05.2018)
- Murphy, S. (2003).** Second language transfer during third language acquisition. *Working Papers in TESOL & Applied Linguistics*, 3 (2), 1-21.
- Nemser, W. (1971 [1992]).** Los sistemas aproximados de los que aprender lenguas segundas. In J.M. Liceras, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 51-61), Madrid: Visor.
- Nichols, J. (1999).** *Linguistic Diversity in Space and Time*. Chicago, IL: University of Chicago Press.

- Nissen, U. K. (2002).** Gender in Spanish. In M. Hellinger & H. Bußmann (Eds.), *Gender Across Languages*, Vol. 2, (pp. 251-280). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Norman, J. (1988).** *Chinese*. Cambridge: Cambridge University Press.
- Nunes, J. J. (1975).** *Compêndio de gramática histórica portuguesa: (fonética-morfologia)*. 8ª edição. Lisboa: Livraria Clássica Editora de A. M. Teixeira
- Odlin, T. (1989).** *Language Transfer: Cross Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Odlin, T. (2013).** Cross-linguistic Influence (CLI). In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 151-155). London & New York: Routledge.
- Oliphant, K. (1997).** Acquisition of grammatical gender in Italian as a foreign language. *Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center*, 1-33. Disponível em: <http://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/8947/NW07.pdf?sequence=1> (consultado em 02.03.2014).
- Paradis, M. (2004).** *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamin Publishing Company.
- Pfau, R. (2009).** *Grammar as Processor. A Distributed Morphology account of spontaneous speech errors*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Pienemann, M. (2008a).** Processability Theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition – an Introduction* (pp. 137-154). New York/London: Routledge & Francis Group.
- Pienemann, M. (2008b).** A Brief Introduction to Processability Theory. In J. Keßler (Ed.), *Processability Approaches to Second Language Development and Second Language Learning* (pp. 9-30). Cambridge Scholars Publishing, 9-30.
- Pinto, J. (2012).** Transferências lexicais na aquisição de português como língua terceira ou língua adicional. Um estudo com alunos universitários em Marrocos. *Diacrítica*, 26 (1), 171-187.
- Pinto, J. (2015).** A aquisição do género e da concordância de género em português língua terceira ou língua adicional. In P. Osório & M.J. Grosso (Eds.), *Teorias e Usos Linguísticos – Aplicações ao Português Língua Não Materna* (pp. 91-110). Lisboa: Lidel.

- Português Fundamental (1984).** *Vocabulário e Gramática*, tomo I, *Vocabulário*. Lisboa: Instituto Nacional de Investigação Científica, Centro de Linguística da Universidade de Lisboa.
- Quirk, R.; Greenbaum, S.; Leech, G. & Svartvik, J. (1985).** *A comprehensive grammar of the English Language*. London/New York: Longman.
- RAE (Real Academia Española) (2001).** *Diccionario de la lengua española* (22<sup>a</sup> ed.), Volumes 1 e 2. Madrid: Real Academia Española.
- RAE (Real Academia Española) (2009).** *Nueva Gramática de la Lengua Española. morfología y Sintaxis*. Madrid: Espasa.
- Ramsey, S. R. (1987).** *The Languages of China*. New Jersey: Princeton University Press.
- Raposo, E. B. P. (2013).** Estrutura da frase. In E. B. P. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura & A. Mendes (Eds.), *Gramática do Português*, Vol. I, (pp. 303-398). Lisboa: Fundação Calouste Gulbenkian.
- Raposo, E. B. P. & Nascimento, M. F. B. (2013).** Nomes Próprios In E. B. P. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura & A. Mendes (Eds.), *Gramática do Português*, Vol. I, (pp. 993-1044). Lisboa: Fundação Calouste Gulbenkian.
- Ringbom, H. (2013).** Linguistic Transfer. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 396-400). London & New York: Routledge.
- Rio-Torto, G.; Rodrigues, A. S.; Pereira, I.; Pereira, R. & Ribeiro, S. (2016).** *Gramática Derivacional do Português*. (2<sup>a</sup> edição). Coimbra: Imprensa da Universidade de Coimbra.
- Rio-Torto, G. & Rodrigues, A. S. (2016).** Formação de nomes. In G. Rio-Torto *et al.*, *Gramática Derivacional do Português*. (2<sup>a</sup> edição) (pp. 135-240). Coimbra: Imprensa da Universidade de Coimbra.
- Ritchie, W. C. & Bhatia, T. K. (1996).** Second Language Acquisition: Introduction, Foundations and Overview. In W.C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 1-46), San Diego, California: Academic Press.
- Ritter, E. (1993).** Where's Gender? *Linguistic Inquiry*, 24 (4), 795-803.
- Rodrigues, A. S. (2016).** Noções basilares sobre a morfologia e o léxico. In Rio-Torto *et al.*, *Gramática Derivacional do Português*. (2<sup>a</sup> edição) (pp. 35-133). Coimbra: Imprensa da Universidade de Coimbra.

- Sabourin, L.; Stowe, L. A. & De Haan, G. D. (2006).** Transfer effects in learning a second language grammatical gender system. *Second Language Research*, SAGE Publications, 2006, 22(1), 1-29.
- Salmons, J. (1993).** The Structure of the Lexicon: Evidence from German gender assignment. *Studies in Language*, 17(2), 411-435.
- Schachter, J. (1974 [1992]).** Un error en el análisis de errores. In J.M. Licerias, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 195-205), Madrid: Visor.
- Schachter, J. (1993).** A new account for language transfer. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning (revised edition)* (pp. 32-46). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Schachter, J. (1996).** Maturation and the issue of Universal Grammar. In W.C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 159-194), San Diego, California: Academic Press.
- Schenke, H. & Seago, K. (2004).** *Basic German: a Grammar and Workbook*, New York: Taylor and Francis Group Routledge.
- Schwartz, B. D. & Sprouse, R. A. (1996).** L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12 (1), 40-72.
- Selinker, L. (1972 [1992]).** La Interlengua. In J.M. Licerias, *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua* (pp. 79-101), Madrid: Visor.
- Selinker, L. & Lakshmanan, U. (1993).** Language Transfer and Fossilization: the multiple effects principle. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning (revised edition)* (pp. 197-216). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Siemund, P. & Dolberg, F. (2011).** From lexical to referential gender: An analysis of gender change in medieval English based on two historical documents. *Folia Linguistica*, 45(2), 489-534.
- Siewierska, A. (1999).** From anaphoric pronoun to grammatical agreement marker: Why objects don't make it. In G. G. Corbett (Ed.), *Folia Linguistica* (pp. 225-251), XXXIII/2 - Special Issue on Agreement.
- Singleton, D. (2005).** The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, 43 (4), 269-285.



- Slabakova, R. (2006).** Is there a critical period for semantics?. *Second Language Research*, 22 (3), 1–37.
- Steele, S. (1978).** *Word order variation: a typology study*. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language Vol. 4: Syntax* (pp. 585-623). Stanford, California: Stanford University Press.
- Suisse, A. (2011).** Os desafios do ensino-aprendizagem do português como L3 no contexto universitário marroquino. *Magriberia*, (4), 169-184.
- Tsimpli, I. M. (2003).** Clitics and Determiners in L2 Greek. In J. M. Liceras *et al.* (Eds.), *Proceedings of the 6<sup>th</sup> Generative Approaches to Second Language Acquisition, Somerville* (pp. 331-339), MA: Cascadilla Proceedings Project.
- Tsimpli, I. M. & Dimitrakopoulou, M. (2007).** The interpretability hypothesis: evidence from wh-interrogatives in second language acquisition. *Second Language Research*, 23 (2), 215-242.
- Tsimpli, I. M. & Mastropavlou, M. (2008).** Feature Interpretability in L2 Acquisition and SLI: Greek Clitics and Determiners. In J. Liceras; H. Zobl & H. Goodluck (Eds.), *The role of formal features in Second Language Acquisition* (pp. 142-183), New York: Taylor and Francis Group Routledge.
- Ullman, M. T. (2001a).** The Declarative/Procedural Model of Lexicon and Grammar. *Journal of Psycholinguistic Research*, 30 (1), 37-69.
- Ullman, M. T. (2001b).** The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, 4 (1), 105-122.
- Ullman, M. T. (2001c).** A Neurocognitive Perspective on Language: The Declarative / Procedural Model. *Neuroscience*, 2, 717-727.
- Ullman, M. T. (2004).** Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- Vainikka, A. & Young-Scholten, M. (1994).** Direct access to X<sup>2</sup>-theory: evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition studies in generative grammar* (pp. 265-316). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- VanPatten, B. (1996).** *Input processing and grammar instruction*. Norwood, New Jersey: Ablex Publishing Co.

- VanPatten, B. (2008).** *Input Processing in Adult Second Language Acquisition.* In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition – an Introduction* (pp. 115-135). New York/London: Routledge & Francis Group.
- Vigliocco, G.; Franck, J. (1999).** When Sex and Syntax Go Hand in Hand: Gender Agreement in Language Production. *Journal of Memory and Language*, 40, 455-478.
- Vilela, M. (1973).** Considerações sobre o género gramatical. *Revista da Faculdade de Letras da Universidade do Porto*, I, Série de Filologia, 139-150.
- Vilela, M. (1999).** *Gramática da língua portuguesa* (2ª edição). Coimbra: Almedina.
- Villalva, A. (1994).** *Estruturas morfológicas. Unidades e Hierarquias nas Palavras do Português.* Lisboa: Fundação Calouste Gulbenkian.
- Villalva, A. (2003).** Estrutura mórfica básica. In Mateus, Maria H.M. *et al.*, *Gramática da Língua Portuguesa* (5ª ed., revista e aumentada) (pp. 919-931). Lisboa: Caminho.
- Villalva, A. (2008).** *Morfologia do Português.* Lisboa: Universidade Aberta.
- Vincent, N. (1988).** Italian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 279-313). New York: Routledge.
- Weinreich, U. (1968)** *Languages in contact: findings and problems* (9<sup>th</sup> Edition). The Hague / Paris: Mouton & Comp.
- Wheeler, B. I. (1899).** The Origin of Grammatical Gender. *The Journal of Germanic Philology*, 2(4), 528-545.
- White, L. (1990).** Another look at the logical problem of foreign language learning: a reply to Bley-Vroman. *Linguistic Analysis*, 20 (1-2), 50-63.
- White, L. (1993).** Universal Grammar: Is It Just a New Name for Old Problems?. In S. Gass & L. Selinker (Eds.), *Language Transfer in Language Learning (revised edition)* (pp. 217-312). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- White, L. (1995).** *Universal Grammar and Second Language Acquisition.* Amsterdam/Philadelphia: John Benjamins Publishing Company.
- White, L. (2003).** *Second Language Acquisition and Universal Grammar.* Cambridge: University Press.
- White, L. (2008).** Linguistic Theory, Universal Grammar, and Second Language Acquisition In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition – an Introduction* (pp. 37-55). New York/London: Routledge & Francis Group.

- White, L.; Valenzuela, E.; Kozłowska-Macgregor, M.; Leung, Y. I. (2004).** Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 2, 105-133.
- Wu, Y. & Bodomo, A. (2009).** Classifiers ≠ Determiners. *Linguistic Inquiry*, 40 (3), 487-507. Disponível em: [https://www.researchgate.net/publication/254956182\\_Classifiers\\_a\\_Determiners](https://www.researchgate.net/publication/254956182_Classifiers_a_Determiners) (consultado em 02.03.2016).
- Zobl, H. (1980).** Developmental and Transfer Errors: Their Common Bases and (Possibly) Differential Effects on Subsequent Learning. *TESOL Quarterly*, 14 (4), 469-479.
- Zubin, D. A. & Köpcke, K. (1984).** Affect Classification in the German Gender System. *Lingua*, 63, 41-96. Disponível em: [https://www.uni-muenster.de/imperia/md/content/germanistik/lehrende/koepcke\\_k/1984\\_zubin\\_koepcke\\_affect\\_classification.pdf](https://www.uni-muenster.de/imperia/md/content/germanistik/lehrende/koepcke_k/1984_zubin_koepcke_affect_classification.pdf) (consultado em 02.03.2018).
- Zubin, D. A. & Köpcke, K. (1986).** Gender and folk taxonomy: the indexical relation between grammatical and lexical categorization. In C. Craig (Ed.), *Noun Classes and Categorization*, Vol. 7 (pp. 139-180). Amsterdam/Philadelphia: John Benjamins Publishing Company.

### Sítios consultados na Internet

Cátedra, Português Língua Segunda e Estrangeira, Instituto Camões, Universidade Eduardo Mondlane, Moçambique. Disponível em: <http://www.catedraportugues.uem.mz/> (consultado em 12.11.2016).

*Corpus* de Produções Escritas por Aprendentes de Português como L2 (PEAPL2). Disponível em: <http://teitok2.iltec.pt/peapl2/> (consultado em 10.11.2018).

Dicionário infopedia da Língua Portuguesa [em linha]. Porto: Porto Editora. Disponível em <https://www.infopedia.pt/dicionarios/lingua-portuguesa/> (consultado em 05.03.2015).

Dicionário PRIBERAM da Língua Portuguesa [em linha], 2008-2013. Disponível em: <https://dicionario.priberam.org/> (consultado em 05.03.2015).

Moçambicanismos, um Glossário com muitas imagens. Disponível em: <http://mocambicanismos.blogspot.com/2009/01/s.html> (consultado em 12.11.2016).

Projeto de Recolha de Dados de Aprendizagem de Português como Língua Estrangeira. Disponível em <http://www.clul.ulisboa.pt/pt/24-recursos/350-recolha-de-dados-de-ple> (consultado em 10.11.2018).

Secção “Produção Escrita de falantes aprendentes adultos” do *Corpus de Aquisição de PL2* (CAL2). Disponível, mediante registo prévio, em: <http://cal2.clunl.edu.pt/> (consultado em 10.11.2018).

*World Atlas of Language Structures Online* [WALS], 30A. Disponível em <http://wals.info/chapter/30> (consultado em 10.06.2014).





