

# ESTUDOS DE LINGUÍSTICA

VOLUME II

CONCEIÇÃO CARAPINHA  
ISABEL A. SANTOS

COORD.



IMPRESA DA UNIVERSIDADE DE COIMBRA  
COIMBRA UNIVERSITY PRESS

# **A MARCAÇÃO DE GÉNERO NOMINAL EM PORTUGUÊS LÍNGUA NÃO-MATERNA: PADRÕES NA AQUISIÇÃO/APRENDIZAGEM**

**Tânia Santos Ferreira**

*Centro de Estudos de Linguística Geral e Aplicada (CELGA)*

*tania.ferreira88@gmail.com*

## **1. Introdução e objetivos**

A categoria de género (GEN) possui um estatuto especial entre as demais categorias, principalmente quando integrada num contexto de aquisição/aprendizagem de uma língua não-materna (LNM). Por não estar presente em todos os sistemas linguísticos do mundo, e dada a variabilidade de critérios de atribuição de GEN existente, esta categoria representa uma das áreas da gramática provavelmente ‘resistente’ ao processo de aquisição/aprendizagem de PLNM (português língua não-materna)<sup>53</sup>.

No presente artigo, apresentamos os resultados de um trabalho cujo principal objetivo é a identificação de padrões variáveis na aquisição/aprendizagem da marcação do GEN dentro do grupo nominal, a partir da análise de desvios produzidos por alunos estrangeiros de diferentes níveis de proficiência em PLNM. Mais especificamente, pretende-se apurar, tendo em conta as particularidades da categoria de GEN em português, que tipo de características formais e semânticas dos nomes suscita maiores

---

<sup>53</sup> O presente trabalho concentra-se no tratamento de alguns aspetos estudados no âmbito da tese de Mestrado intitulada *Padrões na aquisição/aprendizagem da marcação do género nominal em português como L2*, apresentada à Faculdade de Letras da Universidade de Coimbra, em novembro de 2011.

dificuldades aos aprendentes durante o processo de aquisição/aprendizagem da marcação de GEN.

Assim, no que respeita ao enquadramento teórico, considerou-se, em primeiro lugar, o conceito de *Interlândia* (IL) proposto por Larry Selinker ([1972] 1992) e, em seguida, procedeu-se à descrição das propriedades do sistema de marcação de GEN em português, seguindo atentamente o trabalho de G. Corbett (1991).

## 2. Pressupostos teóricos

### 2.1. Interlândia

Diversos investigadores (Corder, 1992; Selinker, 1992) consideraram que o sistema linguístico atualizado pelos aprendentes tardios de uma LNM apresenta características próprias. Segundo estes autores, os ‘erros’<sup>54</sup> cometidos pelos falantes aprendentes não devem ser considerados fruto do acaso e, conseqüentemente, necessitam de ser analisados, no sentido de comprovarem a existência de um sistema linguístico intermédio, condicionado por regras próprias. A partir deste ponto de vista, Selinker propõe, no artigo seminal *Interlanguage*, publicado em 1972, o conceito de *Interlândia* para designar esse sistema linguístico intermédio, independente e único, com uma organização baseada em regras e princípios (Selinker, 1992), evidenciado pelos aprendentes de uma LNM.

Considera-se que as IL correspondem a sistemas linguísticos, dado que durante a sua construção, os fenômenos desviantes produzidos pelos aprendentes em relação à língua-alvo são, de certo modo, previsíveis.

---

<sup>54</sup> Convém salientar que nem todos os erros cometidos pelo falante devem ser objeto de investigação. Segundo Corder (1992) é necessário estabelecer a distinção entre erros não sistemáticos e sistemáticos, sendo que os primeiros – designados como ‘faltas’ – correspondem a lapsos casuais produzidos pelos aprendentes e “no tienen ninguna importancia para el proceso del aprendizaje de la lengua” (1992: 37) e os segundos – referidos como ‘erros’ – designam os desvios recorrentes na produção dos falantes, devendo ser, por isso, analisados. Todavia, Corder sublinha que distinguir as ‘faltas’ de ‘erros’ pode ser uma tarefa muito complicada para o investigador e implica uma análise exaustiva dos dados, concluindo que só o contexto da produção permite indicar a presença de um ‘erro’ (1992: 39).

Tendo em conta este facto, no presente trabalho pretende-se verificar se os desvios produzidos pelos alunos cujas produções constituem o *corpus* analisado permitem identificar o processo de construção da IL, i.e., se é possível inferir diferentes fases do desenvolvimento linguístico do falante com base na correlação entre fenómenos desviantes observados na marcação do GEN nominal e o nível de proficiência/competência linguística dos aprendentes de PLNM.

Em boa verdade, à semelhança do que ocorre na aquisição da língua-materna, os desvios sistemáticos dos falantes serão representativos das diferentes etapas por que os aprendentes passam ao longo da aquisição/aprendizagem de uma LNM, i.e., evidenciam um processo gradual, transicional, reflexo do que Corder (1992) designa por 'built-in-syllabus'. O processo de construção da IL é, então, encarado como um *continuum* que perpassa por distintos estágios de desenvolvimento (Ellis, 2003).

Convém salientar que as IL são sistemas que não correspondem, integralmente, a um conjunto de estruturas desviantes em relação à língua-alvo. Trata-se de sistemas flutuantes, em permanente reestruturação e resultam, essencialmente, do facto de os aprendentes irem colocando hipóteses em relação ao *input* linguístico que vão recebendo (Corder, 1992).

Selinker (1992) identifica vários processos atuantes na construção dos sistemas interlinguísticos. Um desses processos designa-se por transferência de instrução, que se caracteriza por manifestações desviantes em relação à língua-alvo atribuíveis a hábitos ou métodos de ensino (Selinker, 1992: 88). Outro fator corresponde às estratégias de aprendizagem da língua-alvo, observáveis no facto de o aprendente tender à simplificação do material linguístico da L2 (Selinker, 1992: 88) em que evidencia, por exemplo, o não emprego dos artigos, a ausência de marcação do plural ou a omissão de formas verbais<sup>55</sup>.

---

<sup>55</sup> Como exemplo de estratégia conducente à simplificação do material linguístico, Selinker (1992: 88) refere que: "(...) los resultados de esta estrategia se manifiestan en la totalidad de los niveles de la sintaxis de la IL de los hablantes del inglés de la India. (...) si el alumno ha adoptado la estrategia de que todos los verbos son o transitivos o intransitivos, puede producir en su IL oraciones como (4) I am feeling thirsty (...); con lo que parece haber adoptado la estrategia de que la realización de la categoría *aspecto* en su forma progresiva aparece sempre con la marca *-ing* en la estructura superficial (...)"

A sobregeneralização das regras da língua-alvo atua, igualmente, na construção das IL e pode ser vista como resultando da estratégia de simplificação da língua-alvo pelos aprendentes. O produto de tais sobregeneralizações verifica-se, sobretudo, nos planos morfológico e morfo-sintático através da regularização dos paradigmas flexionais (Selinker, 1992). Quando o aprendente assume ter atingido, de forma satisfatória, os objetivos comunicativos da língua-alvo, pode considerar que a sua IL está apta a desempenhar as funções comunicativas dessa língua e, nesse caso, tende a ‘estabilizar’ determinadas estruturas, cessando o desenvolvimento da IL (Selinker, 1992).

Outro processo indicado por L. Selinker (1992) corresponde à transferência linguística. De um modo geral, a transferência designa a presença de informação linguística da língua-materna nos enunciados da L2. Embora este fenómeno seja amplamente discutido nos trabalhos desenvolvidos sobre Aquisição de Línguas Segundas<sup>56</sup>, para este estudo não foi possível evidenciar e analisar detalhadamente a presença deste processo na aquisição/aprendizagem da marcação de GEN nominal. Assim, só os processos anteriormente referidos serão considerados durante a apreciação dos resultados.

## **2.2. O sistema de marcação de género nominal em português**

goria com relevância para a classificação dos nomes<sup>57</sup>. O GEN representa, sem dúvida, uma das categorias gramaticais mais singulares, visto que não possui um estatuto universal. Nichols (1999), ao analisar um total de 174 línguas, concluiu que apenas uma pequena parte possuía um sistema de GEN ou outro tipo de classificação nominal. Também Corbett

---

<sup>56</sup> Para uma revisão aprofundada do fenómeno de transferência linguística, veja-se Gass, 1996: 317-345.

<sup>57</sup> Convém salientar que a marcação do GEN pode, nas várias línguas do mundo, afetar diferentes classes de palavras por via da sua necessidade de concordância em GEN com o nome. Em português, por exemplo, a marcação do GEN afeta – além do nome – determinantes, quantificadores, adjetivos e alguns pronomes. Já noutras línguas, o GEN pode ser marcado também em verbos, tal como demonstra o estudo sobre a diversidade linguística de Nichols (1999: 161).

(1991) afirma que, enquanto o GEN é central em algumas línguas, noutras está completamente ausente.

G. Corbett (1991: 2) verifica que o número de categorias de GEN é variável de língua para língua. Para além da oposição masculino/feminino que se encontra, por exemplo, em português e no castelhano, há línguas que admitem o contraste masculino/feminino/neutro, existindo ainda idiomas que introduziram, em função de diferentes critérios, novas subclasses de GEN.

O GEN possui, ainda, um estatuto particular visível pela multiplicidade de critérios que cada língua seleciona e que servem de base à definição da categoria. Segundo a proposta de Corbett (1991), a marcação do GEN de um nome está dependente de dois tipos de informação: o significado do item nominal e a respetiva forma. Tais propriedades (semânticas e formais), que servem de base à definição do conceito de GEN gramatical, só podem ser consideradas numa língua se, nela, espoletarem o fenómeno de concordância sintática. Este facto permite distinguir os sistemas de classes de GEN dos sistemas de classes nominais<sup>58</sup> que se caracterizam pela presença de partículas morfológicas agregadas ao nome cuja função é a de ‘transportar’ informação relativa ao número (singular/dual/plural/etc.), aos traços semânticos, etc. Além disso, tais classificadores não desencadeiam um processo de concordância. Deste modo, e nas palavras de Hockett (1958: 231), as categorias de GEN “are classes of nouns reflected in the behavior of associated words”.

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

---

<sup>58</sup> Nas línguas africanas, por exemplo, predomina o sistema de classes nominais que se caracteriza pela presença dos classificadores, i.e., partículas morfológicas (por ex. prefixos) agregadas ao nome e que identificam a classe a que estes pertencem (cf. Payne, 2003: 109).

No que respeita aos critérios semânticos e, no caso dos nomes com referentes sexuados, por oposição aos [-animados], existe, na verdade, uma certa base semântica ‘natural’ na atribuição do GEN aos nomes, tendo em conta as características da entidade designada pelo item nominal. Porém, a distinção entre masculino e feminino que corresponde a uma distinção de *género natural* não deve ser confundida com o conceito de GEN gramatical. O *género natural* corresponde ao sexo das entidades com o traço semântico [+animado], ao passo que o GEN gramatical, além da ocasional correspondência com o dos seres sexuados, representa um sistema de classificação gramatical dos nomes que pode, conforme as línguas, não ter qualquer ligação com o *género natural*. De facto, em português se existem nomes [+animados] de GEN masculino por se referirem a uma entidade do sexo masculino (*professor, homem*) e nomes femininos cujo referente é uma entidade do sexo feminino (*professora, mulher*), também existem nomes cujo valor de GEN não corresponde necessariamente ao sexo dos seus referentes. Há itens nominais [+animados] que não apresentam esta correspondência, como os sobrecomuns, que são sempre ou masculinos ou femininos independentemente do sexo dos referentes (*a pessoa, a vítima, o cônjuge*). Já no alemão, por exemplo, o neutro abrange alguns nomes cujo referente é sexuado (*das Kind* ‘a criança’, *das Fräulein* ‘a menina, não casada’) (Bechara, 1999).

No que aos critérios formais diz respeito, Corbett (1991) indica que estes são constituídos tanto por regras do tipo morfológico – que têm em conta a estrutura mórfica da palavra e compreendem mecanismos como a flexão e a derivação –, como por regras do tipo fonológico. i.e., casos em que a posição da sílaba tónica pode condicionar o valor de GEN (Corbett, 1991: 33).

No caso do português, as regras morfológicas podem condicionar a marcação do GEN dos nomes [-animados]. É comum associar-se ao Índice Temático (IT)<sup>59</sup> -o o valor de GEN masculino e ao IT -a o valor de GEN

---

<sup>59</sup> O Índice Temático (IT) corresponde à vogal átona que surge na fronteira direita do nome e, à semelhança do que ocorre com as vogais temáticas dos verbos, funciona como um meio de classificação dos itens nominais. Em português, os nomes distribuem-se por diferentes classes (temáticas) e podem ser de tema em -o, -a, -e, de tema Ø (quando termi-

feminino. Em boa verdade, se existem nomes em que é possível estabelecer essa correlação (*o livro, a caneta*), casos há em que não se pode estabelecer tal ligação (*o mapa, a rádio* ‘estação’). Acrescente-se ainda a existência de nomes com diferentes terminações e que podem estar associados aos dois valores de GEN (*a ponte / o pente; o corrimão / a mão; o cobertor / a dor*).

Na tentativa de apurar a distribuição relativa dos nomes de GEN masculino e feminino por classe temática, procedeu-se à análise dos índices de frequência dos nomes em português a partir das ocorrências dos lemas recenseados no *corpus* de frequências lexicais do português europeu contemporâneo intitulado *Léxico Multifuncional Computorizado do Português Contemporâneo* (CORLEX). No que respeita à correlação nome masculino com IT -o e nome feminino com IT -a, verificou-se que esta ‘regra’ de marcação de GEN atua em cerca de 60% das ocorrências nominais em português<sup>60</sup>. Assim sendo, poder-se-á esperar que para o aprendente de PLNM este possa ser um critério que auxilie o processo de escolha do do valor de GEN destes itens.

### 3. Trabalho empírico

#### 3.1. Descrição do *corpus*

Os textos escritos que servem de base empírica a esta investigação foram selecionados a partir de um conjunto alargado de produções que integram o “*Corpus* de Produções Escritas de Aprendentes de PL2” (PEAPL2)<sup>61</sup> (coordenado por Cristina Martins). Procedeu-se a uma análise rigorosa da totalidade das produções escritas disponíveis no *Corpus* PEAPL2, de

---

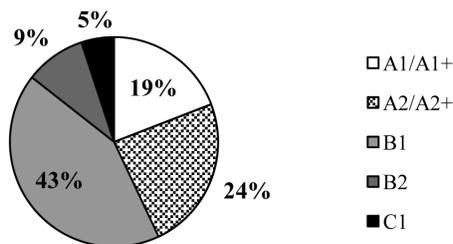
nam em consoante) ou atemáticos (quando terminam em vogal tónica). Sobre este assunto, veja-se Villalva, 2003: 921-925.

<sup>60</sup> Para informação mais detalhada sobre este estudo, veja-se Ferreira, 2011.

<sup>61</sup> Todas as produções que compõem o *corpus* PEAPL2, bem como o perfil dos informantes, estão disponíveis no endereço: <http://www.uc.pt/fluc.rcpl2/>. Para mais informações sobre o método de recolha dos textos, veja-se Martins (2013).

modo a verificar que textos apresentavam desvios em relação à marcação do GEN dentro do grupo nominal. Este processo foi essencial para, numa fase seguinte, se proceder à criação de uma tipologia de desvios. Assim, do conjunto total de 546 textos produzidos por 391 aprendentes de PLNM que, entre maio de 2009 e maio de 2010, frequentaram cursos de PLNM na Faculdade de Letras da Universidade de Coimbra (FLUC), selecionaram-se 323 textos com desvios de marcação de GEN nominal, produzidos por 256 informantes.

O conjunto de informantes é muito diversificado<sup>62</sup>. Trata-se, sobretudo, de aprendentes tardios, com diferentes línguas maternas e que se distribuem por diversos cursos de PLNM da FLUC, que correspondem a diferentes níveis de proficiência. Deste modo, foi possível estabelecer a correspondência entre o nível de cada turma e os níveis do *Quadro Europeu Comum de Referência para as Línguas (QECRL)*<sup>63</sup>. A determinação dos diferentes níveis de proficiência linguística foi muito importante, dado que é sobre esta variável que os dados deste estudo foram analisados. O Gráfico 1 apresenta a distribuição dos informantes por nível de proficiência linguística.



**Gráfico 1** – Distribuição dos informantes por nível comum de referência do QECRL

<sup>62</sup> Para uma consulta detalhada do perfil dos informantes deste estudo, veja-se Ferreira (2011: 46-50).

<sup>63</sup> De modo a avaliar em qual dos níveis de proficiência em PLNM se encontravam e em que turma/nível se ajustariam melhor, os aprendentes que frequentavam o *Curso Anual* e o *Curso de Férias de Língua e Cultura Portuguesas para Estrangeiros* deveriam responder a um teste diagnóstico de português. No caso específico dos alunos *Erasmus*, estes eram integrados nas turmas de acordo com o nível de aprendizagem formal do português que frequentavam nos seus países de origem. Todavia, convém salientar que não foi possível garantir que todos os alunos ficassem inseridos nas turmas mais adequadas ao seu nível de proficiência.

Como se pode constatar, os informantes distribuem-se pelos seguintes níveis QECRL: os níveis elementares (A1 e A2); os níveis intermédios (B1 e B2) e o nível proficiente (C1), sendo o nível B1 o mais representativo, com cerca de 43% dos informantes.

### 3.2. Tipologia de desvios

Como se explicou para a seleção dos textos do *corpus* que serviram de base à investigação, foram consideradas todas as estruturas com desvios quanto à marcação de GEN em português dentro do grupo nominal, na medida em que serão reveladoras dos valores de GEN atribuídos aos nomes pelos aprendentes de PLN. Do *corpus* em estudo apurou-se um conjunto total de 689 desvios (cf. Quadro 1) e, em seguida, procedeu-se à criação de uma tipologia dos casos desviantes, de modo a verificar quais os tipos de nomes mais afetados. Esta tipologia baseia-se nas características da estrutura do GEN do português, língua-alvo, que foram consideradas no ponto 2.2 deste trabalho.

Desta forma, registaram-se os desvios de atribuição do valor de GEN nominal observáveis quer através da forma morfológica do nome, quer através das demais palavras que com este deverão concordar em GEN, coocorrendo no grupo nominal: determinantes, quantificadores e adjetivos. Para a organização dos dados, criaram-se cinco códigos representativos de cinco categorias de desvio de marcação de GEN nominal.

Assim, o código:

- (1) identifica os desvios relativos a nomes [+ animados] (*o menino, a menina*), em que o valor de GEN coincide com o sexo do referente;
- (2) identifica os desvios relativos a nomes com referentes extralinguísticos [- animados] cujo GEN seria corretamente dedutível a partir do IT: se IT *-o*, o nome é masculino (*o barco, o mercado*); se IT *-a*, o nome é feminino (*a lua, a casa*);
- (3) identifica os desvios relativos a nomes com referentes extralinguísticos [- animados] cujo GEN não é dedutível a partir do IT: se

IT é *-o*, o nome é feminino (*a tribo, a foto*); se o IT é *-a*, o nome é masculino (*o mapa, o dia*);

- (4) identifica os desvios relativos a nomes com referentes sexuados cujo valor de GEN não coincide, necessariamente, com o sexo do referente (*a pessoa, a criança; o cônjuge*);
- (5) identifica os desvios relativos a nomes com referentes [- animados] cujo GEN somente é inferível a partir de marcadores sintáticos (determinantes, adjetivos, etc.). É o caso dos nomes de tema em *-e* e dos atemáticos (*o pente, a ponte; o camião, a mão*).

Em cada categoria existem quatro subcategorias que distinguem os diferentes constituintes do grupo nominal nos quais se evidencia(m) a(s) marca(s) do desvio: [a] o desvio é marcado na forma morfológica do nome; [b] o desvio ocorre no determinante; [c] o desvio ocorre no adjetivo; e [d] o desvio ocorre no quantificador. No Quadro seguinte apresentam-se alguns exemplos de desvio extraídos do *corpus*, organizados por categoria e subcategoria<sup>64</sup>:

Categoria de desvio	Subcategoria de desvio	Exemplos
1	1a	“ <i>Os alemães</i> chegam (frequentemente) em ponto (...)” [UC. ER.LPIII.A.12.09.74.50.2L] <sup>1</sup>
	1b	“Outras pessoas e <i>meus amigos</i> falam a mim.” [UC. CA.E.B.11.09.07.1.1A]
	1c	“Tenho um <i>irmã</i> mais <i>novo</i> ” [UC.CA.E.A.12.09.04.1.1A]
	1d	“Tenho <i>um irmã</i> mais novo” [UC.CA.E.A.12.09.04.1.1A]
2	2a	“Um outro <i>coiso</i> que eu costumo fazer no meu tempo livre é a de cozinha.” [UC. ER.LPIII.F.06.09.09.33.1J]
	2b	“Ele sabe algum palavras <i>no meu língua</i> .” [UC. CA.E.A.12.09.07.1.1A]
	2c	“Agora estou a estudar <i>Língua português</i> em Coimbra.” [UC. CA.E.A.12.09.01.1.1A]
	2d	“Agora em Natal volto para casa <i>dois semanas</i> (...)” [UC. ER.LPII.A.12.09.08.1.1A]

<sup>64</sup> Todos os exemplos que foram objeto de análise podem ser consultados em Ferreira (2011: volume II, Anexo III).

3	3a	Não há casos deste tipo.
	3b	“Acho que <i>esta dia</i> foram mais que mil pessoas!” [UC.ER.LPIII.F.06.09.11.33.1J]
	3c	“Quando <prec> preciso de um <i>dia calma</i> (...)” [UC.ER.LPIII.F.06.09.28.33.1J]
	3d	“Há <i>muitas</i> (...) <i>cinemas</i> (...)”. [UC.CA.S.A.12.09.16.69.3Q]
4	4a	“A cidade (...) não é cheia de <i>turistas</i> (...)” [UC.ER.LPIV.A.06.09.39.69.3Q]
	4b	“Mas houve <i>uns pessoas</i> (...)” [UC.CA.E.A.04.10.01.77.3T]
	4c	“ <i>Português pessoas</i> são muito simpáticos” [UC.CA.E.A.05.09.1.33.1J]
	4d	“ <i>Alguns pessoas</i> dizem viagem (...)” [UC.CA.E.B.11.09.08.75.3S]
5	5a	“Agora as minhas <i>examas</i> (...)” [UC.ER.LPI.A.05.10.10.6.1B]
	5b	“Quando eu tenho <i>o possibilidade</i> (...)” [UC.ER.LPIII.F.06.09.12.33.1J]
	5c	“São <i>materiais muito diversas</i> (...)” [UC.CA.S.A.05.09.07.6.1B]
	5d	“Tenho <i>muito saudade</i> para ti!” [UC.ER.LPI.A.05.10.31.6.1B]

**Quadro 1** – Tipologia de desvios

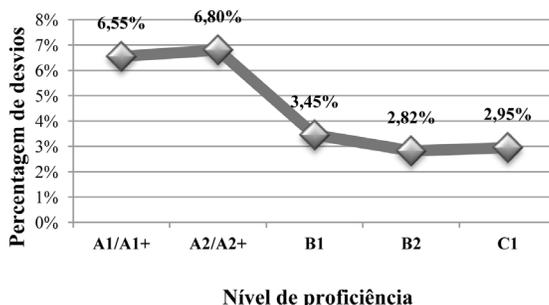
Definidas as características da tipologia de desvios de marcação de GEN nominal, apresentam-se, no próximo ponto deste trabalho, os resultados da análise dos dados apurados, tendo em conta as categorias e subcategorias de desvios criadas para esta investigação e dando particular relevo à distribuição dos desvios por nível de proficiência dos aprendentes. Pretende-se, assim, extrair conclusões sobre o desenvolvimento das respetivas interlínguas, bem como identificar que características formais e semânticas dos nomes suscitam maiores dificuldades na aquisição/aprendizagem da marcação do GEN.

### 3.3. Tratamento e análise dos dados

Num primeiro momento de quantificação dos resultados, verificou-se a frequência dos casos desviantes em função dos níveis comuns de referência do QECRL representados pelos aprendentes que produziram os dados do *corpus*. Deste modo, a partir da totalidade dos nomes produzidos pelo conjunto dos aprendentes de um dado nível (Quadro 2), procedeu-se ao cálculo da respetiva percentagem de desvios (Gráfico 2), de modo a tornar possível a confrontação dos dados em função desta variável.

Nível do QECRL	# Nomes por nível	# Desvios
A1/A1+	2.432	159
A2/A2+	2.443	166
B1	8.425	293
B2	1.589	45
C1	879	26
<b>TOTAL</b>	<b>15.768</b>	<b>689</b>

**Quadro 2** – Frequência de desvios produzidos por nível

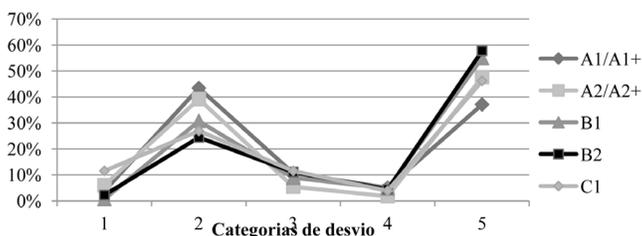


**Gráfico 2** – Percentagem de desvios em função dos nomes produzidos por nível de proficiência em PL2

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

Tendo em conta estes resultados, apresenta-se em seguida a análise detalhada dos desvios produzidos por nível do QECRL representado neste *corpus*, em função das cinco categorias e subcategorias de desvios da tipologia criada (cf. secção 3.2).

No Gráfico 3 regista-se a percentagem de desvios por categoria calculada em função do número total de desvios registados em cada nível de proficiência representado neste estudo (cf. Quadro 2).



**Gráfico 3** – Percentagem total de categorias de desvios produzidos por nível de proficiência dos aprendentes

Mediante a leitura do Gráfico, constata-se que uma das categorias de desvio menos afetadas é a Categoria (1). Trata-se de desvios relativos a nomes [-animados] cujo valor de GEN corresponde ao sexo do seu referente. Assim, poder-se-á admitir que, durante a aquisição/aprendizagem da marcação de GEN os aprendentes de PLNM reconhecem mais facilmente os valores de GEN dos nomes cuja distinção masculino/feminino corresponda a uma diferença de sexo, do tipo *menino/menina*.

Pelo contrário, a categoria de desvio mais afetada em todos os níveis é a Categoria (5), i.e., os desvios relativos a nomes [-animados] cujo valor de GEN somente é visível a partir de marcadores sintáticos. Uma vez que os valores de GEN destes nomes não são visíveis a partir de um conjunto de critérios que permita aos aprendentes de PLNM estabelecer corretamente o seu valor de GEN, a percentagem de desvios em todos os níveis é, como seria de esperar, consideravelmente alta.

A segunda categoria de desvio mais afetada é a categoria 2 que representa os desvios relativos a nomes [-animados], cujo valor de GEN seria corretamente dedutível a partir do IT. Tal resultado é bastante surpreendente e, para justificar estes valores, várias hipóteses podem ser equacionadas.

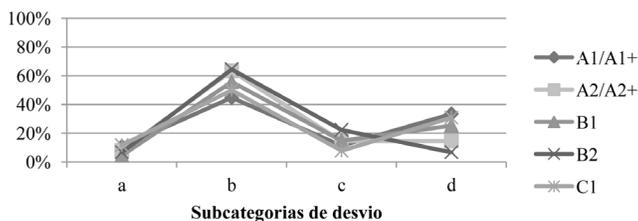
Em primeiro lugar, estes dados parecem indiciar que muitos aprendentes não incorporaram, de modo satisfatório, algumas pistas formais de marcação de GEN, não sendo cabalmente capazes de estabelecer, à medida que produzem estes itens nominais, a correspondência entre os valores de GEN e respectivos IT. Porém, deve equacionar-se que o aluno pode saber as regras de atribuição do valor de GEN aplicáveis em portu-

guês, mas como não reconhece o GEN de determinado nome, é provável que lhe atribua um IT errado ou marque mal o constituinte dentro do grupo nominal (quantificador, determinante ou adjetivo). Em suma, o aluno pode saber a regra de marcação de GEN, mas não conhece o valor de GEN associado a estes nomes.

Outra hipótese explicativa pode estar correlacionada com o facto de os aprendentes apresentarem uma certa ‘desconfiança’ face aos indícios morfológicos, como o IT, para atribuir corretamente os valores de GEN aos nomes. De facto, e como se viu a partir do estudo da base de dados da CORLEX referido na secção 2.2, embora este tipo de nomes seja muito frequente no *input* dos aprendentes (em cerca de 60% das frequências nominais do português), pode não ser, todavia, suficiente para os alunos confiarem plenamente nestas pistas de atribuição de GEN.

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

No que concerne às subcategorias de desvio, observe-se o Gráfico 4 com os valores percentuais calculados em função do total de desvios produzido por nível de proficiência em PLNM (cf. Quadro 2).



**Gráfico 4** – Percentagem total das subcategorias de desvio produzidas por nível de proficiência

Da análise do Gráfico, observa-se que uma das subcategorias menos afetada é a subcategoria [c] que diz respeito aos desvios marcados nos adjetivos.

Em contrapartida, as subcategorias de desvio mais afetadas são as subcategorias [b] e [d] que correspondem, respetivamente, aos desvios de

marcação de GEN no determinante e no quantificador. Trata-se de classes gramaticais que, provavelmente terão, no entender do aprendente, pouco valor do ponto de vista comunicativo, pois este acaba por constatar que, mesmo ‘ignorando’ estas marcas, consegue atingir os seus objetivos comunicativos. Com efeito, e como se assinalou na secção 2.1, durante a construção da IL, os aprendentes tendem a simplificar o material linguístico da língua-alvo optando, assim, por ‘dispensar’ estas classes.

#### **4. Considerações finais**

As correlações estabelecidas entre tipos de desvios, marcas de GEN e nível de proficiência permitem concluir que em todos os níveis de proficiência em PLN analisados se registam desvios de marcação de GEN. Embora os índices de desvio sejam menores nos níveis de proficiência mais elevados, a verdade é que os desvios de atribuição de GEN nunca chegam a ser erradicados, ou seja, dificilmente os aprendentes conseguem, de modo pleno, atribuir corretamente os valores de GEN aos nomes.

Quanto ao tipo de nome mais afetado, pôde constatar-se que os itens cuja marcação de GEN não depende de critérios de natureza quer semântica quer formal suscitam maiores dificuldades aos aprendentes, registando-se igualmente uma grande percentagem de desvios relacionados com itens nominais [-animados] em que o valor de GEN é inferível a partir do seu IT. Estes resultados sugerem, assim, uma certa ‘desconfiança’ do aprendente em relação à fiabilidade do IT como marcador de GEN.

Por fim, o avultado número de casos de atribuição de GEN nominal desviantes parece sugerir que, no processo de aquisição/aprendizagem, a assimilação do valor de GEN de um nome é progressiva e relativamente lenta.

#### **Referências bibliográficas**

- Bechara, Evanildo (1999): *Moderna Gramática Portuguesa*, Rio de Janeiro: Editora Lucerna.  
Corbett, Greville G. (1991): *Gender*, Cambridge: Cambridge University Press.

- Corder, S.P. (1992): *La importancia de los errores del que aprende una lengua segunda*. In: Licerias, J.M., *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*, Madrid: Visor, 31-40.
- Ellis, Rod (2003): *Second Language Acquisition* (9ª edição), Oxford: Oxford University Press.
- Ferreira, Tânia Santos (2011): *Padrões na aquisição/aprendizagem da marcação do género nominal em português como L2*. Dissertação de mestrado em Português Língua Estrangeira/Língua Segunda, apresentada à Faculdade de Letras da Universidade de Coimbra.
- Gass, S. (1996): *Second Language Acquisition and Linguistic Theory: the Role of Language Transfer*. In: Bhatia, Tej. K. & Ritchie, William C. (Eds.), *The Handbook of Second Language Acquisition*, San Diego, California: Academic Press, 317-345.
- Hockett, Charles F. (1958): *A Course in modern linguistics*, New York: The Macmillan Company.
- Martins, Cristina (2013): *O Corpus de Produções Escritas de Aprendentes de PL2 (PEAPL2/CELGA). Caracterização e desenvolvimento de uma infraestrutura de investigação*. In: Flores, Cristina (Ed.). *Português Língua Não Materna: Investigação e ensino*, Lisboa: Lidel, 70-80.
- Nichols, J. (1999): *Linguistic Diversity in Space and Time*, Chicago, IL: University of Chicago Press.
- Payne, Thomas E. (2003): *Describing morphosyntax: a guide for field linguistics*, Cambridge: Cambridge University Press.
- Selinker, L. (1992): *La Interlengua*. In: Licerias, J.M., *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*, Madrid: Visor, 79-101.
- Villalva, Alina (2003): *Estrutura mórfica básica*. In: Mateus, Maria H. M., et. *alli*, *Gramática da língua portuguesa* (5ª edição revista e aumentada), Lisboa: Caminho, 919-931.