

Aprendentes espanhóis de português língua estrangeira: atribuição de género gramatical em palavras sufixadas

Spanish learners of Portuguese as a non-native language: grammatical gender assignment in suffixed words

Tânia Santos Ferreira^{1,a} , Graça Rio-Torto^{1,b} 

¹ Centro de Estudos de Linguística Geral e Aplicada (CELGA-ILTEC), Universidade de Coimbra, Portugal

✉ ^ataniasantos@fl.uc.pt

✉ ^briortorto@fl.uc.pt

Recebido: 11/05/2022; Aceito: 09/11/2022

Resumo

Este estudo visa analisar a atribuição de género gramatical em derivados sufixados do português por aprendentes tardios, nativos de espanhol, a frequentar turmas de diferentes níveis de aprendizagem, de A1 a C2. No inquérito aplicado, foram selecionados nomes de evento, de estado e/ou de propriedade, portadores de sufixos inscritos em classes temáticas marcadas por grande opacidade, como os nomes atemáticos, sufixados em *-agem*, os de Tema \emptyset ([ẽw]), sufixados em *-ão*, *-ção* e *-s[z]ão* e os de Tema *-e* (invariável), sufixados em *-ice* e *-idade*. Partindo da análise dos dados empíricos, registou-se, em termos globais, uma diminuição considerável de desvios à medida que os aprendentes progredem na aprendizagem do português, já que no segmento dos informantes dos níveis C1-C2, os valores de desvios apurados foram residuais. Verificou-se ainda, e sobretudo nos níveis iniciais e intermédios, uma elevada incidência de desvios nos nomes femininos sufixados em *-agem* que estará correlacionada com o facto de o cognato espanhol *-aje* originar nomes de género masculino. Assim, esta assimetria traduz-se, especialmente em fases iniciais da aprendizagem, por efeitos de transferência linguística no desempenho destes aprendentes. Além disso, os elevados índices de desvios apurados entre os nomes sufixados em *-ice* e também entre os itens sufixados em *-ão* são reveladores de que o grau de representatividade e de opacidade do sufixo no *input* condicionam a atribuição de valores de género aos nomes.

Palavras-chave: aquisição de género; língua portuguesa; língua espanhola; nomes derivados; português como língua estrangeira; interlíngua.



Abstract

This study aims to analyze the grammatical gender assignment of Portuguese suffixed nouns by late native Spanish learners, attending classes of different learning levels, from A1 to C2. In the survey, we selected event, state and/or abstract names with suffixes belonging to thematic classes with a low degree of transparency regarding grammatical gender attribution, such as athematic names, suffixed in *-agem*, those with theme \emptyset ([ẽw]), suffixed in *-ção*, *-s[z]ão* and *-ão*, and those with theme *-e* (invariable), suffixed in *-ice* and in *-idade*. The analysis of the empirical data shows that there was a considerable decrease in deviations as learners progressed in their learning of Portuguese, since in the C1-C2 segment, the deviation values were residual. There was also a high incidence of deviations in feminine nouns suffixed in *-agem*, especially in the beginners and intermediate levels, which may be correlated to the fact that the Spanish cognate *-aje* creates masculine nouns. Thus, this asymmetry translates, especially at early stages of learning, into language transfer effects in the performance of these learners. Furthermore, the high deviation rates found among the names suffixed in *-ice* and also among the items suffixed in *-ão* shows that the degree of representativeness and opacity of the suffix in the input constrain the assignment of gender values to nouns.

Keywords: gender acquisition; Portuguese language; Spanish language; derived names; Portuguese as a foreign language; interlanguage.

SUMÁRIO

1. Introdução.
2. Objetivos e perguntas de investigação.
3. Enquadramento teórico.
 - 3.1. Proximidade tipológica e congruência.
 - 3.2. O papel dos conhecimentos linguísticos prévios.
4. Procedimentos metodológicos.
 - 4.1. Informantes.
 - 4.2. Inquérito.
 - 4.3. Tratamento dos dados.
5. Resultados e discussão.
6. Considerações finais.

1. INTRODUÇÃO

A atribuição de género gramatical é um domínio crítico para qualquer aprendente de uma língua estrangeira (LE), independentemente do grau de proximidade tipológica da língua materna (L1) e da língua-alvo de aprendizagem, já que, mesmo em níveis mais avançados de proficiência linguística, se verificam ocorrências desviantes relativas a esta estrutura.

O estudo que aqui se empreende visa investigar o comportamento de falantes nativos de espanhol na codificação do género gramatical de nomes sufixados do português como LE (PLE), que põe em interação dois sistemas (L1 e LE) de género gramatical ‘rule-governed’ (Tucker, Lambert & Rigault 1977; Audring 2016) de grande proximidade. Por via de regra, as dificuldades e os desvios¹ na atribuição de género gramatical são tanto maiores (i) quanto

maior a assimetria entre L1 e LE nas estratégias de codificação de valores de género, (ii) quanto menor é a proficiência em LE e (iii) quanto maior a ausência de pistas semânticas e/ou formais dos nomes². Ora, os nomes aqui investigados não facultam pistas nem formais nem semânticas nessa matéria, pois são nomes de eventos, de estados e de propriedades, logo não associáveis a qualquer traço de animacidade, e pertencentes a algumas das classes temáticas mais opacas, como são a que termina em *-e* (invariável), a atemática e a de tema \emptyset . Já há muito Taft & Meunier (1998: 29) assinalaram que: “[w]hen that ending was rarely associated with the gender of the word, it was relatively difficult to identify that gender”.

O presente estudo tem a seguinte estrutura: após esta breve introdução, em 2. descrevem-se os objetivos específicos em articulação com os dados empíricos elicitados e as perguntas de investigação que norteiam toda a reflexão. O enquadramento teórico deste trabalho é apresentado em 3. Em 4. descrevem-se os procedimentos metodológicos seguidos, no que concerne ao inquérito construído, ao universo dos informantes e ao tratamento dos dados empíricos. Em 5. procede-se à apresentação dos resultados, sua discussão e as considerações finais são apresentadas em 6.

2. OBJETIVOS E PERGUNTAS DE INVESTIGAÇÃO

O presente estudo procura investigar o comportamento de falantes nativos de espanhol na codificação do género gramatical de nomes sufixados do português como LE (PLE), partindo da aplicação de um inquérito no qual se solicitou aos aprendentes que explicitassem o valor de género gramatical associado a 30 nomes portadores de sufixos presentes em 30 construções frásicas. No âmbito do quadro teórico de referência, que recua à teoria seminal de Selinker (1972), encontram-se os conceitos centrais de interlíngua (cf. também White 2003; Selinker & Rutherford 1992, entre outros), sistema linguístico construído pelo aprendente a partir da organização linguística da sua L1 e do *input* que recebe da língua-alvo, e o de transferência linguística (cf. Gass & Selinker 1993). No que diz respeito ao género, são incontornáveis os trabalhos de Corbett (1991, 2013a, 2013b) e Corbett & Fraser (2000). Os dados sobre género e sobre sistema derivacional da L1 dos aprendentes inquiridos (espanhol) e os da língua-alvo de aprendizagem, o português, são espaldados em Ambadiang (1999), Santiago Lacuesta & Bustos Gisber (1999), Real Academia Española y Asociación de Academias de la Lengua Española (2009), Mattoso Câmara (1970, 1975), Nunes (1989), Rio-Torto *et al.* (2016) e Mota (2020).

Em espanhol, como em português, os nomes distribuem-se por duas grandes classes de género gramatical: os femininos e os masculinos. Nos nomes sufixados, o seu género gramatical está fortemente dependente dos sufixos usados. Em espanhol e em português, a presença de um sufixo derivacional nominal pode estar associada a um valor de género

¹ O termo “desvio” emprega-se, neste estudo, como equivalente a “erro”, i.e., “transgresión involuntária de la «norma»” que, no contexto da aprendizagem linguística, corresponde a um “índice y estrategia en el proceso” (Fernández López 1995: 204). Por conseguinte, os desvios ou erros assinalados neste trabalho constituem evidências das distintas fases do processo de construção do sistema interlinguístico do aprendente à medida que progride na aprendizagem. Com efeito, e como tem sido amplamente descrito pela bibliografia de referência (Corder 1967; Selinker 1972; Lott 1983; Larsen-Freeman & Long 1991; Gass & Selinker 2008), os erros intralinguísticos ou evolutivos integram a construção da interlíngua, sendo alguns transitórios e outros permanentes, i.e., fossilizados. No entanto, tendo em conta a amostra recolhida no presente trabalho, os dados analisados e a nossa experiência, cremos que, na maior parte dos casos, estamos perante desvios/erros que tendem a não atingir uma etapa de fossilização (cf. Long 2003: 521).

² Cf. Ferreira (2019: 198): “A distribuição dos desvios por categoria revela uma maior incidência de desvios em nomes cujo valor de género não se infere a partir de ‘pistas’ semânticas, nem a partir de ‘pistas’ formais”.

gramatical específico e não variável, o chamado género ‘inerente’. É desta classe de sufixos associados a um único valor de género, masculino (MASC) ou feminino (FEM), de que aqui nos ocupamos. Assumimos, assim, conjuntamente com [Ambadiang \(1999: 4876\)](#) que “[e]l afijo impone su rasgo de género al conjunto así formado, el género de los nombres derivados se predice a partir del de su último constituyente sufijal.”

O grau de exposição dos aprendentes de PLE ao uso do sufixo, e nomeadamente nas palavras escrutinadas, tem tanto mais peso quanto maior for a opacidade semântica e formal dos sufixos seleccionados e das classes temáticas em jogo. Assim, pretende-se averiguar se a proximidade tipológica no que diz respeito, nomeadamente, à similaridade dos sistemas de género das duas línguas, a espanhola e a portuguesa, e o conhecimento prévio dos aprendentes inquiridos facilita ou dificulta o grau de acerto/desvio na atribuição de género, em palavras derivadas portadoras de sufixos relativamente ‘opacos’ quanto à atribuição de género gramatical.

Uma das questões que se coloca consiste em averiguar se, no universo de aprendentes envolvidos, se confirma o que os estudos disponíveis revelam, e que se traduz pelo facto de, no tocante à atribuição de valores de género aos nomes, à medida que os aprendentes progredem na aprendizagem do português se verificar uma diminuição gradual dos desvios.

Outra questão em análise prende-se com a relação entre índice de desvio e classe formal do sufixo derivacional. Assim, procura-se verificar, no trabalho empírico, se existe correlação entre a classe temática a que pertencem os sufixos e o processo de assimilação/atribuição de valores de género, visto que estudos anteriores mostram que se observa “uma clara relação entre o sufixo derivacional do nome e respetivos índices de desvios, já que nomes cujos sufixos são menos representados na língua apresentam uma maior incidência de desvios” ([Rio-Torto & Ferreira 2022](#)).

Por fim, tratando-se de línguas da mesma família tipológica, importa apurar se, havendo assimetrias entre os operadores sufixais, estas se traduzem ou não, por transferências positivas entre a L1 dos aprendentes, o espanhol, e a língua-alvo de aprendizagem, o português.

Assim, formulam-se as seguintes questões:

- (1) A incidência de desvios diminui na razão inversa da progressão para o nível mais avançado de aprendizagem?
- (2) Há relação entre incidência de desvios registada nos diferentes grupos de informantes e classe temática do nome?
- (3) Há relação entre índice de desvio e natureza do sufixo usado?
- (4) Há relação entre a frequência do item lexical e o desempenho dos aprendentes na atribuição de género?
- (5) As assimetrias entre a L1 dos aprendentes e a língua-alvo traduzem-se por efeitos de transferência sensíveis?

3. ENQUADRAMENTO TEÓRICO

3.1. Proximidade tipológica e congruência

As classes temáticas dos itens nominais seleccionados para este estudo, de Tema \emptyset ([$\tilde{e}\tilde{w}$]), Tema *-e* (invariável) e Atemática, são as mais opacas em termos de codificação de género

gramatical, pois não apresentam pistas formais que permitam orientar o aprendiz quanto ao género gramatical do nome derivado.

Estudos anteriores levados a cabo sobre universos linguisticamente mais heterogêneos de aprendentes revelam que pode haver correlação entre desvio e classe temática do nome³ (Pinto 2017, Ferreira 2021). Pinto (2017: 100) constatou que aos nomes femininos terminados em *-ão* é tendencialmente associado o valor de género masculino pelos aprendentes de PLE, podendo este comportamento justificar-se pelo facto de os alunos interpretarem o *-o* final como um índice temático. Também Ferreira (2021) verificou que os aprendentes, independentemente da L1, tendem a associar *-o* e *-a* finais aos valores de género masculino e feminino, respetivamente (Ferreira 2021: 223).

Em trabalho precedente, envolvendo os mesmos sufixos, Rio-Torto & Ferreira (2022: 73) concluem que:

no domínio de uma mesma classe temática, como a de nomes de tema \emptyset , os valores percentuais de desvio podem oscilar, em termos globais, entre 50.6% para os derivados em *-ão*, 6.4% para os nomes em *-s[z]ão* e 3.3% para os derivados em *-ção*. O mesmo padrão se aplica a nomes de tema *-e*, para os quais a proporção de desvios se situa nos 3.1% nos nomes em *-idade* e 35.6% nos nomes em *-ice*.

Vamos observar, com detalhe, as classes temáticas em análise, presentes no inquérito elaborado⁴, preenchido por aprendentes tardios de PLE⁵ e que têm o espanhol como língua materna (cf. Quadro 1).

Quadro 1. Classes temáticas e classes de género gramatical dos nomes e sufixos do português presentes no inquérito

Género dos N* sufixados	Classes temáticas dos N*			
	Tema \emptyset ([ẽw̃])	Atemático	Tema <i>-e</i> (invariável)	
Feminino	<i>-ção, -s[z]ão</i>	<i>-agem</i>	<i>-ice</i>	<i>-idade</i>
Masculino	<i>-ão</i>			

*N = Nomes

Como já foi afirmado, as classes temáticas em análise são das mais opacas da língua portuguesa, seja a de Tema \emptyset ([ẽw̃]), a de Tema *-e* (invariável) e a Atemática. A proximidade tipológica e derivacional entre as línguas não impede que haja diferenças entre ambas, quer na marcação de género dos sufixos envolvidos (português: *-agem* (FEM) vs. espanhol: *-aje* (MASC)), quer na classe temática (português: *-idade* vs. espanhol: *-idad*), quer na configuração de alguns sufixos, como *-ice* que não tem correspondência atual em espanhol.

³ Também Leiria (2006: 241-242) constatou que do total de 319 desvios de atribuição de valores de género nominal registados, cerca de 24% recaem apenas em cinco nomes: dois de tema *-e* (*cidade* e *gente*), dois de tema *-a* (*problema* e *pessoa*) e um atemático (*viagem*).

⁴ Gostaríamos de agradecer aos alunos que colaboraram, anonimamente, neste projeto e aos docentes que o integraram na sua atividade letiva.

⁵ Neste trabalho, a expressão “aprendente tardio” é utilizada para designar o falante que inicia o processo de assimilação de uma língua em fases posteriores ao denominado “período crítico” da aquisição linguística (Lenneberg 1967).

Não obstante a elevada cognatividade dos sufixos presentes em ambas as línguas e o carácter internacional de alguns sufixos (português: *-idade*, espanhol: *-idad*; português: *-ção*, espanhol: *-ción*), a verdade é que as peculiaridades associadas a sufixos mais idiossincráticos, como *-ice*, são certamente fonte acrescida de dificuldades no processamento e na aquisição das condições do seu uso (cf. Quadro 2). O conhecimento prévio (Caffarra *et al.* 2015; Sá-Leite *et al.* 2020) aponta para um processamento mais célere em construções menos opacas e mais lento em construções mais opacas, porque menos representadas e mais idiossincráticas. A relevância do processamento da (a)tipicidade formal no que ao género gramatical diz respeito tem sido comprovada através de diferentes modalidades de experiências, seja na produção, seja em juízos de (a)gramaticalidade. Como assinalam Bordag, Opitz & Pechmann (2006: 1090), a propósito das reações de falantes ingleses, aprendentes de alemão como LE: “the L2 word's termination plays a role in L2 gender processing. Participants were fastest when producing gender-marked noun phrases containing a noun with a gender typical termination and slowest when the noun had a gender atypical termination.”

Quadro 2. Sufixos do português presentes no inquérito e equivalentes cognatos em espanhol

Valor de género do nome derivado	Espanhol (L1)		Português (LE)	
	MASC.	FEM.	MASC.	FEM.
Sufixos	<i>-aje</i>		<i>-agem</i>	
	<i>-ícia</i> (versão erudita)		<i>-ice</i>	
	<i>-idad</i>		<i>-idade</i>	
	<i>-ción, -sión</i>		<i>-ção, -s[z]ão</i>	
	<i>-ón</i>		<i>-ão</i>	

É conhecida a controvérsia sobre o “efeito de congruência de género (“gender congruency effect”) na atribuição expectável dos valores de género gramatical, em função da proximidade tipológica dos sistemas de género das L1 e LE em interação, sendo que esse efeito varia em função de fatores diversos, tais como:

the phonological gender transparency of both languages (whether or not they have phonological gender cues associated with the ending letter — e.g., “-a” for feminine words and “-o” for masculine words in Romance languages), the degree of L2 proficiency, and task requirements (naming vs. translation) (Sá-Leite *et al.* 2020: 677).

Existe grande proximidade e paralelismo entre a maior parte dos sufixos do português e do espanhol presentes nas palavras do inquérito aplicado, quer em termos formais, uma vez que o étimo latino é comum a cada par de sufixos (*-ão / -ón*; *-ção, -s[z]ão / -ción, -sión*; *-idade / -idad*), quer no que diz respeito ao valor de género das palavras derivadas com cada um destes sufixos, dado que o género gramatical é comum ao espanhol e ao português em diversos pares de sufixos enunciados (cf. Quadro 2). As grandes diferenças residem no par *-agem* e *-aje*, pois em português os nomes em *-agem* são marcados como FEM e em espanhol os derivados em *-aje* são associados ao MASC. No que diz respeito ao sufixo *-ice*, este não tem paralelo no espanhol, língua na qual se regista apenas a variante erudita *-ícia*, também presente na língua portuguesa mais antiga, mas não aqui explorada.

O sufixo *-agem*, que remonta ao latim *-Ārīcu-*, terá sido importado do provençal, do catalão *-atge* e/ou do francês *-age* (Pharies 2002: 52-53). Os empréstimos no castelhano e no

português começam a fazer-se sentir desde cedo (séc. XIII), apresentando o género MASC em espanhol. No português, após um período de oscilação, fixou-se o género FEM, segundo alguns autores talvez por contaminação com o FEM *-agem* de valor coletivo (a *criadagem*, a *folhagem*). A autonomização do sufixo *-aje* dos radicais estrangeiros teve lugar no espanhol a partir do século XVII. A influência do género MASC que o cognato *-aje* tem na gramática mental prévia dos aprendentes tardios espanhóis faz-se certamente sentir no seu desempenho aquando da atribuição de género aos derivados em *-agem* do português. Como tal, a hipótese de uma interferência na interlíngua tem grande robustez. Como afirmam [Bordag & Pechmann \(2007: 310-311\)](#):

gender-interference effect from L1 is a robust effect which can be replicated under various conditions. [...] the L2 gender production is affected not only by the L1 grammatical gender, but also by the phonological form of the L2 noun, especially by its termination: Subjects had the least difficulty with nouns with a gender-typical termination and the most difficulty with nouns with a gender-atypical termination. These results speak in favor of a model, which allows interaction between the levels of phonological encoding and grammatical encoding, e.g., the Interactive Activation model of Dell (1986) or, in this respect, the Independent Network model of Caramazza (1997).

O sufixo português *-ice* remonta ao latim *-ITĪA* ([Meyer-Lübke 1895: 567-569](#)) ou *-ITĪES* ([Diez 1874: 336](#)), mas a configuração *-ice* é hoje uma idiosincrasia da língua portuguesa, uma vez que o italiano *-izia* (*avarizia*), o francês *-ice* (*avarice*) e o espanhol *-ícia* (*avaricia*) não vingaram como produtivos nas respetivas línguas, o que acontece com o português, como [Diez \(1874: 336\)](#) já assinala. O mesmo acontece com as configurações eruditas do português *-ícia*, presente em nomes antigos (*avaricia*, *blandícia*, *imundícia*, *puerícia*, *vespícia*) e *-ície* (*calvície*). Por isso, alguns autores admitem ([Meyer-Lübke 1895: 569](#)) que em português *-ice* tenha sofrido influência eventual do francês *-ise* (*bêtise*, *couardise*, *franchise*, *gourmandise*).

3.2. O papel dos conhecimentos linguísticos prévios

No âmbito da investigação sobre a aquisição/aprendizagem tardia⁶ da categoria de género gramatical, diferentes trabalhos têm procurado compreender de que forma a configuração do conhecimento linguístico prévio dos aprendentes, proveniente da L1 ou de LE anteriormente conhecidas, atua ao longo das distintas fases de assimilação desta estrutura gramatical ([White et al. 2004](#); [Franceschina 2005](#); [Sabourin, Stowe & Haan 2006](#); [Martins 2015](#); [Lacsán 2016](#); [Pinto 2017](#); [Ferreira 2019](#)).

A transferência linguística é um dos processos que atuam no âmbito da construção dos sistemas interlinguísticos dos aprendentes ([Selinker 1972: 216](#)), e que muito contribui para a formulação de hipóteses sobre o *input* linguístico durante o processo de aquisição/aprendizagem de uma LE. Por se tratar de um processo mental que atua ao longo das distintas fases do processo de aquisição, a transferência linguística apresenta quer efeitos positivos, quer efeitos negativos ([Alonso Alonso 2020](#)). Por um lado, os conhecimentos linguísticos prévios podem conduzir à ocorrência de desvios, tratando-se, assim, de um típico fenómeno de transferência negativa ou de interferência. Em contrapartida, o conhecimento linguístico prévio pode também constituir um fator positivo na medida em que auxilia a assimilação das estruturas da língua-alvo de aprendizagem (cf. [Alonso Alonso 1999](#); [White 2003](#); [Ellis 2000](#)).

⁶ A expressão “aquisição/aprendizagem tardia” designa o processo de assimilação linguística que se inicia em fases posteriores ao denominado “período crítico” em [Lenneberg \(1967\)](#).

Fatores diversos podem contribuir para a influência linguística ao longo do processo de assimilação das estruturas de uma LE, entre os quais se destaca a proximidade tipológica dos sistemas, i.e., da L1 e da língua-alvo de aprendizagem (Alonso Alonso 1999; Corder 1993; Schachter 1993), assumindo-se que uma maior proximidade estrutural potencia a ocorrência de fenómenos típicos de transferência linguística. As questões relativas à proximidade tipológica têm sido especialmente relevantes para os estudos sobre a aprendizagem de uma LE por falantes multilingues, sugerindo-se que os aprendentes não transferem, necessariamente, as propriedades linguísticas do seu idioma nativo, mas de outros sistemas previamente conhecidos que sejam tipologicamente mais próximos dos da língua-alvo (veja-se, por exemplo, os trabalhos de DeAngelis 2005; Sabourin, Stowe & Haan 2006 e, para o português, os estudos de Lacsán 2016 e Pinto 2017).

Em trabalhos relativos à aquisição/aprendizagem de PLE (Leiria 2006; Lacsán 2016; Martins 2015; Pinto 2017; Ferreira 2019) verifica-se que, mesmo nos níveis mais avançados, os desvios nunca chegam a ser erradicados por completo das produções dos aprendentes, independentemente de terem representada na sua L1 a categoria gramatical de género (cf. Leiria 2006: 241; Martins 2015: 44; Lacsán 2016: 122). No estudo de Ferreira (2019), procedeu-se à análise de desvios de atribuição e de concordância nominal de género em produções escritas por aprendentes, falantes nativos de espanhol, italiano, alemão, inglês e chinês. Os resultados deste trabalho são reveladores de que, nos níveis iniciais, a proporção de desvios é consideravelmente maior nas produções dos aprendentes cuja L1 não possui um sistema de género gramatical. Por sua vez, nos níveis mais avançados, registaram-se incidências de desvios semelhantes em todos os segmentos da amostra. Além disso, entre os falantes nativos de espanhol, verificou-se que, embora os índices de desvio sejam baixos em todos os níveis de proficiência, no nível mais avançado, C1, não se observa uma considerável melhoria do desempenho linguístico destes informantes. A autora conclui que emerge, desde cedo, um certo efeito de estabilização da aprendizagem neste grupo de aprendentes e “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 L[íngua]-A[lvo]” (Ferreira 2019: 292). Por conseguinte, infere-se que além dos conhecimentos linguísticos prévios, outros fatores do próprio desenvolvimento interlinguístico terão igualmente um papel determinante para a assimilação dos valores de género nominal em PLE.

Também no estudo de Martins (2015) verificou-se que o facto de a língua materna do falante aprendente exibir um sistema de género não auxilia a assimilação de valores de género nominal em português. Os resultados apurados neste trabalho são reveladores de que os falantes nativos do chinês, idioma sem a categoria de género gramatical, registam um melhor desempenho global face ao registado nos restantes grupos de informantes, falantes nativos de alemão e de espanhol. Por fim, e nas palavras da autora, é possível inferir nos dados apurados um certo efeito de transferência negativa “não tanto direta de valores de género da L[íngua] M[aterna] para a L[íngua]-A[lvo] em vocábulos semanticamente equivalentes (...), mas mais propriamente sob a forma de uma atitude de insegurança por parte do aprendente nestas circunstâncias” (Martins 2015: 44).

4. PROCEDIMENTOS METODOLÓGICOS

4.1. Informantes

A base empírica deste estudo é composta por dados obtidos através da aplicação de um questionário *online*⁷ junto de 127 informantes, aprendentes tardios de PLE. O inquérito foi aplicado em turmas de aprendentes adultos, a frequentar aulas de língua portuguesa em instituições de ensino em Portugal, nomeadamente na Universidade de Coimbra (UC), e fora de Portugal, em Espanha, na Escola Oficial de Idiomas de Cáceres (EOI), na Universidade da Estremadura (UEX) e na Universidade de Alcalá e Henares (UAH). Tendo em conta os objetivos do presente estudo, selecionaram-se para análise os inquéritos submetidos pelos informantes, falantes nativos de espanhol⁸.

Os participantes têm idades maioritariamente compreendidas entre os 20 e os 25 anos, havendo apenas um falante com 73 anos e outro com 16. A amplitude de idades representada neste conjunto é justificável pelo facto de terem sido testados informantes inseridos em distintos contextos de formação linguística. Tipicamente, os aprendentes que frequentam cursos de português nas universidades fazem-no na continuidade da sua formação académica superior, sendo a cadeira de português uma unidade curricular obrigatória no seu plano de estudos. Um outro subconjunto de aprendentes frequenta cursos de português no âmbito de uma formação livre, não inserida, portanto, nos *curricula* de cursos académicos conferentes de grau. A média (\bar{X}) de idades dos dois grupos é substancialmente diferente, sendo de 20 anos para informantes inseridos no contexto universitário e de 44 anos para os que frequentam aulas de português num contexto não universitário (cf. Quadro 3). Não obstante, uma vez que o universo dominante de aprendentes é largamente homogéneo em termos etários, o nível etário dos informantes não foi, todavia, considerado como uma variável para este estudo empírico.

Quadro 3. Distribuição dos informantes por contexto de formação em língua portuguesa e média de idades

Contexto da formação em língua portuguesa	n	Idade \bar{X}
Universitário: UC / UEX / UAH	52	20 (19-73)
Não Universitário: EOI	75	44 (16-70)

Os informantes encontram-se distribuídos em função do nível de proficiência linguística da turma frequentada, definido de acordo com o *Quadro Europeu Comum de Referência para as Línguas* (QECRL) (Conselho da Europa 2001). Obtiveram-se, assim, respostas de três grupos distintos: um primeiro que compreende os aprendentes dos níveis de iniciação, que frequentam turmas de nível A1 e turmas de nível A2, um segundo que integra aprendentes dos

⁷ O inquérito foi elaborado através da plataforma Microsoft Forms®, tendo sido o respetivo *link* enviado aos informantes por *email*.

⁸ Por questões de natureza metodológica, selecionaram-se apenas os inquéritos de informantes que assinalaram como L1 o espanhol, tendo-se excluído da análise os inquéritos submetidos por aprendentes bilingues, i.e., informantes que, além do espanhol, referiram ser falantes nativos de outro idioma. Registaram-se, com efeito, submissões de inquéritos de falantes nativos de: espanhol e asturiano (n=1); espanhol e basco (n=4); espanhol e catalão (n=3) e espanhol e francês (n=1). A opção de exclusão destes inquéritos prende-se com o facto de o número de falantes bilingues ser reduzido e, portanto, a sua fraca representatividade não permitir uma comparação robusta relativamente ao desempenho observado entre os falantes monolingues e falantes bilingues.

níveis intermédios, B1-B2, e um último com aprendentes a frequentar turmas de níveis mais avançados, C1-C2. Os dados foram recolhidos uma só vez, em momentos únicos, e não em momentos diferenciados e sequenciais, pelo que não foi efetuado um estudo cronologicamente longitudinal, ou seja, não se analisaram produções de diferentes ou sucessivos intervalos periódicos. A recolha foi efetuada junto de aprendentes que frequentavam turmas de vários níveis de competência e de proficiência linguísticas em português, tratando-se, portanto, de amostras *cross-sectional*.

O **Quadro 4** apresenta informação relativa à distribuição dos informantes por níveis de proficiência (nível QECRL), bem como o respetivo tempo médio de resposta aos inquéritos. Dada a significativa homogeneidade dos tempos de resposta nos diferentes níveis de proficiência, a variável ‘tempo médio de resposta’ não foi considerada neste estudo.

Quadro 4. Distribuição do número de informantes por nível QECRL e tempo médio de resposta ao inquérito

Nível QECRL	N.º de informantes	Tempo médio de resposta hh:mm:ss
A1-A2	77	00:12:44
B1-B2	44	00:11:53
C1-C2	6	00:12:55
Σ	127	00:12:27

Como se pode constatar da leitura do **Quadro 4**, os 127 informantes não se distribuem equitativamente pelos vários níveis de proficiência. Com efeito, regista-se uma maior concentração de aprendentes nos níveis de iniciação por oposição ao registado nos níveis mais avançados. No que diz respeito ao tempo médio de resposta por inquérito registado em cada grupo, não se observam diferenças consideráveis. Tendo em conta que os inquéritos foram preenchidos em aula e sem recurso a material de apoio (dicionário ou outros recursos), seria expectável que os informantes dos níveis iniciais precisassem de mais tempo, visto que algumas construções frásicas são extensas e com léxico mais complexo. O facto de neste conjunto (A1-A2) de informantes não se ter verificado um maior tempo de resposta poderá estar correlacionado com a proximidade tipológica da L1 dos sujeitos, o espanhol, e a língua-alvo de aprendizagem, o português.

Tal como se explicita em 4.3., as variáveis consideradas no estudo empírico foram o nível de proficiência e de competência linguísticas das turmas frequentadas pelos informantes, as classes temáticas dos nomes sufixados e o valor de frequência lexical do nome.

4.2. Inquérito

Solicitou-se aos informantes que identificassem o valor de género gramatical de 30 itens nominais sufixados (22 formas atestadas e 8 formas possíveis⁹), distribuídos equitativamente por 6 grupos estipulados em função do sufixo derivacional: *-ão*, *-ção*, *-s[z]ão*, *-agem*, *-ice* e *-idade* (cf. **Quadro 5**).

⁹ Uma forma atestada é uma forma documentada em produtos lexicais da língua. Por sua vez, uma forma possível corresponde a uma palavra que não consta dos reportórios lexicais de uma língua (no português, por exemplo, “meigosa”), mas é construída de acordo com os princípios que regulam morfofonologicamente essa língua (cf. [Rodrigues 2016: 80-87](#)).

Quadro 5. Distribuição das formas nominais aplicadas no inquérito

Valor de género		MASCULINO			FEMININO		
		Sufixos	-ão	-ção	-s[z]ão	-agem	-ice
Derivados	atestados	<i>abanão</i> <i>arrastão</i> <i>empurrão</i> <i>puxão</i>	<i>confraternização</i> <i>evolução</i> <i>preocupação</i> <i>satisfação</i>	<i>alusão</i> <i>confusão</i> <i>decisão</i> <i>revisão</i>	<i>abordagem</i> <i>lavagem</i> <i>passagem</i>	<i>criancice</i> <i>velhice</i> <i>vigarice</i>	<i>cumplicidade</i> <i>exclusividade</i> <i>falsidade</i> <i>finalidade</i>
	possíveis	<i>resfrião</i>	<i>avisação</i>	<i>precisão</i>	<i>levantagem</i> <i>tratagem</i>	<i>contentice</i> <i>vitalice</i>	<i>prontidade</i>

No inquérito, os aprendentes tinham de seleccionar, em cada construção frásica apresentada, a forma adequada do artigo definido (*o, a, os, as*) que concorda em género e em número com o respetivo nome sufixado. Atendendo aos sufixos derivacionais seleccionados para este estudo, somente os nomes sufixados em *-ão* são marcados como masculinos e as restantes formas sufixadas são marcadas como femininas. Deste modo, criaram-se construções frásicas que integram itens nominais derivados, quer na forma singular quer na forma plural, em *-ão* (1), em *-ção* (2), em *-s[z]ão* (3), em *-agem* (4), em *-ice* (5) e em *-idade* (6).

- (1) Ouvir esta palestra foi (o/a/os/as) *empurrão* de que precisava para iniciar o meu próprio negócio.
- (2) Muitos pais não são capazes de esconder (o/a/os/as) *preocupações* que têm com o futuro dos filhos.
- (3) O orientador ainda não enviou (o/a/os/as) *revisões* finais do relatório de estágio ao Simão.
- (4) Este detergente é ótimo para (o/a/os/as) *lavagem* de cobertores e tapetes.
- (5) Muitas pessoas preparam-se, cada vez mais cedo, para (o/a/os/as) *velhice*.
- (6) Qual é (o/a/os/as) *finalidade* do projeto proposto hoje na Assembleia Geral?

Tendo em conta os objetivos do presente estudo (cf. Secção 2.), consideraram-se apenas as respostas relativas à atribuição do valor de género nominal, tendo-se optado por não analisar as respostas que envolvem a atribuição do valor de número.

4.3. Tratamento dos dados

Os dados obtidos do inquérito foram inseridos e codificados numa base de dados com vista ao seu tratamento e análise estatística. Registaram-se, por inquérito submetido, as ocorrências desviantes, i.e., os casos em que a seleção da forma do artigo na frase não correspondia ao valor de género do item nominal com o qual se relacionava. Deste modo, para aferir padrões de atribuição de valores de género nominal em diferentes fases do desenvolvimento interlinguístico, analisaram-se os valores das médias (em percentagens) das respostas desviantes dadas por segmento da amostra, em função do nível QECRL dos informantes e atendendo às seguintes variáveis linguísticas: (i) classe temática e sufixo derivacional do item; e (iii) valor da frequência lexical do nome.

O **Quadro 6** apresenta a distribuição do número absoluto de respostas obtidas (R) e de desvios (D), distribuídos em função (i) do nível QECRL da turma frequentada pelos informantes, (ii) da classe temática e (iii) do respetivo sufixo derivacional dos nomes presentes no inquérito. Assim, do conjunto dos 127 inquéritos submetidos, apurou-se um total de 3810 ocorrências de atribuição de valores de género, tendo-se assinalado 751 ocorrências desviantes.

Quadro 6. Distribuição do número de respostas e de desvios por afixos e nível QECRL

Nível QECRL	Classe temática do nome sufixado						Σ R/D
	Tema \emptyset ([ēw̃])			Atemático	Tema -e (invariável)		
	-ão R/D*	-ção R/D	-s[z]ão R/D	-agem R/D	-ice R/D	-idade R/D	
A1-A2(n=77)	385/85	385/56	385/61	385/222	385/111	385/20	2310/555
B1-B2 (n=44)	220/39	220/18	220 / 9	220/69	220/50	220/5	1320/190
C1-C2 (n=6)	30/2	30/0	30/1	30/0	30/2	30 /0	180/ 6
Σ	635/126	635/74	635/71	635/291	635 /163	635/25	3810/ 751

* R (respostas) / D (desvios)

5. RESULTADOS E DISCUSSÃO

Em primeiro lugar, considerem-se os valores das médias de ocorrências totais desviantes registadas por segmentos da amostra empírica em estudo.

Quadro 7. Médias (em percentagens) das ocorrências desviantes totais apuradas por grupos de informantes

Nível QECRL	Número absoluto de respostas	Número absoluto de desvios	Média de desvios (%)
A1-A2*	2310	555	24.03
B1-B2**	1320	190	14.39
C1-C2***	180	6	3.33
Σ	3810	751	19.71

* (n=77)

** (n=44)

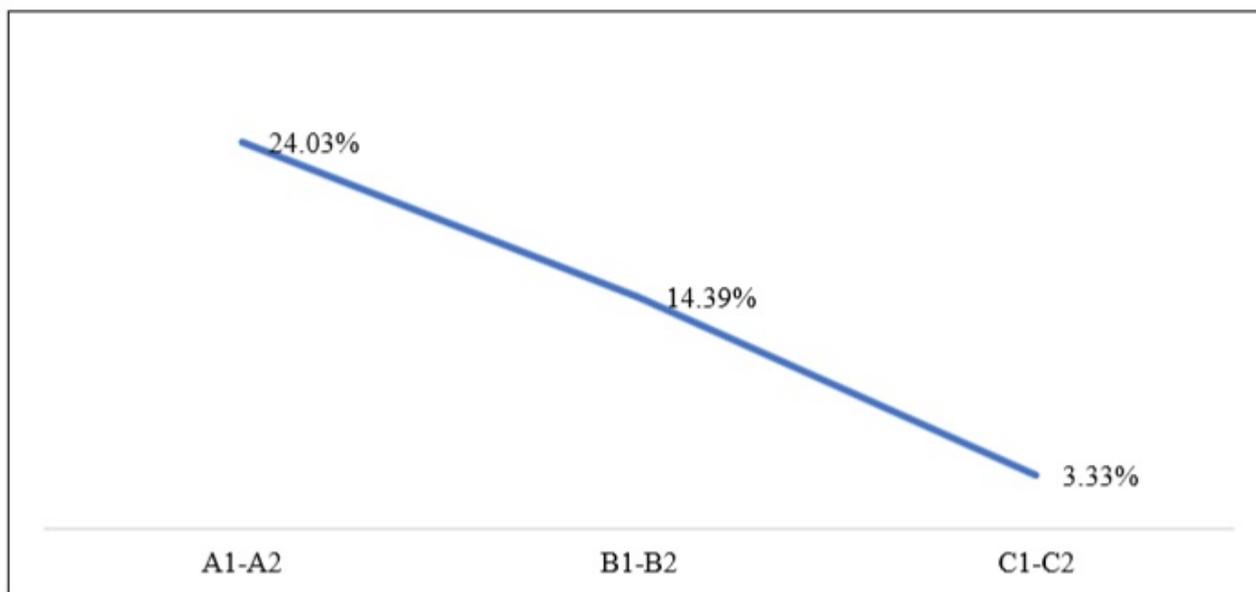
*** (n=6)

Em função do número absoluto de respostas obtido, constata-se que cerca de 20% correspondem à atribuição desviante de valor de género nominal. Assim, e de um modo geral, verifica-se uma taxa de acertos considerável nestes aprendentes (entre 76% a 96%).

No que concerne à distribuição total de desvios por níveis de proficiência, constata-se que os níveis de iniciação (A1-A2) concentram a maior proporção da média desvios (24.03%). Na passagem para os níveis intermédios, observa-se uma descida da proporção de desvios (14.40%) e os níveis mais avançados (C1-C2) apresentam valores da média de desvios residuais (3.33%) verificando-se, assim, e como seria expectável, uma considerável melhoria

no desempenho destes aprendentes relativamente à associação de valores de género aos nomes sufixados (cf. Gráfico 1).

Gráfico 1. Distribuição dos valores das médias globais de desvios apurados por níveis de proficiência



Com base nestes resultados constata-se ainda que as maiores diferenças de desempenho se observam entre os níveis de iniciação (A1-A2) e os níveis mais avançados (C1-C2). Assim, quanto mais baixo é o nível, maior é a proporção global de desvios. Portanto, estes dados indicam que ao longo da aprendizagem do português, os aprendentes, falantes nativos de espanhol, vão assimilando adequadamente o valor de género associado ao item nominal sufixado, estando ainda em consonância com os resultados reportados em outros estudos (cf. White *et al.* 2004: 127; Isabelli-García 2010: 299; Martins 2015: 42; Lacsán 2016: 83; Ferreira 2019: 199).

Em seguida, analisa-se o desempenho linguístico dos diferentes grupos de aprendentes em função da classe temática dos itens nominais afetados. O Quadro 8 apresenta os valores das médias de desvios (em percentagens) obtidos em função da classe temática.

Quadro 8. Médias (em percentagens) das ocorrências desviantes apuradas em cada segmento da amostra por classes temáticas

Nível QECRL	Classe temática do nome e sufixo		
	Tema \emptyset ([ẽw]) -ão, -ção, -s[z]ão	Atemática -agem	Tema -e (invariável) -ice, -idade
A1-A2 (n=77)	17.49	57.66	17.01
B1-B2 (n=44)	10.00	31.36	12.50
C1-C2 (n=6)	3.33	3.33	3.33
\bar{X}	10.27	30.78	10.95

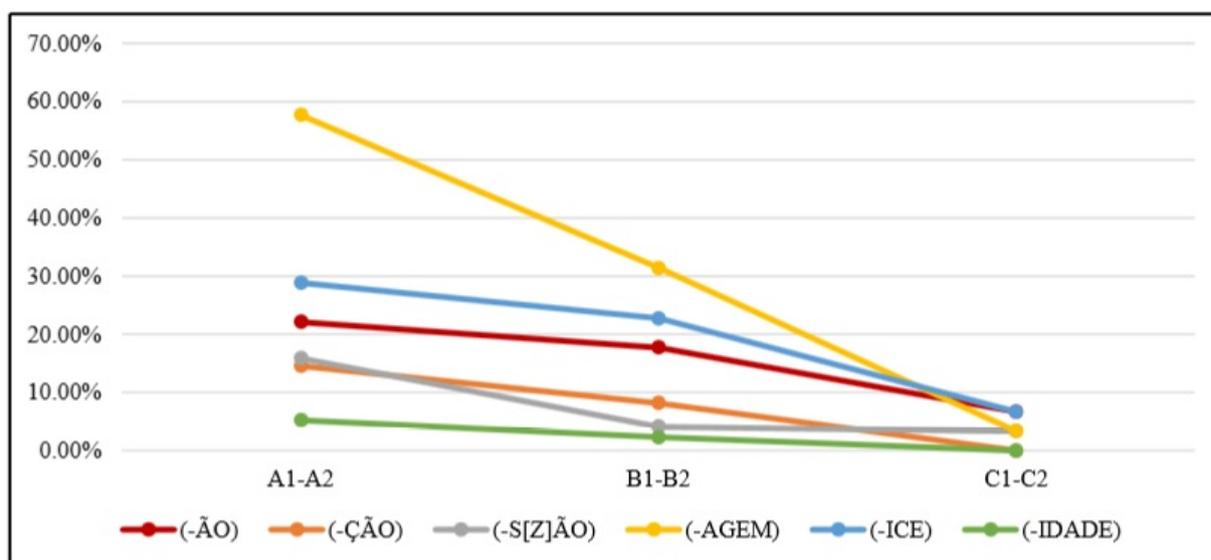
Partindo da análise das médias (em percentagens) das ocorrências desviantes (Quadro 8), verifica-se que nos níveis mais avançados (C1-C2) não há diferenças significativas de desempenho em função da classe temática do nome, dado que em todas se apurou o mesmo valor da média de desvios (3.33%). Em contrapartida, nos níveis de iniciação (A1-A2) e

intermédios (B1-B2), regista-se uma incidência significativa de desvios (57.66% e 31.36%) na classe dos nomes atemáticos, i.e., dos sufixados em *-agem*. Nas restantes classes temáticas, os valores das médias de desvios são mais próximos, mas evidenciam, em todo o caso, pequenas diferenças de desempenho nos dois segmentos (A1-A2 vs. B1-B2). Com efeito, nos níveis de iniciação o valor da média de desvios em nomes da classe de Tema \emptyset ([ẽw]) é praticamente idêntico (17.49%) ao registado entre os nomes de Tema *-e* (invariável) (17.01%). Já nos níveis intermédios, a incidência de desvios nos nomes de Tema *-e* (invariável) (12.50%) é ligeiramente superior à registada nos nomes de Tema \emptyset ([ẽw]) (10.00%). De forma a compreender estas diferenças de desempenho, é preciso analisar as incidências de desvios atendendo a outras variáveis, nomeadamente o sufixo derivacional e a frequência do *input* do item lexical.

Quadro 9. Médias (em percentagens) das ocorrências desviantes apuradas em cada segmento da amostra por sufixos

Nível QECRL	Sufixo derivacional					
	<i>-ão</i>	<i>-ção</i>	<i>-s[z]ão</i>	<i>-agem</i>	<i>-ice</i>	<i>-idade</i>
A1-A2 (n=77)	22.07	14.55	15.84	57.66	28.83	5.19
B1-B2 (n=44)	17.72	8.18	4.09	31.36	22.72	2.27
C1-C2 (n=6)	6.67	.00	3.33	3.33	6.67	.00
\bar{X}	19.84	11.65	11.18	45.98	25.67	3.94

Gráfico 2. Distribuição das médias percentuais de desvios por sufixo derivacional apuradas em função do nível QECRL dos informantes



Considerando o desempenho dos informantes por nível QECRL, nos níveis de iniciação A1-A2 a maior incidência de desvios recai nos nomes sufixados em *-agem* (57.66%), *-ice* (28.83%) e *-ão* (22.07%) (cf. Quadro 9 e Gráfico 2). Também nos níveis intermédios se verifica a mesma tendência, ou seja, há uma maior frequência de desvios em nomes sufixados em *-agem* (31.36%), *-ice* (22.72%) e *-ão* (17.72%). Já nos níveis C1-C2, a maior incidência de desvios regista-se nos sufixados em *-ão* e *-ice* (ambos com 6.67%). Os nomes sufixados em

-agem e em *-s[z]ão* apresentam a mesma proporção de desvios (3.33%) e os em *-ção* e *-idade* registam .00% de desvios.

O [quadro seguinte](#) procura explicitar a gradiência existente entre os três conjuntos de sufixos aqui analisados, a qual é mais acentuada nos níveis de proficiência mais baixos.

Quadro 10. Valores absolutos (em percentagens) das ocorrências desviantes de cada sufixo apuradas por nível QECRL

Nível QECRL	← Proporção mais elevada de desvio			→ Proporção mais baixa de desvio		
	A1-A2	-agem (57.66%)	-ice (28.83%)	-ão (22.07%)	-s[z]ão (15.84%)	-ção (14.55%)
B1-B2	-agem (31.36%)	-ice (22.72%)	-ão (17.72%)	-ção (8.18%)	-s[z]ão (4.09%)	-idade (2.27%)
C1-C2	-ão (6.67%)	-ice (6.67%)	-agem (57.66%)			-ção (.00%)
			-s[z]ão (3.33%)			-idade (.00%)

Pela análise do [Quadro 10](#), constata-se que, em termos globais, a maior proporção de desvios se concentra nos itens sufixados em *-agem* (de 31.36% a 57.66%), em *-ice* (de 22.72% a 28.83%) e em *-ão* (de 17.72% a 22.07%). Por oposição, os nomes sufixados em *-s[z]ão*, *-ção* e *-idade* apresentam índices de desvios globalmente mais baixos: *-ção* e *-s[z]ão* têm proporções de desvios globais de .00% (no nível C1-C2) a 11.65%, e de 3.33% a 15.84% respetivamente, e o sufixo *-idade* regista o valor percentual global de desvios mais baixo, de .00% a 5.19%. Assim, verifica-se, no desempenho destes aprendentes, uma escalaridade decrescente nos desvios, que é transversal aos níveis QECRL A1-A2 e B1-B2, e que envolve a classe dos nomes:

- (i) atemáticos (sufixados em *-agem*);
- (ii) de Tema *-e* (invariável) e sufixados em *-ice*;
- (iii) de Tema \emptyset [ẽw], sufixados em *-s[z]ão* e *-ção* (alternando as posições destes nos valores de desvio registados nos níveis QECRL A1-A2 e B1-B2); e
- (iv) de Tema *-e* (invariável), sufixados em *-idade*.

A classe dos nomes atemáticos é, de longe, a mais propensa a dificuldades na atribuição de género gramatical. Não obstante, dentro de uma mesma classe temática, as características de cada sufixo são prevalentes no tocante à proporção de desvios. A classe dos nomes de Tema *-e* revela que não existe uma correlação direta entre classe temática do nome e as ocorrências de desvios, havendo, em contrapartida, padrões distintos em função do sufixo derivacional. O contraste entre o valor da média global de desvios registado nos sufixados em *-ice* (25.67%) e em *-idade* (3.94%), ambos de Tema *-e* (invariável), é revelador de que o grau de idiomatismo e/ou de representatividade do sufixo na língua e/ou no *input* ao qual o aprendente está exposto têm um papel relevante no índice de desvios encontrado. A polaridade mais acentuada situa-se, assim, entre *-agem* e *-ice*, por um lado, e os exemplares *-ção* e *-idade*, por outro: sendo sufixos de classes temáticas nada transparentes e canónicas no tocante à atribuição de género, os dois primeiros contrapõem-se aos dois últimos no (in)acerto de atribuição de valor de género, uma vez que são marcados por graus diversos de

representatividade na língua e certamente também no *input* a que os aprendentes estão expostos.

Na classe dos atemáticos explorados, o facto de *-agem* ser cognato de *-aje*, com valor de género antagónico e de classe temática diferente, mostra, pelos desvios verificados (45.98%), que o peso da transferência da L1 na associação de valores de géneros aos nomes em português é enorme, corroborando, assim, o chamado “gender congruency effect” (Sá-Leite *et al.* 2020) na atribuição expectável dos valores de género gramatical, em função da proximidade tipológica dos sistemas de género das L1 e LE em interação. Em espanhol, por sua vez, os nomes terminados em *-aje* como *viaje*, *abordaje* e *aprendizaje*, por exemplo, são de género masculino, e em português as formas nominais correspondentes, *viagem*, *abordagem* e *aprendizagem*, são marcadas como femininas. Portanto, a proximidade formal entre os vocábulos poderá, assim, contribuir para que os aprendentes, sobretudo em fases iniciais da aprendizagem, assumam que os nomes formalmente idênticos possuam o mesmo valor de género nominal. No caso deste sufixo, o efeito da proximidade formal dos idiomas na atribuição de valores de género já não é tão expressivo em fases mais avançadas da aprendizagem, já que nos níveis C1-C2 apenas se regista a atribuição de valor de género desviante a um item lexical sufixado em *-agem*. Estes resultados vão ainda ao encontro dos reportados por Lacsán (2016: 90), em que se registou uma incidência de desvios de atribuição de valor de género a nomes deverbiais sufixados em *-agem* nos inquéritos recolhidos junto do grupo de aprendentes de PLE, de nível A2, que são falantes nativos de húngaro com conhecimentos prévios de uma língua românica, espanhol e italiano. Em contrapartida, nos níveis mais avançados, a taxa de acertos apurada é elevada, sugerindo-se, portanto, que na assimilação dos valores de género nominal, sobretudo em fases iniciais, os aprendentes transferem as propriedades de outras línguas previamente conhecidas, tipologicamente mais próximas da língua-alvo de aprendizagem.

No que concerne aos nomes sufixados em *-ice*, marcados como femininos, a incidência de desvios dever-se-á ao facto de a configuração de *-ice* fazer dele um sufixo autóctone e idiossincrático da língua-alvo (cf. Secção 3.1.). Ademais, a sua fraca representatividade no *input* a que estão expostos poderá contribuir para que os informantes desconheçam o valor de género do nome e optem, assim, por marcá-los como nomes masculinos. O recurso a uma estratégia de marcação *default* da forma masculina, forma não-marcada, é, aliás, uma tendência observada em outros estudos sobre a aquisição de género gramatical por aprendentes tardios¹⁰ (Bruhn de Garavito & White 2002: 170; White *et al.* 2004: 128).

Já entre os nomes pertencentes à classe temática \emptyset ([ẽw]), apenas os nomes sufixados em *-ão* são marcados, na língua portuguesa, como masculinos, e aos restantes itens sufixados em *-ção* e *-s[z]ão* está associado o valor de género feminino. Nestes casos, a presença de um *-o* final poderia indiciar ao aprendente que se trata de formas de género masculino, segundo, aliás, a tendência verificada em outros estudos empíricos sobre associação de valor de género (cf. Pinto 2017; Ferreira 2019), mas, na verdade, a tendência é a de associar a todos estes nomes sufixados o valor de género feminino. O comportamento desviante aqui assinalado é ainda mais surpreendente se atentarmos ao facto de que existe uma grande proximidade entre

¹⁰ Partindo da análise de dados relativos à concordância nominal por aprendentes de espanhol como L2, Bruhn de Garavito & White (2002: 170) registam que “(...) when feminine gender agreement is found it is largely appropriate: given a feminine article or a feminine adjective, the noun will be feminine. It is masculine that is overgeneralized, such that masculine forms of determiners and adjectives can occur with feminine nouns (though correct agreement is predominant) (...) this reflects the fact that masculine is the default in the domain of gender in Spanish (cf. Harris 1991), just as infinitives often serve as defaults in the domain of tense”.

estes sufixos no português e no espanhol, quer no que respeita à forma quer no que concerne aos valores de género associados (cf. Quadro 2). O desempenho destes informantes pode estar, assim, correlacionado com a representatividade dos sufixos no *input* a que estão expostos. Como é conhecido, *-ão* é um sufixo menos representado do que *-ção* e *-s[z]ão* (Rio-Torto & Rodrigues 2016) e, além disso, existem na língua portuguesa nomes terminados em *-ão* marcados como FEM (a *mão*, a *paixão*, a *opinião*, a *união*) o que poderá levar os aprendentes, sobretudo em fases iniciais da aprendizagem, a atribuir o mesmo valor de género a itens formalmente idênticos.

Por fim, o sufixo *-idade* regista, em todos os segmentos, os valores de incidências de desvios mais baixos. O facto de se tratar de um sufixo de circulação internacional e de haver, em espanhol, nomes formalmente idênticos (terminados em *-dad*) e marcados como femininos pode contribuir para um efeito de transferência direta de valores de género da L1 para a língua-alvo de aprendizagem.

Consideremos, agora, a proporção de desvios por itens, tendo em conta os valores de frequência lexical das 22 formas atestadas (cf. Quadro 11). Tomou-se como referência os valores globais de frequência por milhão registados no *Corpus do Português – Web / Dialects* (CdP)¹¹ de Mark Davies, nas variedades europeia e brasileira do português.

Quadro 11. Valores de frequência dos nomes por milhão registados no *Corpus do Português* (CdP) de Mark Davies

	Nomes	Valores de frequência por milhão (CdP)
+ frequente	<i>evolução</i>	453.6
	<i>decisão</i>	194.34
	<i>passagem</i>	88.13
	<i>preocupação</i>	67.21
	<i>abordagem</i>	45.10
	<i>satisfação</i>	44.66
	<i>finalidade</i>	42.74
	<i>revisão</i>	38.39
	<i>confusão</i>	37.33
	<i>lavagem</i>	13.27
	<i>velhice</i>	9.50
	<i>exclusividade</i>	8.57
	<i>falsidade</i>	8.05
	<i>alusão</i>	7.07
	<i>cumplicidade</i>	6.83
	<i>confraternização</i>	2.11
	<i>pução</i>	1.62

¹¹ O *Corpus do Português – Web / Dialects* (CdP) (disponível para consulta mediante registo prévio em <https://www.corpusdoportugues.org/web-dial>) de Mark Davies integra, aproximadamente, mil milhões de palavras recolhidas em cerca de um milhão de páginas *web* de quatro países de Língua Oficial Portuguesa, a saber: Portugal, Brasil, Angola e Moçambique. Neste *corpus*, é possível pesquisar por palavras e aceder aos respetivos índices de frequência lexical por item.

	<i>empurrão</i>	1.54
	<i>arrastão</i>	0.82
	<i>vigarice</i>	0.59
	<i>abanão</i>	0.20
—frequente	<i>criancice</i>	0.12

A partir da distribuição dos itens por frequência lexical, constata-se que, entre os nomes da classe temática \emptyset ([ẽw]), todos os sufixados em *-ão* apresentam índices de frequência baixos, inferiores a dois mil por milhão (*puxão* (1.62), *empurrão* (1.54), *arrastão* (0.82) e *abanão* (0.20)). Já os sufixados em *-ção* e *-s[z]ão* registam frequências lexicais superiores. Com efeito, e à exceção de *confraternização*, cujo valor de frequência lexical é ligeiramente superior a dois mil por milhão (2.11), as restantes formas sufixadas em *-ção* têm valores de frequência superiores a 40 mil por milhão (*evolução* (453.6); *preocupação* (67.21) e *satisfação* (44.66)). Entre os sufixados em *-s[z]ão*, excetuando o nome *alusão*, cujo valor de frequência é de, aproximadamente, sete mil por milhão, os restantes itens detêm valores de frequência superiores a 30 mil por milhão (*decisão* (194.34), *revisão* (38.39) e *confusão* (37.33)). Atendendo aos valores de frequência lexical registados por itens sufixados confirma-se, por via indireta, a fraca representatividade do sufixo *-ão* face aos sufixos *-ção* e *-s[z]ão*. Relativamente aos nomes atemáticos, sufixados em *-agem*, constata-se que todas as formas atestadas apresentam valores de frequência na língua superiores a 10 mil por milhão (*passagem* (88.13), *abordagem* (45.1) e *lavagem* (13.27)) e, por conseguinte, poder-se-á afirmar que, em termos absolutos, este sufixo apresenta uma forte representatividade no *input*. Por fim, nos nomes de tema em *-e* (invariável), os sufixados em *-idade* têm, globalmente, valores de frequência superiores aos registados entre os sufixados em *-ice*. Com efeito, e à exceção do nome *velhice*, cuja frequência é de 9 mil por milhão, as formas *vigarice* e *criancice* registam uma frequência lexical inferior a mil por milhão. Já todas as formas sufixadas em *-idade* apresentam índices de frequência lexical superiores a 8 mil por milhão.

De forma a averiguar se existe correlação entre a proporção de desvios e a frequência lexical dos itens, consideram-se, num primeiro momento, os resultados relativos aos itens considerados frequentes, cujo valor de frequência por milhão é superior a 10 mil e, em seguida, os resultados apurados nos nomes menos frequentes, em que o valor de frequência por milhão é inferior a 10 mil. Por fim, apresentam-se e discutem-se os resultados obtidos relativamente às 8 pseudopalavras.

Os nomes sufixados em *-ção*, *-s[z]ão* e *-agem* apresentam, cada um, três formas lexicais consideradas frequentes, i.e., com valores acima dos 10 mil por milhão. Neste conjunto também se inclui um nome sufixado em *-idade* (cf. Quadro 12).

Quadro 12. Distribuição das respostas, do número absoluto de ocorrências desviantes e das respetivas médias (percentagens) apuradas nos nomes cujos valores de frequência por milhão são superiores a 10 mil por milhão

Nome	Frequência CdP/ milhão	Nível QECL									TOTAIS		
		A1-A2			B1-B2			C1-C2			# R	# D	%
		# R*	# D**	%	# R	# D	%	# R	# D	%			
<i>evolução</i>	453.6	77	9	11.69	44	0	.00	6	0	.00	127	9	7.09
<i>decisão</i>	194.34	77	9	11.69	44	0	.00	6	0	.00	127	9	7.09
<i>passagem</i>	88.13	77	49	63.64	44	14	31.82	6	0	.00	127	63	49.61

Nome	Frequência CdP/ milhão	Nível QECRL									TOTAIS		
		A1-A2			B1-B2			C1-C2			# R	# D	%
		# R*	# D**	%	# R	# D	%	# R	# D	%			
<i>preocupação</i>	67.21	77	2	2.60	44	0	.00	6	0	.00	127	2	1.57
<i>abordagem</i>	45.1	77	48	62.34	44	15	34.09	6	0	.00	127	63	49.61
<i>satisfação</i>	44.66	77	7	9,09	44	0	.00	6	0	.00	127	7	5.51
<i>finalidade</i>	42.74	77	2	2,60	44	0	.00	6	0	.00	127	2	1.57
<i>revisão</i>	38.39	77	7	9,09	44	0	.00	6	1	16.67	127	8	6.30
<i>confusão</i>	37.33	77	12	15,58	44	1	2.27	6	0	.00	127	13	10.24
<i>lavagem</i>	13.27	77	38	49.35	44	13	29.55	6	0	.00	127	51	40.16
\bar{X}	102.477	77	18.3	23.77	44	4.3	9.77	6	0.10	1.67	127	22.7	17.87

* R (respostas)

** D (desvios)

A média global de desvios entre os nomes com valores de frequência acima dos 10 mil por milhão é de 22.7 desvios por item. Nos níveis iniciais (A1-A2), a média de desvios é mais expressiva (18.3 desvios por item) por oposição ao valor médio de desvios registado nos níveis mais avançados (C1-C2) que é de 0.10 desvios por item. Também é de salientar uma diminuição acentuada das médias de desvios por item lexical registadas entre os níveis A1-A2 e B1-B2, já que nos níveis intermédios a média global de desvios apurada por item é de 4.3. Portanto, é a partir dos níveis intermédios que se verifica uma redução significativa de desvios de atribuição de género relativamente aos nomes considerados frequentes.

À exceção do grupo de informantes do segmento da amostra de C1-C2 – que apenas apresenta uma ocorrência desviante relativa ao nome *revisão* –, a maior proporção de desvios nos restantes segmentos recai em três formas nominais, sufixadas em *-agem*. Com efeito, os itens *passagem*, *abordagem* e *lavagem* registam, nos níveis A1-A2, os valores percentuais de desvios mais elevados de 63.64%, 62.34% e 49.35%, respetivamente. Já nos níveis intermédios, embora se registre uma diminuição considerável dos índices de desvios em todos os itens, os nomes sufixados em *-agem* apresentam uma proporção média de desvios de, aproximadamente, 30%. Portanto, estes resultados reforçam ainda a ideia de que o sufixo derivacional é determinante para condicionar a atribuição do valor de género, independentemente da frequência do item lexical, sobretudo numa fase inicial da aprendizagem. Além disso, estes dados evidenciam que não existe uma relação direta entre a frequência lexical e a proporção de desvios, pois em nomes com maiores valores de frequência a incidência de desvios é superior à registada nos nomes com frequências mais baixas.

Considere-se, em seguida, os resultados apurados entre os nomes cujo valor de frequência lexical é inferior a 10 mil por milhão ([Quadro 13](#)).

Quadro 13. Distribuição das respostas, do número absoluto de ocorrências desviantes e das respetivas médias (percentagens) apuradas nos nomes cujos valores de frequência por milhão são inferiores a 10 mil por milhão

Nome	Frequência CdP/ milhão	Nível QECRL									TOTAIS		
		A1-A2			B1-B2			C1-C2			# R	# D	%
		# R*	# D**	%	# R	# D	%	# R	# D	%			
<i>velhice</i>	9.5	77	30	38.96	44	12	27.27	6	0	.00	127	42	33.07
<i>exclusividade</i>	8.57	77	2	2.60	44	0	.00	6	0	.00	127	2	1.57
<i>falsidade</i>	8.05	77	9	11.69	44	3	6.82	6	0	.00	127	12	9.45
<i>alusão</i>	7.07	77	19	24.68	44	5	11.36	6	0	.00	127	24	18.90
<i>cumplicidade</i>	6.83	77	1	1.30	44	0	.00	6	0	.00	127	1	0.79
<i>confraternização</i>	2.11	77	10	12.99	44	1	2.27	6	0	.00	127	11	8.66
<i>puxão</i>	1.62	77	12	15.58	44	10	22.73	6	0	.00	127	22	17.32
<i>empurrão</i>	1.54	77	14	18.18	44	5	11.36	6	1	16.7	127	20	15.75
<i>arrastão</i>	0.82	77	33	42.86	44	17	38.64	6	1	16.7	127	51	40.16
<i>vigarice</i>	0.59	77	20	25.97	44	6	13.64	6	0	.00	127	26	20.4
<i>abanão</i>	0.2	77	18	23.38	44	6	13.64	6	0	.00	127	24	18.90
<i>criancice</i>	0.12	77	21	27.27	44	9	20.45	6	0	.00	127	30	23.62
\bar{X}	3.92	77	15.75	20.45	44	6.17	14.02	6	0.10	2.78	127	22.08	17.39

* R (respostas)

** D (desvios)

Entre os itens considerados menos frequentes, regista-se uma média global de 22.08 desvios por nome, sendo que o valor apurado não é substancialmente diferente do registado nos nomes considerados mais frequentes (22.7, cf. [Quadro 12](#)). Por conseguinte, este é um indicador de que a frequência lexical não determina o desempenho dos informantes. É também no grupo dos informantes dos níveis A1-A2 que se regista a média de desvios por nome mais alta (15.75), por oposição aos valores da média de desvios apurados nos níveis B1-B2 (6.17) e nos níveis mais avançados (0.10). Com efeito, também aqui e à semelhança do que se verificou nos nomes considerados mais frequentes, nos níveis C1-C2, as incidências de desvios são residuais (cf. [Quadro 13](#)).

Os derivados em *-ice* (*velhice*, *criancice* e *vigarice*) e em *-ão* (*alusão*, *puxão*, *empurrão*, *arrastão* e *abanão*) registam, globalmente, maiores incidências de desvios, independentemente da sua frequência lexical. Com efeito, o item *velhice* tem um valor de frequência (9.5 por milhão) ligeiramente superior ao registado nos sufixados em *-idade*, mas, ainda assim, apresenta valores da média de desvios globais consideravelmente mais altos (33.07%), do que os valores globais apurados em *exclusividade* (1.57%), *falsidade* (9.45%) e *cumplicidade* (0.79%).

Estes dados confirmam, uma vez mais, que a baixa representatividade e/ou a idiosincrasia do sufixo condiciona a associação de valores de género aos nomes que o apresentam. No caso de *velhice*, a singularidade do sufixo no panorama derivacional do português ([Rio-Torto 2022](#)), que o contraste com o equivalente espanhol *vejez* ilustra, não favorece uma estratégia de transferência interlinguística.

Em seguida, consideramos os valores apurados entre as pseudopalavras criadas para este estudo (cf. Quadro 14).

Quadro 14. Distribuição das respostas, do número absoluto de ocorrências desviantes e das respetivas médias (percentagens) apuradas nos nomes possíveis/pseudopalavras sufixadas

Nomes possíveis	Nível QECRL									TOTAIS		
	A1-A2			B1-B2			C1-C2			# R	# D	%
	# R*	# D**	%	# R	# D	%	# R	# D	%			
<i>resfrião</i>	77	8	10.39	44	1	2.27	6	0	.00	127	9	7.09
<i>avisação</i>	77	31	40.26	44	17	38.64	6	0	.00	127	48	37.80
<i>precisão</i>	77	14	18.18	44	3	6.82	6	0	.00	127	17	13.39
<i>levantagem</i>	77	39	50.65	44	14	31.82	6	1	16.67	127	54	42.52
<i>tratagem</i>	77	48	62.34	44	13	29.55	6	1	16.67	127	62	48.82
<i>contentice</i>	77	34	44.16	44	22	50.00	6	1	16.67	127	57	44.88
<i>vitalice</i>	77	6	7.79	44	1	2.27	6	0	.00	127	7	5.51
<i>prontidade</i>	77	6	7.79	44	2	4.55	6	0	.00	127	8	6.30
\bar{X}	77	23.25	30.19	44	9.13	20.74	6	0.38	6.25	127	32.75	25.67

* R (respostas)

** D (desvios)

Entre as formas nominais possíveis, constata-se que a média de desvios é a mais alta em termos globais (32.75 desvios por nome) (cf. Quadro 14). Neste conjunto, destaca-se uma maior proporção global de desvios nas pseudopalavras em *-agem* (*levantagem* apresenta um valor global de desvios de 42.52% e *tratagem* regista 48.82%) e na pseudopalavra *contentice* (44.8%), havendo ainda uma proporção global considerável de desvios no nome *avisação* (37.80%).

Se atentarmos ao desempenho registado por segmento da amostra, constata-se que nos níveis iniciais, A1-A2, a maior proporção de desvios recai nas pseudopalavras em *-agem* (*levantagem* (50.65%) e *tratagem* (62.34%)), na pseudopalavra *contentice* (44.16%) e em *avisação* (40.26%). A mesma tendência observa-se nos níveis intermédios, ou seja, a maior incidência de desvios recai nestes mesmos nomes. Além disso, e à exceção de *contentice* que apresenta, em B1-B2, uma incidência de desvios superior à registada em A1-A2 (50%), todos os restantes itens apresentam uma diminuição gradual da proporção de desvios.

Em C1-C2, as ocorrências desviantes manifestam-se nas pseudopalavras *levantagem*, *tratagem* e *contentice* que registam o mesmo índice proporcional de desvios, 16.67%, sendo este valor correspondente a um desvio por item lexical. Além disso, o facto de neste segmento da amostra se terem registado nos nomes possíveis a média percentual de desvios mais elevada por item (6.25%) relativamente à apurada nos outros conjuntos de nomes (1.67% para os nomes mais frequentes (cf. Quadro 12) e 2.78% para os nomes menos frequentes (cf. Quadro 13)), permite-nos aferir que, em fases mais avançadas da aprendizagem linguística, os aprendentes assimilaram o valor de género do item nominal, mas nas formas não atestadas acusam ainda um efeito de transferência linguística direta de atribuição de valores de género.

Por fim, os dados apurados nos diferentes segmentos da amostra empírica confirmam ainda um certo 'efeito de sufixo' no momento da decisão da atribuição de um valor de género,

pois são poucos expressivos os desvios que recaem sobre as formas possíveis sufixadas em *-idade* face à proporção de desvios assinalada nas pseudopalavras sufixadas em *-ice* e em *-agem* (cf. Quadro 14).

Os dados apurados convocam uma discussão em várias frentes, que sumariamente vamos aqui empreender.

A atribuição de género, nomeadamente por falantes de uma língua não nativa, rege-se por variáveis que também estão presentes no processo de atribuição aqui estudado, e que convergem largamente com o que se preconiza na ‘Canonical approach’ da aquisição de género, cunhada por Corbett (2005), que aponta para uma codificação mais célere e mais bem-sucedida em construções menos opacas e mais lenta e mais desviante em construções mais opacas, porque menos representadas e mais idiossincráticas. A similaridade tipológica entre a língua-fonte e a língua-alvo, o nível de proficiência do aprendente, a maior/menor representatividade, canonicidade/perifericidade e opacidade do sufixo usado, bem como a respetiva classe temática do nome, são fatores que influem no processo de construção interlinguístico das manifestações de género na língua-alvo. Como se verificou, os dados apurados atestam que (i) a incidência de desvios diminui na razão inversa da progressão para o nível mais avançado de aprendizagem, e que (ii) há relação entre incidência de desvios nos diferentes conjuntos de informantes e maior/menor representatividade, maior/menor opacidade do sufixo usado e respetiva classe temática do nome. Estas considerações estão, assim, em consonância com o que diversos estudos comprovam sobre as variáveis mais influentes na atribuição de valores de género por aprendentes tardios de uma LE (cf. Bruhn de Garavito & White 2002; White *et al.* 2004; Franceschina 2005; Sabourin, Stowe & Haan 2006; Isabelli-García 2010; Martins 2015; Lacsán 2016; Pinto 2017; Ferreira 2019; Rio-Torto & Ferreira 2022).

Na aprendizagem do espanhol como LE, por falantes de línguas nativas tipologicamente muito distintas da castelhana, como árabe, malaio, ou basco, verifica-se que a atribuição de género é um dos domínios de maior fonte de erros na interlíngua que vai sendo construída pelos aprendentes. Mas, como o estudo de Ferreira & Elejalde (2017) mostra, também em falantes de línguas tipologicamente próximas do espanhol, depois da omissão do til nas esdrúxulas “*la falsa selección de género* corresponde al error más frecuente de todo el corpus con un total de 968 (30%) ocurrencias” (Ferreira & Elejalde 2017: 527), num universo de aprendentes de espanhol como língua estrangeira cujas línguas maternas são o alemão (32%), o francês (27%), o inglês (27%), o português (3%), o sueco (3%), o checo (3%), o italiano (2%), e o russo (2%) (Ferreira & Elejalde 2017: 517). Por seu turno, em composições de estudantes malaios aprendentes de espanhol LE, os erros mais frequentes na atribuição de género situam-se nos itens lexicais globalmente considerados menos unívocos, tais como os que têm como fronteira direita *-e* (**el torre, *el gente, *el calle, *la parque, *la paisaje, *la aire*), nasal (**una pan, *la jardin, *un investigación*) e os terminados em *-r* ou *-s* (**la lugar, *la país, *los flores*), por contraste com os itens lexicais que terminam em *-o* ou em *-a*, genericamente menos propensos a opções divergentes (Mohd Hayas 2006: 33-34). Estes mesmos padrões também se verificam na aquisição do género na língua materna. Como mostra Imaz Agirre (2016: 523), na aquisição de género na língua materna por falantes de espanhol e por falantes bilingues deste idioma e de basco, “[c]anonicity seems to be a strong cue for both groups, since all participants were more accurate and faster with canonical word endings”, o que confirma o peso decisivo da configuração morfofonológica do sufixo (e, portanto, da sua classe temática) no (in)sucesso da atribuição de género.

Na atribuição do valor de género em contexto de aprendizagem de uma LE, a influência da frequência do item lexical sufixado tem menos peso que a familiaridade do padrão sufixal em

jogo no léxico mental do aprendente. De igual modo, permanece controverso o efeito da proximidade linguística (Ellis 2002a, 2002b) entre a língua-fonte e a língua-alvo no sucesso da codificação do género, mesmo em palavras com estruturas semelhantes ou cognatas, obrigando assim a relativizar o “efeito de congruência de género (“gender congruency effect”) que, todavia, desempenha um papel relevante na construção da interlíngua, e poderá ocorrer na transferência do género associado a *-aje* (esp.) para o de *-agem* (port.).

6. CONSIDERAÇÕES FINAIS

Neste estudo, procurou-se averiguar padrões relativos à atribuição de valores de género nominal a nomes sufixados do português por aprendentes tardios, falantes nativos de espanhol, a frequentar turmas de diferentes níveis QECRL, de A1 a C2. A base empírica deste trabalho resultou da aplicação de um questionário *online*, no qual se solicitou aos informantes que associassem a 30 formas nominais distintas, portadoras dos sufixos *-ão*, *-ção*, *-s[z]ão*, *-agem*, *-ice* e *-idade*, o valor de género correspondente, selecionando, para cada item, a forma adequada do artigo definido.

Os resultados globais apurados evidenciam, e à semelhança de estudos anteriores (Ferreira 2019; Martins 2015; Lacsán 2016; Pinto 2017) uma diminuição gradual do número de desvios à medida que os aprendentes progridem na aprendizagem do português. Estes dados são, assim, reveladores de que estamos perante erros transitórios (ou de desenvolvimento) característicos do processo de construção do sistema interlinguístico do aprendente. Com efeito, nos níveis mais avançados, C1-C2, regista-se uma taxa de desvios muito baixa. Nos níveis A1-A2 a proporção de desvios registada foi de 24.03% (cf. Gráfico 1), sendo este valor um indício de que, mesmo numa fase inicial da aprendizagem os aprendentes, falantes nativos do espanhol, atribuem adequadamente os valores de género aos itens nominais sufixados em português. Verificou-se ainda que uma parte substancial dos nomes com maior incidência de desvios nos níveis iniciais não apresentam desvios nos níveis mais avançados, sugerindo-se, assim, que estes aprendentes vão assimilando, com razoável grau de sucesso, os valores de género desses itens.

No que respeita à relação entre a ocorrência de desvios e a classe temática dos nomes, os dados apurados não evidenciam diferenças de desempenho estatisticamente relevantes no segmento dos aprendentes a frequentar turmas dos níveis C1-C2, já que nestes se apurou, por classe temática, o mesmo valor percentual da média de desvios (3.33%). Nos restantes segmentos da amostra empírica, os valores apurados revelam uma maior taxa de desvios na classe dos atemáticos, sufixados em *-agem*. Com efeito, nestes nomes regista-se, globalmente, uma média de desvios de 45.98%, sendo a incidência de casos desviantes mais expressiva nos níveis A1-A2 (57.66%) (cf. Quadro 9). Por se tratar de um sufixo cognato de *-aje* com um valor de género antagónico, os resultados apurados apontam para um efeito de transferência de valor de género, sobretudo nos aprendentes a frequentar turmas de níveis mais baixos. Contudo, os efeitos da influência interlinguística deixam de se fazer sentir de modo expressivo nos níveis mais avançados. Assim, a proximidade formal entre os sufixos poderá contribuir para que os aprendentes, sobretudo em fases iniciais da aprendizagem, assumam que nomes formalmente idênticos são marcados com o mesmo valor de género. Num estágio mais avançado da aprendizagem, as assimetrias entre a L1 e a LE parecem não condicionar, em grande medida, o comportamento dos aprendentes no que à associação de valores de género aos nomes sufixados diz respeito.

Os dados apurados indiciam ainda uma relação entre os índices de desvio e a natureza [\pm opaca] ou [\pm idiossincrática] do sufixo em causa. Por outras palavras, o desempenho

registado nestes aprendentes parece ser sensível à opacidade da classe temática do nome e do respetivo sufixo em causa. O contraste entre os valores das incidências de desvios globais apuradas nos nomes sufixados em *-ice* (25.67%) e *-idade* (3.94%), pertencentes à classe dos nomes de Tema *-e* (invariável), ou entre os nomes portadores de *-ão* (19.84%), *-ção* (11.65%) e *-s[z]ão* (11.18%), da classe dos itens de Tema \emptyset ([ẽ̃]), revela que o grau de idiomaticidade e/ou de representatividade do sufixo no *input* a que o aprendente está exposto tem um papel relevante para os índices de desvios apurados.

A singularidade de *-ice* no panorama derivacional do português, a sua idiosincrasia face a outras línguas românicas, como o espanhol, e o seu reduzido índice de frequência no *input* ao qual os aprendentes estão expostos, produzem um ‘efeito de sufixo’ que se traduz por um elevado índice de desvios, dada a sua total opacidade formal no que respeita ao sentido de atribuição de valores de género gramatical. O mesmo já não se verifica de forma tão expressiva entre os nomes sufixados em *-idade*, dada a proximidade formal deste sufixo com o correspondente espanhol (*-dad*) e de ambos originarem, nos dois idiomas, formas nominais derivadas marcadas com o valor de género feminino.

No que concerne ao peso da frequência lexical dos itens, verifica-se que não existe uma relação direta entre a frequência do item lexical e a proporção de desvios apurada, pois em nomes com maiores valores de frequência, a incidência de desvios é superior à registada nos nomes com frequências mais baixas. Mais uma vez, o que parece ser determinante para o comportamento destes aprendentes é o grau de opacidade do sufixo e o nível de aprendizagem em que se encontram. O caso de *confraternização* é lapidar, pois sendo um nome de baixa frequência no português (cf. Quadro 11), portador de um sufixo muito representado, e com estrutura derivacional e formal idêntica à do homólogo espanhol (*confraternización*), regista uma taxa de desvios de 12.99% nos aprendentes de nível A1-A2, de 2.27% nos de nível B1-B2 e de .00% nos de nível C1-C2. O efeito de frequência faz-se sentir não tanto em relação a cada palavra, mas antes em relação a cada afixo, ou seja, ao nível do morfema, o que constitui um enorme desafio para as teorias morfológicas e do processamento morfolexical.

Por fim, este trabalho permitiu aferir que ao longo do processo de aquisição/aprendizagem de valores de género dos nomes sufixados (atemáticos, de Tema \emptyset e de Tema *-e* (invariável)) por aprendentes tardios de PLE, falantes nativos de espanhol, (i) a proximidade formal dos sistemas linguísticos, nomeadamente nos domínios derivacional e de atribuição de género, por um lado, e (ii) o grau de opacidade do sufixo, por outro, são determinantes.

Agradecimentos

Queremos manifestar o nosso profundo agradecimento aos dois avaliadores anónimos, cujas considerações muito contribuíram para melhorar a qualidade da reflexão empreendida. Os erros remanescentes são da nossa exclusiva responsabilidade.

Financiamento

O presente trabalho é financiado por fundos nacionais portugueses, através da FCT (Fundação para a Ciência e Tecnologia), como parte do projeto do Centro de Estudos de Linguística Geral e Aplicada (CELGA-ILTEC), UID FCT 4887, da Universidade de Coimbra.

Referências bibliográficas

- Alonso Alonso, María Rosa. 1999. A relación interlingua-transferencia. *Revista Galega do Ensino* (25), 137-147.
- Alonso Alonso, María Rosa. 2020. *La transferencia lingüística: perspectivas actuales*. Granada: Editorial Comares.
- Ambadiang, Théophile. 1999. La Flexión nominal: género y número. En Ignacio Bosque & Violeta Demonte (orgs.), *Gramática Descriptiva del Español*, vol. III. 4843-4913. Madrid: Gredos.
- Audring, Jenny. 2016. *Gender*. <https://doi.org/10.1093/acrefore/9780199384655.013.43>
- Bordag, Denisa & Thomas Pechmann. 2007. Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, 10(3), 299-314.
- Bordag, Denisa, Andreas Opitz & Thomas Pechmann. 2006. Gender processing in first and second languages: the role of noun termination. *Journal of experimental psychology. Learning, memory, and cognition*, vol. 32(5), 1090-1101.
- Bruhn de Garavito, Joyce & Lydia White. 2002. The second language acquisition of Spanish DPS: the status of grammatical features. En Ana Teresa Pérez-Leroux & Juana Muñoz Licerias (eds.), *The acquisition of Spanish morphosyntax: the L1/L2 connection*. 153-178. Dordrecht: Kluwer.
- Caffarra, Sendy, Anna Siyanova-Chanturia, Francesca Pesciarelli, Francesco Vespignani & Cristina Cacciari. 2015. Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian? *Psychophysiology* 52(8), 1019-1030.
- Conselho da Europa. 2001. *Quadro Europeu Comum de Referência para as Línguas. Aprendizagem, ensino, avaliação*. Porto: Edições ASA.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville. 2005. The canonical approach in typology. En Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic Diversity in Language Theories*. 25-49. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Corbett, Greville. 2013a. *The Expression of Gender. The Expression of Cognitive Categories* (ECC). Vol. 6. Berlin: De Gruyter Mouton.
- Corbett, Greville. 2013b. Number of Genders. En Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30> [18/03/2022]
- Corbett, Greville & Norman Fraser. 2000. Gender assignment: a typology and a model. En Gunter Senft (ed.), *Systems of nominal classification*. 293-325. Cambridge: Cambridge University Press.

- Corder, Stephen Pit. 1967. The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(1-4), 161-170.
- Corder, Stephen Pit. 1993. A role for the mother tongue. En Susan Gass & Larry Selinker (eds.), *Language transfer in language learning (revised edition)*. 18-31. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Davies, Mark. *Corpus do Português - Web / Dialects* [CdP]. <https://www.corpusdoportugues.org/web-dial> [29.09.2021]
- DeAngelis, Gessica. 2005. Interlanguage Transfer of Function Words. *Language Learning*, 55(3), 379-414.
- Diez, Frederich. 1874 (1836-1843). *Grammaire des langues romanes*. 3.. ed. Paris: A. Franck.
- Ellis, Nick C. (2002a). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, Nick C. (2002b). Reflections on Frequency Effects in Language Processing. *Studies in Second Language Acquisition*. 24, 297-339.
- Ellis, Rod. 2000. *The study of second language acquisition*. Oxford University Press.
- Fernández López, Sonsoles. 1995. Errores e interlengua en el aprendizaje del español como lengua extranjera. *Didáctica. Lengua y Literatura*, 7, 203-215.
- Ferreira, Tânia. 2019. *Aquisição/Aprendizagem do sistema de atribuição de género nominal em PLNM*. Coimbra: Universidade de Coimbra. [Tese de doutoramento inédita].
- Ferreira, Tânia. 2021. A morfologia dos desvios de género gramatical em PLNM. *Diadorim*, 23 (2), 210-226.
- Ferreira Cabrera, Anita; Jéssica Elejalde Gómez. 2017. Análise de erros recorrentes em um corpus de Aprendizes de Espanhol como Língua Estrangeira (Corpus CAELE). *Revista Brasileira de Linguística Aplicada*, 17(3), 509-537.
- Franceschina, Florencia. 2005. *Fossilized second language grammars: the acquisition of grammatical gender*. Amsterdam / Philadelphia: John Benjamins.
- Gass, Susan M.; Larry Selinker. 2008 (1994). *Second Language Acquisition: an introductory course*. 3rd edition. London / New York: Routledge Taylor & Francis Group.
- Imaz Agirre, Ainara. 2016. The processing of gender assignment in spanish. Does canonicity have an effect on Basque speakers? *Revista Española de Lingüística Aplicada / Spanish Journal of Applied Linguistics*, 29(2), 523-543.
- Isabelli-García, Christina. 2010. Acquisition of Spanish gender agreement in two learning contexts: study abroad and at home. *Foreign Language Annals*, 43 (2), 289-303.

- Lacsán, Veronika. 2016. *The acquisition of gender agreement in L2 Portuguese by adult L1 Hungarian speakers*. Lisboa: Universidade de Lisboa. [Dissertação de Mestrado]. <https://repositorio.ul.pt/handle/10451/24501>
- Larsen-Freeman, Diane; Michael H. Long. 1991. *An introduction to second language research*. London / New York: Routledge Taylor & Francis Group.
- Leiria, Isabel. 2006. *Léxico, aquisição e ensino do Português europeu língua não materna*. Lisboa: FCG/FCT.
- Lenneberg, Eric H. 1967. The Biological Foundations of Language. *Hospital Practice*, 2 (12), 59-67.
- Long, Michael H. 2003. Stabilization and Fossilization in Interlanguage Development. En Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*. 487-535. Malden: Blackwell.
- Lott, David. 1983. Analysing and counteracting interference errors. *English Language Teaching Journal*, 37 (3), 256-262.
- Martins, Cristina. 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.
- Mattoso Câmara, Joaquim. 1970. *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.
- Mattoso Câmara, Joaquim. 1975. *História e Estrutura da Língua Portuguesa*. Rio de Janeiro: Padrão.
- Meyer-Lübke, Walter. 1895. *Grammaire des langues romanes*. Traduction par Auguste Doutrepoint et Georges Doutrepoint. Tome II *Morphologie*. Paris: H. Welter Éditeur.
- Mohd Hayas, Kasma. 2006. *Errores léxicos en las composiciones de los estudiantes malayos en el aprendizaje del español como lengua extranjera*. Salamanca: Universidad de Salamanca. [Memória do Máster de Enseñanza de Español como Lengua Extranjera].
- Mota, Maria Antónia. 2020. Morfologia do nome e do adjetivo. En Eduardo Buzaglo Paiva Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, Amália Mendes & Amália Andrade (orgs.), *Gramática do português*. Vol. III. 2835-2930. Lisboa: Fundação Calouste Gulbenkian.
- Nunes, José Joaquim. 1989. *Compêndio de Gramática Histórica Portuguesa*. Lisboa: Clássica Editora.
- Pharies, David. 2002. *Diccionario etimológico de los sufijos españoles*. Madrid: Gredos.
- Pinto, Jorge. 2017. A aquisição do género e da concordância de género em português língua terceira ou língua adicional. En Paulo Osório (coord.), *Teorias e Usos Linguísticos*. 91-110. Lisboa: Lidel.

- Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa.
- Rio-Torto, Graça. 2022, Peculiarities of Portuguese Word-Formation. En Michele Loporcaro & Francesco Gardani (eds.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. <https://oxfordre.com/linguistics/page/romance-linguistics/>
- Rio-Torto, Graça & Alexandra Rodrigues. 2016 (2013). Formação de nomes. En Graça Rio-Torto, Alexandra Soares, Isabel Pereira, Rui Pereira & Sílvia Ribeiro (orgs.), *Gramática derivacional do português*. 2.^a ed. 135-240. Coimbra: Imprensa da Universidade de Coimbra.
- Rio-Torto, Graça, Alexandra Soares, Isabel Pereira, Rui Pereira & Sílvia Ribeiro. 2016 (2013). *Gramática derivacional do português*. 2.^a ed. Coimbra: Imprensa da Universidade de Coimbra.
- Rio-Torto, Graça & Tânia Ferreira. 2022. Atribuição de Género Gramatical por Aprendentes de PLNM em Palavras Sufixadas. *Confluência*, 63, 44-76.
- Rodrigues, Alexandra Soares. 2016. Noções basilares sobre morfologia e o léxico. En Graça Rio-Torto Alexandra Soares, Isabel Pereira, Rui Pereira & Sílvia Ribeiro (orgs.), *Gramática derivacional do português*. 2. ed. 35-133. Coimbra: Imprensa da Universidade de Coimbra.
- Sabourin, Laura, Laurie A. Stowe & Ger J. de Haan. 2006. Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22 (1), 1-29.
- Sá-Leite, Ana Rita, Karlos Luna, Isabel Fraga & Montserrat Comesaña. 2020. The gender congruency effect across languages in bilinguals: A meta-analysis. *Psychonomic Bulletin & Review* 27, 677-693.
- Santiago Lacuesta, Ramón & Eugénio Bustos Gisber. 1999. La derivación nominal. En Ignacio Bosque & Violeta Demonte, *Gramática descriptiva de la lengua española*, vol. 3. 4505-4594. Madrid: Espasa Calpe.
- Schachter, Jacquelyn. 1993. A new account for language transfer. En Susan Gass & Larry Selinker (eds.), *Language Transfer in Language Learning* (revised edition). 32-46. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209-231.
- Selinker, Larry & William E. Rutherford. 1992. *Rediscovering Interlanguage*. London and New York: Routledge.
- Taft, Marcus & Fanny Meunier. 1998. Lexical representation of gender: A quasiregular domain. *Journal of Psycholinguistic Research*, vol. 27(1), 23-45.
- Tucker, G. Richard, Wallace E. Lambert & André Rigault. 1977. *The French Speaker's Skill with Grammatical Gender: An Example of Rule-governed Behavior*. The Hague: Mouton.

- White, Lydia. 2003. On the Nature of Interlanguage Representation: Universal Grammar in the Second Language. En Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*. 18-42. Blackwell Publishing.
- White, Lydia, Elena Valenzuela, Martyna Kozłowska-Macgregor & Yan-Kit Ingrid Leung. 2004. Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics* 25(1), 105-133.