

## ATRIBUIÇÃO DE GÊNERO GRAMATICAL POR APRENDENTES DE PLNM EM PALAVRAS SUFIXADAS

Graça Maria de Oliveira e Silva Rio-Torto (UC, CELGA-ILTEC), Prof. Catedrática de Linguística da Faculdade de Letras da Universidade de Coimbra, membro do CELGA-ILTEC, [gracart@gmail.com](mailto:gracart@gmail.com), <https://orcid.org/0000-0002-1525-0737>

Tânia Santos Ferreira (UC, CELGA-ILTEC), Investigadora Doutorada da Faculdade de Letras da Universidade de Coimbra, membro do CELGA-ILTEC, [tania.ferreira@fl.uc.pt](mailto:tania.ferreira@fl.uc.pt), <https://orcid.org/0000-0002-8045-3506>

## GRAMMATICAL GENDER ASSIGNMENT BY PORTUGUESE NON-NATIVE LANGUAGE LEARNERS IN SUFFIXED WORDS

### RESUMO:

Neste estudo, pretende-se indagar em que medida a atribuição de género gramatical em nomes sufixados, por aprendentes de Português europeu como Língua não Materna (PLNM), varia em função do sufixo e, mais especificamente, da classe formal deste e da marca de género associada. A base empírica deste estudo foi extraída de um inquérito, no qual se solicitava a aprendentes tardios, falantes nativos de chinês, a frequentar turmas de diferentes níveis de aprendizagem (de B1 a C1), que explicitassem o valor de género nominal de nomes sufixados em *-ção*, *-s[iz]ão*, *-ão*, *-agem*, *-idade* e *-ice*. Registaram-se padrões idênticos em todos os níveis de aprendizagem, mas há uma incidência assinalável de desvios em formas sufixadas em *-ão* e *-ice*. Estes resultados mostram que a atribuição de um valor de género ao nome depende do grau de representatividade e de opacidade do sufixo no *input* e do respetivo valor de género associado.

**PALAVRAS-CHAVE:** Género gramatical. Aquisição lexical. Sufixação. Derivação. Português como língua não materna.

### ABSTRACT:

This study aims to investigate to what extent the grammatical gender assignment of suffixed nouns by learners of European Portuguese as a non-native language (PNNL) varies according to the suffix and, more specifically, its formal class and the gender's marker associated to them. The empirical basis of this study was drawn from a survey, in which late learners, native speakers of Chinese, who attend classes of different levels of learning (from B1 to C1), were asked to explicit the nominal gender value of suffixed nouns in *-ção*, *-s[iz]ão*, *-ão*, *-agem*, *-idade* and *-ice*. Identical patterns were found at all levels of learning, but there is a notable incidence of deviations in suffixed forms in *-ão* and *-ice*. These results show that the assignment of nouns to genders depends on the degree of representativeness and opacity of the suffix in the input and its associated gender value.

**KEYWORDS:** Grammatical gender. Lexical acquisition. Suffixation. Derivation. Portuguese as a non-native language.

Commented [ET1]: Incluir nome completo do(s) autor(es), vínculos institucionais, emails, Orcid. Informar biodados atualizados para constar na seção Colaboradores deste número.

## Introdução

A atribuição do género gramatical dos nomes da língua portuguesa rege-se por padrões que Corbett (1991) designa como '(predominantemente) semânticos' e 'formais'. O género gramatical em português compreende duas grandes classes, a de MASCULINO (doravante MASC) e a de FEMININO (doravante FEM), sendo residuais as manifestações de neutro (MATTOSO CÂMARA, 1970). Como em outras línguas, a atribuição do género gramatical está muitas vezes relacionada com as marcas formais (fónicas e/ou morfológicas) da fronteira direita da palavra, mas estas não são unívocas e previsíveis sobretudo quando estão em jogo nomes de entidades não animadas.

O género gramatical é uma propriedade lexical do nome que lhe é intrínseca (CHOMSKY, 1995). Como já Mattoso Câmara (1975, p. 77) postula "[o] caráter masculino ou feminino da palavra está imanente na palavra e é de natureza lexical, não flexional." Trata-se, portanto, de uma propriedade nominal obrigatória (ou seja, não opcional), que pode ser codificada formalmente/morfológicamente (*ancião* vs. *anciã*<sup>1</sup>; *czar* vs. *czarina*; *doutor* vs. *doutora*) e/ou sintaticamente (*o/a atleta*, *o/a turista*, *o/a viajante*).

Quando a palavra é sufixada, o género gramatical da unidade lexical está fortemente dependente do sufixo usado. A relevância destes e do grau de exposição dos aprendentes tardios de Português como Língua não Materna (doravante PLNM) ao seu uso é tanto maior quanto maior a opacidade semântica e formal dos sufixos selecionados e das classes temáticas em jogo.

O presente estudo encontra-se organizado da seguinte forma: na secção 1 apresentam-se os objetivos, os dados empíricos usados no inquérito, as perguntas de investigação formuladas, tendo em conta as classes temáticas dos nomes e os sufixos em jogo, bem como as variáveis já testadas na atribuição de género gramatical por aprendentes de PLNM. Na secção 2, procede-se à descrição da influência do sufixo na atribuição de valor de género gramatical em Português. As secções 3 e 4 são consagradas à apresentação e discussão dos dados inquiridos, em vista à validação das hipóteses formuladas.

### 1. Objetivos, dados empíricos e perguntas de investigação

Este estudo tem por objetivo analisar em que medida a atribuição de género gramatical em nomes sufixados, por aprendentes tardios de PLNM, varia em função do sufixo usado, da classe formal deste e da marca de género que lhe está associada. Para avaliar a fiabilidade desta hipótese, foi elaborado um inquérito no qual se solicitava a aprendentes tardios<sup>2</sup> de PLNM a frequentar turmas dos níveis B1, B2 e C1 (segundo a escala do *Quadro Europeu Comum de Referência para as Línguas* (QECL) (CONSELHO DA EUROPA, 2001)) que explicitassem qual o género gramatical de nomes portadores de sufixos com acentuada opacidade para a atribuição de género. Foram, pois, selecionados sufixos de diferentes classes temáticas que estudos anteriores revelaram apresentarem significativas dificuldades de atribuição de género por falantes de PLNM (MARTINS, 2015; FERREIRA, 2019).

<sup>1</sup> Sobre a explicação histórica do mecanismo de marcação de género nestes nomes ver Rio-Torto, 2002.

<sup>2</sup> Os informantes que integram a base empírica deste estudo são alunos universitários que, no momento da recolha dos dados, se encontravam a frequentar cursos de língua portuguesa para estrangeiros. O nosso agradecimento aos alunos que, anonimamente, colaboraram neste projeto e aos docentes que o integraram na sua atividade letiva.

Os desvios na atribuição de género gramatical por aprendentes de PLNM em palavras não sufixadas variam em função de fatores diversos. Entre esses fatores, destacam-se:

- (i) a inviabilidade de se estabelecer uma correlação entre o valor de género gramatical e
  - a. a classe semântica do nome, como nos nomes de entidades não animadas: pt. [mar]<sup>MASC</sup> (vs. fr. [mer]<sup>FEM</sup>), pt. [dor]<sup>FEM</sup>; pt. [arroz]<sup>MASC</sup>, pt. [roz]<sup>FEM</sup>; pt. [sal]<sup>MASC</sup>, pt. [cal]<sup>FEM</sup>; e
  - b. a classe formal da fronteira direita dos nomes, nomeadamente a natureza do índice temático, quando existe e/ou está explícito, e o facto de se tratar de um tema  $\emptyset$  ou de um nome atemático.

Como assinala Ferreira (2019), os desvios são tanto mais acentuados quanto o nome (não especificamente sufixado) se caracteriza por ser atemático, de tema  $\emptyset$  ou de tema *-e*, uma vez que este índice temático é o mais ambivalente<sup>3</sup>. Importa apurar se assim também ocorre em nomes sufixados.

O quadro de referência das classes temáticas dos nomes do Português utilizado neste estudo é o de Mota (2020). Nele, os nomes distribuem-se por três grandes classes temáticas:

- (i) nomes de tema *-a, -o, -e*;
- (ii) nomes de tema  $\emptyset$ , como os terminados em *-ão* ([ $\text{ẽw}$ ] fonético); e
- (iii) nomes atemáticos.

O quadro seguinte contém os sufixos usados no inquérito distribuídos pelas classes temáticas e pelos respetivos valores de género gramatical envolvidos nos nomes presentes no inquérito aplicado.

Classe temática do N*	Nome deverbal e sufixo		Nome deadjetival e sufixo	
	Tema $\emptyset$ ([ $\text{ẽw}$ ] fonético)	Atemático	Tema <i>-e</i> (invariável)	
Feminino	-ção, -s[z]ão	-agem	-ice	-idade
Masculino	-ão	-	-	-

\*N - Nome

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

No Português hodierno, o género associado a cada um destes sufixos é fixo e invariável. Assim, não há nomes sufixados em *-idade* ou em *-ice* que não sejam FEM, do mesmo modo que não há nomes portadores do sufixo *-ão* que não sejam MASC. No que diz respeito a *-agem*, e não obstante as variações observadas na história da língua no tocante ao género dos nomes em que está presente (cf. Secção 2), hoje em dia apenas subsiste oscilação na atribuição de feminino, ou de masculino (como em francês) a *a/o personagem*. Nesta matéria a língua portuguesa é bastante mais unívoca que outras de

<sup>3</sup> A este respeito, observa Ferreira (2019, p. 198) o seguinte: “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”.

morfologia igualmente rica. Na língua grega, por exemplo, o sufixo *-os* pode ser marcado como MASC, FEM e NEUTRO (MASTROPAVLOU & TSIMPLI, 2011), em função da base a que se conecta, não sendo, portanto, subespecificado quanto ao género.

Os nomes que foram objeto de escrutínio encontram-se explicitados no quadro seguinte. São 30 nomes (24 atestados e 6 possíveis) distribuídos equitativamente por 6 grupos, em cada um dos quais existe uma forma sufixal específica: *-ão, -ção, -s[z]ão, -agem, -ice* e *-idade*.

Derivados	DEVERBAIS				DEADJETIVAIS	
	MASCULINO		FEMININO			
Género	-ão	-ção	-s[z]ão	-agem	-ice	-idade
Sufixos	-ão	-ção	-s[z]ão	-agem	-ice	-idade
Palavras atestadas	abandão arrastão empurrão estício puxão	confraternização evolução preocupação satisfação	adesão alusão confusão decisão revisão	abordagem lavagem passagem	eriancice velhice vigarice	cumplicidade exclusividade falsidade finalidade
Palavras possíveis	-	avisação	-	levantagem tratagem	contentice vitalice	prontidade

**Quadro 2.** Nomes presentes no inquérito aplicado

Como se observa no quadro anterior, este contém diferentes classes de nomes portadores de sufixos, a saber:

- (i) nomes derivados no português em *-ão, -ção, -agem, -ice* e *-idade*;
- (ii) nomes possíveis em *-ção, -agem, -ice* e *-idade*; e
- (iii) nomes portadores de *-s[z]ão* de origem latina, mas encaráveis como igualmente sufixados.

Os nomes selecionados não têm todos o mesmo estatuto dentro da língua portuguesa. Desde logo, os nomes em *-s[z]ão* têm origem latina, mas não deixam de ser sufixados e facilmente relacionáveis com as respetivas bases verbais. Ademais, trata-se de nomes cujas versões românica e inglesa são de circulação internacional, pelo que não é despendida a probabilidade de contato com configurações próximas da portuguesa. Como em inglês os nomes equivalentes não são marcados por género gramatical, a importância de tal eventual *input*, no caso mais implícito que explícito (sobre as manifestações e implicações deste, ver Martins, 2022), só se fará sentir em falantes com contacto com alguma língua românica.

De todos os nomes e sufixos selecionados, apenas os que são portadores de *-ão* são marcados como MASC; todos os demais são FEM. Como tem sido abundantemente defendido, em várias línguas o género gramatical MASC pode ser encarado como o género gramatical ‘por defeito’<sup>4</sup> (HARRIS, 1991; ALARCÓN, 2004; WHITE *et al.* 2004; MARIOTTO & LOURENÇO-GOMES 2013; MARTINS 2015; FERREIRA 2019). Veremos se, perante os dados apurados neste estudo, tais suposições se comprovam, ou não.

<sup>4</sup> Reportando-se a nomes do espanhol de tema em *-o* não especificamente sufixados, Harris (1991, p. 29) considera que “The class of forms with final vowel (“word marker”) *-o* represents the default case, corresponding to the lack of any lexical mark.” Por sua vez, Ferreira (2019, p. 204) observa que “em todos os grupos de informantes se regista o mesmo padrão, i.e., a maior parte dos desvios recai em formas nominais femininas, havendo, portanto, uma preferência pela atribuição por *default* do valor de género masculino. \*(...) por isso, eu faço **este** **atividade** nos meus tempos livres cada vez mais frequente.” (Chinês.A1-A2).”

A frequência destes nomes na língua portuguesa é, segundo dados do *Corpus do Português – Web / Dialects* (doravante CdP)<sup>5</sup>, bastante diversa (cf. Quadro 3). Como se observa no quadro seguinte, que integra índices de frequência quer do Português do Brasil quer do Português Europeu, apenas dois nomes são marcados, globalmente, com valores de frequência por milhão acima dos 100 mil: *evolução* (453.6) e *decisão* (194.34). Há também nove nomes cujos valores de frequência globais por milhão são superiores a 10 mil: *passagem* (88.13); *preocupação* (67.21); *abordagem* (45.10); *satisfação* (44.66); *finalidade* (42.74); *revisão* (38.39); *confusão* (37.33); *adesão* (26.62) e *lavagem* (13.27). Todos os demais 13 nomes apresentam valores de frequência lexical por milhão abaixo de 10 mil. Na secção 4, observaremos em que medida os valores de frequência interferem, ou não, na atribuição de género (sobre os efeitos da frequência em múltiplos aspetos da *performance* linguística veja-se Bybee (2007), entre outros).

Sufixos	Valores de freq. por milhão			
	Nomes	Geral	Portugal	Brasil
-ão	Abanão	0.20	0.91	0.03
	Arrastão	0.82	1.28	1.24
	Esticão	0.13	0.62	0.01
	Empurrão	1.54	2.97	1.85
	Puxão	1.62	1.45	2.63
-ção	Confraternização	2.11	2.30	2.22
	Evolução	453.6	270.2	558.4
	Preocupação	67.21	65.58	67.27
	Satisfação	44.66	43.13	40.13
-agem	Abordagem	45.10	47.40	34.44
	Lavagem	13.27	14.61	17.34
	Passagem	88.13	97.84	88.59
-s[z]ão	Adesão	26.62	31.47	18.23
	Alusão	7.07	6.60	6.39
	Confusão	37.33	50.70	41.96
	Decisão	194.34	186.22	195.54
	Revisão	38.39	38.53	27.72
-ice	Criancice	0.12	0.14	0.16
	Velhice	9.50	10.74	9.35
	Vigarice	0.59	1.64	0.53
-idade	Cumplicidade	6.83	10.78	7.79
	Exclusividade	8.57	7.73	11.20
	Falsidade	8.05	5.68	8.72
	Finalidade	42.74	28.35	40.10

Quadro 3. Valores de frequência dos nomes sufixados por milhão segundo dados do *Corpus do Português* (CdP)

Como já referido, a base empírica deste estudo resultou da recolha de respostas de um inquérito aplicado junto de aprendentes tardios de PLNM. Entre as respostas obtidas,

<sup>5</sup> O *Corpus do Português* (CdP) - *Web / Dialects* (disponível para consulta 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 (Brasil, Portugal, Angola e Moçambique) e permite fazer pesquisas por palavras e aceder aos respetivos índices de frequência.

verificou-se uma predominância de informantes de LM chinesa. Como a língua materna dos aprendentes inquiridos é maioritariamente o mandarim, a diversidade tipológica de línguas-fonte torna inviável averiguar a eventual influência da língua mãe (e os graus de transferência envolvidos) dos aprendentes no processo de atribuição de valores de género aos nomes sufixados. Optou-se, assim, por trabalhar neste estudo unicamente os inquéritos submetidos por falantes de LM chinesa.

Para a análise e discussão dos dados recolhidos, teve-se em conta o nível da turma frequentada pelo informante bem como as especificidades dos itens nominais sufixados, tendo em vista responder às seguintes questões:

- (i) é possível aferir, nas diferentes fases do desenvolvimento interlinguístico dos aprendentes, padrões de atribuição de valores de género em função da classe temática, do sufixo derivacional e respetivo valor de género do item nominal?
- (ii) existe correlação entre acertos/desvios mais acentuados atendendo ao grau de representatividade do sufixo na língua-alvo, como *-ção* face a *-ão*, *-s[z]ão* e *-agem*, ou como *-idade* face a *-ice*?
- (iii) existe correlação entre a incidência de acertos/desvios e frequência de uso no *input* do item nominal?

## 2. Atribuição de Género Gramatical e influência do sufixo

Quaisquer que sejam os quadros teóricos usados sobre o peso que os radicais *e/ou* os sufixos têm no género gramatical dos nomes em estudo (associando o género ao radical *e/ou* ao sufixo), é inequívoco que, em Português, nos nomes sufixados a determinação do seu género gramatical está fortemente dependente do sufixo usado: nos derivados de *[perna]*<sub>FEM</sub>, nome feminino, *[pernita]*<sub>FEM</sub> e *[pernoca]*<sub>FEM</sub> são igualmente femininos, mas *[perni]*<sub>MASC</sub> e *[pername]*<sub>MASC</sub> são masculinos. Os nomes derivados em *-il* são masculinos (cf. o *cabril*, o *carril*, o *gatil*, o *gradil*, o *poldril*, o *trombil*), mesmo tendo por base radicais de nomes femininos (a *cabra*, a *grade*, a *tromba*); também os nomes derivados em *-ame* são masculinos (o *cordame*, o *velame*), podendo a base ser feminina (a *corda*, a *vela*).

Assim, como afirmam Mastropavlou & Tsimpli (2011):

[...] suffixes are stored in the mental lexicon carrying a gender feature specified for a value which is inherited from nouns they frequently and productively co-occur. In that sense and in line with most lexicalist models of morphology (see Lieber 1992; Selkirk 1982, among others), stems and affixes are both stored in the lexicon, specified with respect to a number of features. (MASTROPAVLOU & TSIMPLI, 2011, p.52)

A atribuição do género gramatical canónico a palavras possíveis ou a pseudo-palavras, registada no estudo levado a cabo por estas autoras gregas, corrobora a assunção de que o género associado a um sufixo será ‘herdado’ do dos nomes nos quais ele mais frequentemente e produtivamente ocorre. Para além da frequência e da consistência do *input* envolvido, vários estudos apontam para a consistência de relações (final fónico-lexical de palavra e género, configuração fónico-lexical do artigo e género do nome) e em que assentam as representações que os falantes memorizam (cf. TAFT & MEUNIER, 1998, entre outros), e que estão na base da atribuição das marcas de género a um dado nome<sup>6</sup>.

<sup>6</sup> “The system is trained on a set of input-output correspondences, including both regular and irregular ones, and from this a set of input units becomes linked (via a layer of hidden units) to a set of output units in a particular way, with a range of different link-strengths depending on the frequency of the particular

Em Português, a presença, no nome, de um sufixo derivacional está associada a um valor de género específico e não variável. Com efeito, e à exceção dos sufixos *-ista* e *-nte* que geram produtos nominais com dois valores de género<sup>7</sup>, todos os demais sufixos estão associados a um único valor de género, masculino ou feminino. Assim, assumimos conjuntamente com Ambadiang (1999, p. 4876) que “[e] affijo 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 quadro que se segue traça uma panorâmica da distribuição dos valores de género feminino e masculino pelas diferentes classes temáticas dos nomes em Português, estabelecidas em função dos sufixos nelas presentes.

Classes temáticas	Género do N <sup>o</sup> derivado		MASC	FEM
	Sufixo			
Tema -o	-ago, -ado, -alho, -ário, -clo, -cino, -ismo, -mento, -tório,		[O] calorção, ducado, ramalho, preçário, vinhedo, cinzeiro, provincianismo, conhecimento, dormitório,	
Tema -e	-ame, -ete, -ote, -ume		[O] cordame, joguete, ralhete, capote, malote, azedume	
	-dade, -ice, -itude, -ose			[A] idiotice, tolice, magnitude, penalidade, universalidade, granulose
Tema -a	-aça, -ada, -alha, -aria, -eira, -eta, -eza, -ia, -ina, -ncia, -ura, -dura			[A] ameaça, fumaça, garfada, gentalha, perfumaria, lancheira, cruzeta, certeza, alegria, nectarina, distância, pintura, queimadura
	-oma		[O] fibroma, pleuroma	
Atemático	-agem			[A] aprendizagem, gatunagem
Tema o	-ez			[A] pacatez, sensatez
	-ção, -s[ç]ão			[A] admiração, decisão
	-ão		[O] empurrão, esticão, puxão	
	-dor, -al, -il, -ês		[O] aparador, separador, laranjal, quintal, gatil, eduquês	

\*N - Nome

Quadro 4. Classes temáticas dos sufixos e marcas de género

Como se verifica, no que toca à marcação de género, existem dois grandes grupos de nomes derivados em Português. Um em que os nomes são:

- (i) MASC:
  - (i<sup>1</sup>) nomes de tema em -o átomo e nomes de tema em -a sufixados em -oma;
  - (i<sup>2</sup>) nomes de tema em -e; e
- (ii) FEM (nomes de tema em -a, tema em -e e nomes atemático em -agem).

Um outro grupo de nomes é o dos de tema o, que podem ser de género invariavelmente MASC (-dor, -al, -il, -ês, -ão) ou de género invariavelmente FEM (-ção, -s[ç]ão, -ez).

relationship. When a new input is given, the system uses these links to activate an appropriate output.” (TAFT & MEUNIER, 1998, p. 25)

<sup>7</sup> Estes nomes ambíguos serão objeto de um outro estudo, uma vez que, sendo ‘comuns de dois géneros’, a especificação de género coloca problemas de concordância.

Os nomes de tema em -o e em -a são, no cômputo global, os mais unívocos quanto ao género gramatical, estando associado aos primeiros o género MASC e aos segundos o género FEM. Assim é com nomes não derivados, e a situação aplica-se também aos que são sufixados. Na língua portuguesa, apenas um grupo residual de nomes em -o é marcado como FEM (a *libido*, a *moto*, a *nao* (var. antiga de *navi*), a *virago*, a *rádio*, a *foto*, a *químio* [por *quimioterapia*]), a *pólio* [por *poliomielite*]). Seguem-se os nomes de tema sufixado em -a, maioritariamente FEM, e residualmente MASC, quando o sufixo é -oma. Cerca de seis centenas de nomes em -a, quase todos de origem grega, são MASC (cf. *dia*, *drama*, *mapa*, *profeta*, *poeta*, *genoma*, *telegrama*)<sup>8</sup>. Situação idêntica ocorre em espanhol (HARRIS, 1991).

Também os nomes de tema em -e podem ser MASC ou FEM. Na verdade, a presença de um índice temático -e não fornece pistas sobre o género do derivado, sendo este MASC nos nomes em -ame (o *cordame*) ou -ume (o *azedume*) e FEM nos nomes em -idade (a *portugalidade*) e em -ice (a *doutorice*, a *vedetice*) (cf. Quadro 4).

Os nomes atemáticos em -agem são FEM, na língua portuguesa atual (a *aprendizagem*, a *viagem*). Mas os nomes portadores dos sufixos românicos equivalentes ao port. -agem são masculinos. Assim acontece em Espanhol, com -aje (el *aprendizaje*, el *viaje*), em Francês -age (un *apprentissage*, le *voyage*) e em Italiano -aggio (un *viaggio*). Ou seja, a língua portuguesa diferencia-se neste aspeto, o que pode dificultar a atribuição de género gramatical a nomes em -agem por parte de aprendentes cujas línguas maternas sejam o Castelhana, o Francês ou o Italiano. Todavia, na história da língua portuguesa, períodos houve em que se registava variação no género dos nomes sufixados em -agem. Até ao século XVI documentam-se palavras em -agem marcadas como masculinas, como o testemunho do séc. XIV abona (“ho linhagem umanall”, *Coronica Troiana* 21, p. 18-20), tendo a fixação sido progressivamente consolidada de então para cá. Em francês e em espanhol, preserva-se o género masculino nas palavras cujas terminações remontavam a -aticu-, como em Francês le *passage* (MASC), em Espanhol el *pasaje* (MASC). Em Português, a palavra equivalente é feminina — a *passagem* (FEM) —, pois nesta língua a terminação -agem (<-age) ter-se-á deixado contaminar por -agem, com origem em -agine (*imagine*- > pt. *imagem*, fr. *image*, esp. *imagen*, todas do género FEM), generalizando-se, assim, o género feminino.

Os nomes em -ção e em -ão, não obstante a presença de um -o na sua fronteira direita, também são não indexais do género gramatical do nome em que ocorrem. Em teoria, a presença de -o poderia apontar para o género MASC, mas tratar-se-ia de uma pista enganadora, pois o género dos nomes com -ção [švĩ] e -ão [ĩvĩ] têm um comportamento ambivalente quanto ao género. No caso de -ção, o português apresenta o género FEM<sup>9</sup> comum aos equivalentes do fr. -ation, do esp. -ación, do it. -azione. Já com -ão há atribuição do género MASC. (pt. o *empurrão*, esp. el *empujon*). Consequentemente, o facto de estes nomes terem alguma ressonância com os derivados em -ção pode levar os aprendentes a marcar como FEM o nome *cabeção*, que é MASC. Poder-se-á assumir que estes não são sufixos cujo valor de género seja univocamente identificado e atribuído por aprendentes de PLNM.

Em suma, a ambivalência que se observa nos nomes sufixados de tema -a ou -e não se regista nos nomes sufixados de tema -o, nos atemáticos em -agem e nos de tema o. Os

<sup>8</sup> Não é feita, aqui, menção aos nomes sufixados em -ista pelo facto de serem ambíguos: o/a *artista*; o/a *motorista*.

<sup>9</sup> Como afirma Rodrigues (2016, p. 71) “o sufixo -ção categoriza o produto que é por ele formado como de género feminino (a *destruição*, a *avaliação*).”

nomes de tema *o* são sistematicamente FEM quando sufixados em *-ez*, *-ção* e *-s[z]ão*, e sistematicamente MASC quando sufixados em *-ão*, *-dor*, *-al*, *-il*, *-ês* (cf. Quadro 4).

Assim, sendo arbitrário, o gênero gramatical é mais imprevisível nos nomes não derivados do que nos sufixados, uma vez que nestes a presença de um dado formativo sufixal assegura uma classe de gênero invariável e sistemática (FEM/MASC). Veremos em que medida esta realidade se traduz por acertos ou desvios significativos nos vários casos.

### 3. Procedimentos metodológicos

Os dados que constituem a base empírica do presente estudo foram extraídos de um inquérito aplicado *online*<sup>10</sup> e no qual se solicitou a aprendentes tardios de PLNM que identificassem o valor de gênero de 30 nomes sufixados (24 formas atestadas e 6 possíveis), distribuídos equitativamente por 6 grupos que, por sua vez, foram estipulados em função do sufixo derivacional: *-ão*, *-ção*, *-s[z]ão*, *-agem*, *-ice* e *-idade* (cf. Quadro 2).

Cada item nominal integra uma construção frásica distinta, tendo o informante de selecionar, em função do nome sufixado, a forma adequada do artigo definido (*o*, *a*, *os*, *as*) que, com ele, deve concordar em gênero e em número<sup>11</sup>. À exceção dos nomes sufixados em *-ão*, que são marcados como MASC, os demais nomes sufixados são marcados como FEM (cf. Quadro 2). Foram, assim, elaboradas construções frásicas com nomes sufixados em *-ão* (cf. (1a)); em *-ção* (cf. (1b)); em *-s[z]ão* (cf. (1c)); em *-agem* (cf. (1d)); em *-ice* (cf. (1e)) e em *-idade* (cf. (1f)):

- (1)
- Na semana passada, *o(a/os/as)* 'puxões de orelhas' que a mãe deu ao Filipe foram tão eficazes que agora ele já come a sopa toda.
  - A loja da Maria Alice preocupa-se com *o(a/os/as)* *satisfação* dos seus clientes e, por isso, só vende produtos de qualidade.
  - As estradas de acesso à Torre da Serra da Estrela foram fechadas e instalou-se *o(a/os/as)* *confusão* no trânsito!
  - Este detergente é ótimo para *o(a/os/as)* *lavagem* de cobertores e tapetes.
  - A atitude do João revelou *o(a/os/as)* tal *criançice* de que todos falavam na semana passada!
  - Qual é *o(a/os/as)* *finalidade* do projeto proposto hoje na Assembleia Geral?

Em cada inquérito submetido, recolheram-se ainda informações com vista à caracterização do perfil dos informantes, nomeadamente no que concerne ao seu perfil sociolinguístico: língua(s) materna(s); nível de proficiência linguística do português; outras línguas não-maternas conhecidas além do português e respetivo nível de proficiência; sexo; idade; e instituição de ensino frequentada aquando do preenchimento do inquérito.

<sup>10</sup> O inquérito foi elaborado na plataforma Microsoft Forms<sup>®</sup> e o respetivo *link* para preenchimento foi enviado aos informantes, via *email*. O tempo médio de preenchimento registado foi de 38 minutos e 14 segundos.

<sup>11</sup> O uso de formas nominais masculinas e femininas, no singular e no plural permite averiguar se o informante reconhece o valor de gênero do item, bem como o respetivo valor de número. Neste trabalho em particular, e em função dos objetivos estipulados, analisaram-se unicamente as respostas facultadas em função da atribuição de valor de gênero nominal (convergente ou divergente) aos nomes sufixados.

O questionário foi preenchido por aprendentes que se encontravam a frequentar cursos de português para estrangeiros, ministrados em instituições de ensino superior em Portugal, na China Continental e em Macau. Para este estudo, selecionaram-se as respostas facultadas por falantes nativos de chinês, distribuídos por três níveis de proficiência de português: B1, B2 e C1, que correspondem aos níveis do QECRL da turma frequentada no momento da recolha.

Os informantes são de ambos os sexos (74% do sexo feminino e 26% do sexo masculino) e têm uma média de idades de 20,2 anos (distribuídos entre os 19 e os 31 anos). O Quadro 5 apresenta o conjunto total de dados empíricos apurados, distribuídos em função do nível QECRL da turma frequentada e da classe temática dos nomes sufixados presentes no inquérito.

Informantes		Respostas		Classe temática do nome e sufixo					
				Tema <i>o</i> ([f̥w̃] fonético)		Atemático		Tema <i>e</i> (invariável)	
LM	Nível QECRL	#	#	-ão	-ção	-s[z]ão	-agem	-ice	-idade
Chinês	B1	38	1140	190	190	190	190	190	190
	B2	48	1440	240	240	240	240	240	240
	C1	18	540	90	90	90	90	90	90
Σ		104	3120	520	520	520	520	520	520

Quadro 5. Caracterização da base empírica do estudo: informantes e respostas obtidas por classe temática do nome e respetivo sufixo

Com vista a aferir padrões de aquisição na atribuição de valores de gênero aos nomes sufixados em português que se correlacionem, por um lado, com o perfil do aprendente e, por outro, com as propriedades dos itens nominais sufixados, tiveram-se em conta as seguintes variáveis na análise dos dados empíricos:

- nível de proficiência dos informantes em português (nível QECRL);
- classe temática do nome e respetivo sufixo derivacional; e
- índices de frequência lexical do nome derivado.

### 4. Resultados e discussão

A base empírica deste trabalho é constituída por 3120 ocorrências de atribuição de valor de gênero gramatical, entre as quais se registaram 557 desvios. Assim, e em função do número absoluto de ocorrências, verificou-se que cerca de 18% correspondem a desvios de atribuição de valor de gênero nominal (cf. Quadro 6). Por conseguinte, de um modo geral, estes dados são reveladores de que, entre os níveis QECRL B1 e C1, uma parte substancial (72%) dos aprendentes de LM chinesa seleciona adequadamente o valor de gênero do item nominal sufixado.

Nível QECRL	B1	B2	C1	Σ
# Respostas	1140	1440	540	3120
# Desvios de AGN*	268	213	76	557
% relativa de desvios de AGN	23,5%	14,7%	14,1%	17,9%

\*AGN – Atribuição de gênero nominal

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

Para averiguar se a incidência de desvios diminui em fases mais avançadas do desenvolvimento interlinguístico dos participantes, calcularam-se, para cada segmento da amostra, as respetivas percentagens relativas de desvios em função do número absoluto de respostas submetido por nível QECRL (cf. Quadro 6 e Gráfico 1). Verificou-se que o nível B1 apresenta o maior índice de desvios (23,5%), enquanto os níveis B2 e C1 registam incidências de desvios mais baixas (14,7% e 14,1%, respetivamente).

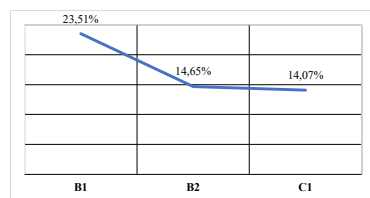


Gráfico 1. Percentagens relativas de desvios por nível QECRL

Embora se verifique uma diminuição da proporção de desvios nos níveis intermédio (B2) e avançado (C1), relativamente ao nível B1, é de assinalar que os valores percentuais apurados nos três segmentos da amostra empírica não são substancialmente diferentes. Em termos quantitativos, o desempenho linguístico dos informantes de B2 e de C1, no que à associação de valor de género aos nomes derivados diz respeito, é praticamente similar, e o diferencial relativamente ao desempenho dos alunos de B1, situa-se abaixo de 10%, pelo que não pode ser considerado avultado.

Além disso, a percentagem relativa de desvios registada no nível C1 não reflete uma melhoria significativa do desempenho destes aprendentes relativamente aos outros níveis (B1 e sobretudo B2) e parece indicar, neste segmento, um certo efeito de estabilização da aprendizagem da atribuição de género nominal aos nomes sufixados em Português. Por conseguinte, estes resultados são reveladores de que mesmo em fases mais avançadas do seu desenvolvimento interlinguístico, os aprendentes de LM chinesa não reconhecem adequadamente o valor de género de todos os nomes sufixados apresentados.

De forma a apurar se existe correlação entre a frequência de desvios e as classes temáticas dos nomes e respetivos sufixos derivacionais, procedeu-se ao cálculo das percentagens relativas de desvios por segmentos da amostra, calculadas em função do número absoluto de respostas submetido por conjunto de nomes sufixados (cf. Quadro 7 e Gráfico 2). Assim, procura-se determinar se a incidência de respostas desviantes depende da classe temática e respetivo sufixo do nome.

Classe Temática	Sufixo	Nível QECRL						TOTALS					
		B1		B2		C1		n.º de respostas	n.º de desvios	%			
		n.º de respostas	%	n.º de respostas	%	n.º de respostas	%						
o ([ð]) fonético	-ão	190	12,2	18,2	241	10,0	42,9	90	0,0	42,9	220	26,2	20,4
	-ão	190	12,2	16,3	241	10,0	42,9	90	0,0	42,9	220	17,7	2,1
	-s[e]ão	190	19	10	241	13	5,4	90	1	1,1	220	33	6,8
Alambrico	-agem	190	21	11,1	241	18	7,5	90	2	2,2	220	41	7,9
	-ade	190	79	41,6	241	73	30,4	90	31	34,4	220	185	84,0
	-idade	190	15	7,9	241	1	0,4	90	0	0	220	15	7,1
Σ		1440	268	23,5	1440	212	14,7	540	14,1	14,1	3128	557	17,9

Quadro 7. Distribuição do número absoluto de respostas, de desvios e respetivas percentagens relativas apuradas por nível QECRL e sufixo derivacional

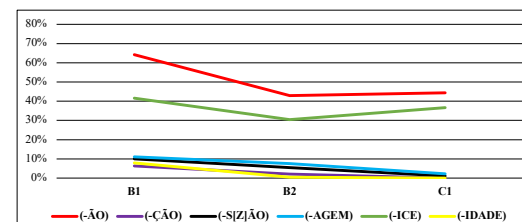


Gráfico 2. Distribuição das percentagens relativas de desvios por sufixo derivacional em função do nível QECRL

Segundo os dados cartografados no Gráfico 2 e os do Quadro 7, a maior proporção de desvios, em todos os segmentos da amostra, regista-se nos nomes sufixados em *-ão* (50.6%) e em *-ice* (35.6%). Em contrapartida, os sufixados em *-agem*, *-ção*, *-[z]são* e *-idade* registam índices de desvio reduzidos: *-agem* (7.9%), *-ção* (3.3%), *-[z]são* (6.4%) e *-idade* (3.1%). Por conseguinte, não existe uma correlação direta entre a classe temática do nome e incidências de desvios. Em contrapartida, os dados revelam padrões distintos atendendo ao sufixo derivacional do nome em causa.

Em termos gerais, os elevados índices de desvios registados nos nomes masculinos sufixados em *-ão* (50.6%) por oposição aos índices de desvios apurados nos restantes nomes, que são todos uniformemente femininos, contrariam, em certa medida, a tendência amplamente assinalada em muitos dos estudos realizados sobre a aquisição tardia da categoria de género gramatical (BRUHN DE GARAVITO & WHITE, 2002; ALARCÓN, 2004; WHITE *et al.* 2004; MARIOTTO & LOURENÇO-GOMES 2013; MARTINS 2015; FERREIRA 2019). Com efeito, nestes trabalhos, constata-se que uma parte substancial dos desvios recai sobre nomes femininos aos quais os aprendentes atribuem o valor de género masculino. Esta tendência corresponde a uma estratégia, adotada pelos aprendentes, de marcação por defeito do género masculino, a forma não-marcada, e é característica do processo de assimilação tardia da categoria de género gramatical, sobretudo em fases iniciais da aprendizagem<sup>12</sup>.

Assim, nos nomes sufixados em *-ção*, *-[z]são* e *-ão*, a presença da vogal *-o* na fronteira direita do nome poderia conduzir o aprendente a assumir que se trata de uma forma masculina, já que o facto de haver, em português, uma correlação parcial entre índices temáticos *-o* e *-a* e valores de género MASC e FEM, respetivamente, poderia contribuir para a assunção, por parte destes aprendentes, de que estas vogais

<sup>12</sup> Assim acontece com os nomes femininos sufixados em *-ice*, nos quais se regista, em B1, uma proporção maior de desvios (41.6%), por oposição aos valores registados nos restantes níveis (30.4% em B2 e 36,7% em C1). Porém, esta estratégia de marcação de masculino por defeito não é tão expressiva nos femininos que terminam em <o>.

correspondem a marcadores morfológicos de género gramatical (cf. Secção 2)<sup>13</sup>. Contudo, não é tão expressiva a associação do MASC a nomes sufixados em *-ção* e *-s[ç]ão* como é a associação do FEM aos nomes sufixados em *-ão* (cf. Gráfico 2). Face a estes resultados, colocamos a hipótese de que a incidência dos desvios nestes nomes esteja correlacionada com a representatividade do sufixo no *input* linguístico ao qual o aprendente está exposto. Ou seja, postula-se que *-ção* e *-s[ç]ão* estejam mais representados no *input* e, por esse motivo, a incidência de desvios de atribuição de valores de género é mais reduzida nos respetivos nomes sufixados. Em contrapartida, a presença no léxico da língua-alvo, o Português, de nomes femininos terminados em <ão>, como, por exemplo, a *região*, a *mão*, a *opinião*, a *união* e a *pensão*, e de todos os sufixados em *-ção*, poderá contribuir para gerar ambiguidade no *input* e, por conseguinte, conduzir o aprendente a assumir que itens formalmente idênticos são marcados pelo mesmo valor de género. Este processo de analogia de uma determinada marca de género a uma determinada configuração formal (no caso <-ão> e género FEM) está atestado na atribuição de género a palavras importadas (POPLACK; PÓUSADA; SANKOFF, 1982), pelo que poderá também não ser descartado.

Entre os nomes de tema em *-e*, sufixados em *-idade* e *-ice*, registam-se índices de desvios diferentes. Mais precisamente, o sufixo *-ice* apresenta, em todos os segmentos da amostra empírica, frequências de desvio de atribuição de género substancialmente mais elevadas (36.7%) do que as registadas nos nomes sufixados em *-idade* (3.10%) (cf. Quadro 7). O diferente desempenho linguístico observado pode estar correlacionado, de novo, com a representatividade do sufixo derivacional em causa. Com efeito, segundo MOITA *et. al.* (2010), na neologia do português, o sufixo *-ice* é menos produtivo do que *-idade*, e é um sufixo com uma configuração autóctone na língua portuguesa, mais opaco e idiossincrático (RIO-TORTO & RODRIGUES, 2016, p. 138). Assim, a baixa representatividade deste sufixo no *input* a que estes aprendentes estão expostos poderá contribuir para uma maior incidência de desvios de atribuição de género a estes nomes derivados. Ademais, sendo *-idade* um sufixo ‘internacional’, o sufixo e os nomes em que ocorre têm muito mais probabilidade de fazerem parte do *input* dos inquiridos, do que a maior parte dos nomes em *-ice*.

Os itens pertencentes à classe atemática, sufixados em *-agem* e marcados como FEM, apresentam frequências de desvios relativamente baixas em todos os segmentos da amostra empírica analisada (7.9%). Como são formas relativamente opacas no que à associação de valores de género diz respeito (por serem atemáticas), seria expectável que apresentassem taxas de desvio consideravelmente altas, tal como atestam alguns estudos sobre a aquisição de género gramatical em PLNM (FERREIRA 2011, 2019; MARTINS 2015; PINTO 2017).

Portanto, a não atribuição sistemática de MASC por defeito aos nomes derivados marcados como FEM (como são os sufixados em *-agem*, *-ção*, *-s[ç]ão* e *-idade*) leva-nos a equacionar a hipótese de que, em fases mais avançadas do desenvolvimento interlinguístico, os informantes estarão menos suscetíveis à ‘estereotipagem linguística’ na atribuição de género, e serão, assim, capazes de aferir os valores de género em função

<sup>13</sup> Pinto (2017, p.100) observou, nos seus dados, que os nomes terminados em *-ão* são tendencialmente associados ao género masculino pelos aprendentes de PLNM, podendo este comportamento justificar-se pelo facto de os alunos interpretarem o *-o* final como um índice temático. Por seu turno, Ferreira (2021), ao analisar desvios morfológicos correlacionados com a atribuição de género nominal, verificou que os aprendentes, independentemente da sua LM, tendem a associar as vogais finais *-o* e *-a* aos valores de género MASC e FEM, respetivamente, e adaptam a forma morfológica dos itens, quer gramaticais quer lexicais, em função deste padrão, conforme ilustram os exemplos seguintes retirados desse estudo: “O *minho quarto* (Alemão.A1-A2); *A minha amiga melhora* (Inglês.A1-A2); os *semestros* (Alemão.B1-B2)” (FERREIRA, 2021, p.223).

da estrutura formal do nome. Por ‘estereotipagem linguística’ entende-se a associação ao nome, e aos itens que com ele estabelecem relações de concordância nominal, do valor de género MASC por defeito, por este ser considerado, em Português como em outras línguas românicas, o valor não-marcado. Além disso, estes dados também indiciam que, em determinados conjuntos de nomes sufixados, os aprendentes memorizaram o respetivo valor de género, podendo essa associação estar correlacionada quer com a representatividade do sufixo quer com a respetiva frequência dos nomes sufixados no *input*.

Considerada a proporção dos desvios por sufixo derivacional, procurou-se, em seguida, apurar se existe correlação entre as taxas de desvios registadas e a frequência dos nomes derivados (segundo dados de frequência apurados no CdP; cf. Secção 2). Neste contexto, coloca-se a hipótese de que nomes menos frequentes no *input* registam uma maior incidência de desvios do que os nomes mais frequentes, já que a frequência lexical pode contribuir, favoravelmente, para a assimilação do item nominal e das suas propriedades morfossintáticas no léxico mental do aprendente. Em conformidade, em nomes não atestados na língua portuguesa, a incidência de desvios será substancialmente elevada.

No quadro seguinte, registam-se os nomes sufixados que foram alvo de escrutínio neste estudo, distribuídos por ordem decrescente dos respetivos valores de frequência gerais (que incluem os valores apurados no subcorpus de Portugal e do Brasil) por milhão registados no CdP.

	Nomes	Valores de frequência por milhão	
+ frequente	evolução	453.6	
	decisão	194.34	
	passagem	88.13	
	prosecução	67.21	
	abordagem	45.10	
	satisfação	44.66	
	finalidade	42.74	
	revisão	38.39	
	confusão	37.33	
	adesão	26.62	
	lavagem	13.27	
	velhice	9.50	
	exclusividade	8.57	
	falsidade	8.05	
	alusão	7.07	
	cumplicidade	6.83	
	confraternização	2.11	
	puxão	1.62	
	empurrão	1.54	
	arrastão	0.82	
	vigarice	0.59	
	- frequente	abanão	0.20
		estício	0.13
		cranceice	0.12

Quadro 8. Distribuição das formas sufixadas atestadas por valores de frequência por milhão registados no CdP

Da leitura do quadro, constata-se que todos os nomes sufixados em *-ão* possuem índices de frequência inferiores a dois mil por milhão (*puxão* (1.62); *empurrão* (1.54), *arrastão* (0.82), *abanão* (0.20) e *estício* (0.13)), sendo que entre os nomes sufixados em *-ção*, apenas *confraternização* regista valores de frequência que rondam os dois mil por

milhão (2.11), e os restantes apresentam valores de frequência consideravelmente superiores (*evolução* (453.6); *preocupação* (67.21); e *satisfação* (44.66)). Estes indicadores parecem, assim, favorecer a hipótese de que os nomes sufixados em *-ção* são mais representativos no *input* do que os sufixados em *-ão*. Do mesmo modo, constata-se que, entre os sufixados em *-idade*, grande parte apresenta valores de frequência consideravelmente altos, por oposição aos derivados em *-ice* que, globalmente, registam valores de frequência mais baixos.

De forma a aferir se existe uma correlação entre a frequência dos itens lexicais e frequência de desvios de atribuição de género, apuraram-se, para cada segmento da amostra, as percentagens relativas de desvios.

Para a análise dos dados, optou-se por considerar, em primeiro lugar, os índices de desvios nos nomes cuja frequência lexical é superior a 10 mil por milhão, segundo o Cdp.

Nome	Valores de Frequência Cdp (por milhão)	Nível QECRL									TOTAIS		
		B1			B2			C1			# de respostas	# de desvios	%
		# de respostas	# de desvios	%	# de respostas	# de desvios	%	# de respostas	# de desvios	%			
exclusivo	453.6	38	3	7.9	48	0	0	18	0	0	104	3	2.9
decisão	194.34	38	4	10.5	48	1	2.1	18	1	5.6	104	6	5.8
ansiedade	58.13	38	1	2.6	48	4	8.3	18	0	0	104	5	4.8
preocupação	67.21	38	2	5.3	48	1	2.1	18	0	0	104	3	2.9
abstração	45.10	38	5	13.2	48	4	8.3	18	2	11.1	104	11	10.6
antifédo	44.66	38	1	2.6	48	1	2.1	18	0	0	104	2	1.9
finalidade	42.74	38	4	10.5	48	0	0	18	0	0	104	4	3.8
revista	38.39	38	5	13.2	48	2	4.2	18	0	0	104	7	6.7
confusão	37.53	38	5	13.2	48	1	2.1	18	0	0	104	6	5.7
alçada	26.62	38	3	7.9	48	5	10.4	18	1	5.6	104	9	8.7
livragem	13.27	38	4	10.5	48	1	2.1	18	0	0	104	5	4.8
%	95.58	28	3.4	8.4	48	2	4.2	18	0.36	2	104	5.7	5.5

Quadro 9. Distribuição do número absoluto de respostas, desvios e respetivas percentagens relativas apuradas por nível QECRL nos itens com valores de frequência superiores a 10 mil por milhão

Conforme se verifica no Quadro 9, o valor médio ( $\bar{X}$ ) da frequência lexical destes itens é de 95.58 por milhão e o valor médio de desvios registado em todos os segmentos da amostra é de 5.7 desvios por item lexical. Em termos percentuais, apurou-se que, em todos os segmentos, entre os nomes deste conjunto, cerca de 5,5% apresentam desvios de atribuição de género.

No que diz respeito às percentagens relativas de desvios, e atendendo aos dados apurados em todos os segmentos, os nomes sufixados em *-ção* registam a menor incidência de desvios (*satisfação* (1.9%); *preocupação* (2.9%) e *evolução* (2.9%)), em contraste com os sufixados em *-s[iz]ão* que apresentam, neste conjunto, a maior incidência de desvios (*adesão* (8.7%); *confusão* (7.7%); *revista* (6.7%) e *decisão* (5.8%)). Além disso, estes dados parecem indicar que o facto de um nome apresentar valores de frequência mais elevados não contribui favoravelmente para uma redução significativa do número de desvios de atribuição de género nominal. Com efeito, apesar de *decisão* possuir valores de frequência lexical elevados (194.34 por milhão), apresenta, em termos globais, uma percentagem relativa de desvios ligeiramente superior à registada em *lavagem*, o item que, neste conjunto, apresenta o valor de frequência lexical mais baixo (13.27 por milhão) e cuja percentagem relativa de desvios apurada em todos os segmentos é de 4.8%.

Relativamente à proporção de desvios por segmentos da amostra, regista-se, no nível B1, a maior incidência de respostas desviantes, havendo uma diminuição gradual das taxas de desvios nos níveis subsequentes (B2 e C1). Além disso, verifica-se que em C1 os desvios são residuais, incidindo apenas em três formas derivadas, sufixadas em *-agem* (*abordagem*) e em *-s[iz]ão* (*decisão* e *alusão*), não havendo, também neste

segmento, evidências de que a maior frequência lexical do item determine menos ocorrências de desvios.

Vejam-se, em seguida, os dados relativos à proporção de desvios apurada nos itens cuja frequência lexical é inferior a 10 mil por milhão.

Nome	Valores de Frequência Cdp (por milhão)	Nível QECRL									TOTAIS		
		B1			B2			C1			# de respostas	# de desvios	%
		# de respostas	# de desvios	%	# de respostas	# de desvios	%	# de respostas	# de desvios	%			
velhice	9.50	38	14	36.8	48	12	25	18	7	38.9	104	33	31.7
exclusividade	8.57	38	1	2.6	48	0	0	18	0	0	104	1	0.9
filosofia	8.05	38	1	2.6	48	0	0	18	0	0	104	1	0.9
plano	7.07	38	2	5.3	48	2	4.2	18	0	0	104	4	3.8
competência	6.83	38	2	5.3	48	1	2.1	18	0	0	104	3	2.9
confiteiramente	2.11	38	3	7.9	48	0	0	18	0	0	104	3	2.9
respeito	1.62	38	21	55.3	48	21	43.8	18	0	0	104	51	49.0
capacidade	1.54	38	22	57.9	48	22	45.8	18	6	33.3	104	50	48.1
arrastado	0.82	38	27	71.1	48	18	37.5	18	0	0	104	54	51.9
vigância	0.99	38	15	39.5	48	13	27.1	18	7	38.9	104	35	33.7
plano	0.20	38	20	52.6	48	19	39.6	18	6	31.3	104	46	43.3
estudo	0.13	38	29	76.3	48	21	43.8	18	10	55.6	104	60	57.7
compaixão	0.12	38	15	39.5	48	14	29.2	18	3	15.8	104	31	29.8
%	1.63	38	13.7	36	48	11	22.9	18	4.5	23.2	104	29.2	28.1

Quadro 10. Distribuição do número absoluto de respostas, desvios e respetivas percentagens relativas apuradas por nível QECRL nos itens com valores de frequência inferiores a 10 mil por milhão

Segundo os dados representados no Quadro 10, o valor médio ( $\bar{X}$ ) da frequência lexical destes itens é de 3.63 por milhão, sendo este valor muito inferior ao registado entre os itens de maior frequência (cf. Quadro 9). No que aos desvios de atribuição de valor de género diz respeito, verificou-se que, em média, ocorrem 29.2 desvios por item lexical. Este valor é também exponencialmente diferente do apurado para os itens mais frequentes.

Em termos globais, constata-se que os nomes sufixados em *-ão* e em *-ice* apresentam maiores índices de desvios em relação às restantes formas sufixadas, independentemente dos respetivos valores de frequência lexical. Por exemplo, a percentagem relativa de desvios registada no nome *vigância* é de 33.7% e é ligeiramente superior à apurada para o nome *velhice*, que é de 31.7%, apesar de o primeiro nome registar valores de frequência lexical consideravelmente mais baixos. De facto, o nome *vigância* regista um valor de frequência de 0.59 por milhão, enquanto *velhice* apresenta um valor de 9.50 por milhão (cf. Quadro 10). Do mesmo modo, se atentarmos nos resultados apurados entre os nomes *velhice* e *exclusividade*, cujos valores de frequência lexical são muito próximos, 9.50 e 8.57, respetivamente, verificamos que a incidência global de desvios em *velhice* é de 31.7%, sendo substancialmente superior à registada em *exclusividade*, que é de 2.9%.

Portanto, estes resultados levam-nos a considerar, de novo, que a frequência lexical do item não parece ser claramente determinante para o processo de assimilação dos valores de género. Além disso, não é certo que um item considerado mais frequente no léxico do Português também o seja no *input* ao qual o aprendente está exposto. Com efeito, a esmagadora maioria destes informantes contacta com a língua apenas em contexto instrucional o que condiciona, por sua vez, o contacto com e/ou a utilização de léxico mais diversificado. Do mesmo modo, os próprios recursos instrucionais usados em sala de aula poderão, em certa medida, condicionar o grau de exposição ao *input* da língua-alvo.

Relativamente à proporção de desvios por segmentos da amostra, os dados coligidos revelam que nos níveis mais avançados, B2 e C1, a proporção de desvios por item mantêm-se consideravelmente elevada, sobretudo nos itens sufixados em *-ice* e *-ão*. Ou seja, não se regista uma diminuição expressiva do número de desvios nestas formas sufixadas, o que nos leva a assumir que, mesmo em fases mais avançadas do



desenvolvimento linguístico, o sufixo derivacional será determinante no processo de assimilação do valor de género.

A fim de determinar se a frequência dos itens lexicais é fator condicionante para a ocorrência de desvios, analisaram-se também as frequências de desvios registadas nas formas possíveis, criadas para este estudo. Convém salientar que nem todos os grupos de nomes sufixados apresentam nomes possíveis (cf. Quadro 2). Com efeito, não foram disponibilizadas formas possíveis de derivados em *-ão* e em *-s/z/ão*, pelo que, na classe temática  $\emptyset$  ([ãw̃] fonético) apenas o sufixo *-ção* apresenta uma forma possível.

Nome	Nível QECRL									TOTALS		
	B1			B2			C1			# de respostas	# de desvios	%
	# de respostas	# de desvios	%	# de respostas	# de desvios	%	# de respostas	# de desvios	%			
*avizalde	38	3	7.9	48	3	6.3	18	0	0	104	6	5.8
*matagem	38	6	15.8	48	4	8.3	18	0	0	104	10	9.6
*levantagem	38	5	13.2	48	5	10.4	18	0	0	104	10	9.6
*vitalice	38	21	55.3	48	19	39.6	18	9	50	104	49	47.1
*contentice	38	16	42.1	48	15	31.3	18	5	27.8	104	36	34.6
*prontidade	38	3	7.9	48	0	0	18	0	0	104	3	2.9
$\Sigma$	38	9.6	25.2	48	8.6	17.9	18	2.7	15.1	104	20.9	20.1

Quadro 11. Distribuição do número absoluto de respostas, desvios e respetivas percentagens relativas apuradas por nível QECRL nas formas possíveis e sem registos de frequência lexical

Considerando a incidência de desvios por formas possíveis (cf. Quadro 11), verifica-se que o padrão é semelhante ao apurado entre as formas atestadas. De facto, neste conjunto de itens, a maior proporção de desvios em todos os níveis QECRL regista-se nos nomes sufixados em *-ice*. Em termos percentuais, e em todos os segmentos da amostra, a proporção de desvios registada entre as formas possíveis *\*vitalice* e *\*contentice* é de 47.1% e 34.6%, respetivamente. Em contrapartida, os nomes possíveis sufixados em *-idade* e *-ção* apresentam incidências de desvios relativamente baixas (*\*prontidade* (2.9%) e *\*avisação* (5.8%)), o que nos leva à assunção de que para estas formas sufixadas os informantes, sobretudo de níveis mais avançados, reconhecem, com razoável grau de sucesso, o valor de género que lhes está associado.

## Conclusões

O estudo levado a cabo permitiu identificar padrões relativos à atribuição de valores de género nominal a nomes sufixados por aprendentes tardios de PLNM, falantes nativos de chinês, a frequentar turmas universitárias dos níveis QECRL B1, B2 e C1. Com base nos resultados globais obtidos, constatou-se uma diminuição gradual dos desvios à medida que os aprendentes progredem na aprendizagem do Português, embora se registre, no nível C1, um certo efeito de estabilização da aprendizagem face a B2, já que no nível mais avançado a incidência de desvios não é significativamente mais baixa do que a registada em B2 (cf. Gráfico 1).

Os dados apurados indiciam ainda 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. Em contrapartida, estes dados demonstram que não existe correlação entre a classe temática a que pertencem os sufixos e o processo de assimilação/atribuição de valores de género. Com efeito, 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* (cf. Quadro 7).

A incidência expressiva de desvios sobre os nomes masculinos derivados em *-ão*, por oposição à registada entre os nomes femininos derivados em *-ção* e *-s/z/ão*, leva-nos a concluir que, em níveis de proficiência linguística mais avançados, os aprendentes tendem menos à 'estereotipagem linguística' (tal como acima descrita, selecionando o MASC por defeito) no momento de atribuição do valor de género nominal, e parecem mais sensíveis à estrutura formal do item sufixado e às relações que se estabelecem entre determinados sufixos derivacionais e respetivos valores de género.

No que à frequência lexical diz respeito, constatou-se que independentemente da sua frequência no *input*, os nomes sufixados em *-ão* e *-ice* apresentam maior proporção de desvios, mesmo em níveis mais avançados de proficiência linguística do português. Este padrão aplica-se não apenas a palavras atestadas, mas também a palavras possíveis. Estes resultados são, assim, reveladores de que o sufixo derivacional da palavra tem mais peso na atribuição de um valor de género do que a própria frequência no *input* do nome derivado.

## REFERÊNCIAS

- ALARCÓN, I. The sequential acquisition of L2 Spanish gender marking: Assignment and agreement. *Indiana University Linguistics Club Working Papers Online*, Indiana, v. 4, n.1, p. 1-23, 2004.
- AMBADIANG, T. La Flexión nominal: género y número. In: BOSQUE, Ignacio; DEMONTE, Violeta (Orgs.). *Gramática Descriptiva del Español*. Vol. III. Madrid: Gredos, 1999. Cap. 76, p. 4843-4913.
- BYBEE, J. Introduction. In: BYBEE, J. (Ed.). *Frequency of Use and the Organization of Language*. New York: Oxford University Press, 2007. p. 7-22.
- BRUHN DE GARAVITO, J.; WHITE, L. L2 acquisition of Spanish DPs: the status of grammatical features. In: PÉREZ-LEROUX, A.T.; LICERAS, J. (Eds.). *The acquisition of Spanish morphosyntax: The L1/L2 connection*. Dordrecht: Kluwer, 2002. p. 153-178.
- CONSELHO DA EUROPA. *Quadro Europeu Comum de Referência para as Línguas. Aprendizagem, ensino, avaliação*. Porto: Edições ASA, 2001.
- CORBETT, G. G. *Gender*. Cambridge: Cambridge University Press, 1991.
- DAVIES, M. *Corpus do Português* [CdP]. Disponível em: <<https://www.corpusdoportugues.org/web-dial/>> /. Acesso em: 29 set. 2021.
- Corónica troiana em linguagem portuguesa: edicion y estudio*. Edição de Ana María García Martín. Salamanca. Luso-Española de Ediciones, 1998.
- FERREIRA, T. S. *Aquisição/Aprendizagem do sistema de atribuição de género nominal em PLNM*. 2019. 343 f. Tese (Doutoramento em Linguística do Português: Investigação e Ensino) – FLUC/UC, Coimbra, 2019.
- \_\_\_\_\_. A morfologia dos desvios de género gramatical em PLNM. *Diadorim*, Rio de Janeiro, v. 23, n.2, p. 210-226, 2021.
- HARRIS, J. W. The Exponence of Gender in Spanish. *Linguistic Inquiry*, v. 22, n.1, p. 27-62, 1991.
- MARTINS, C. O lugar do *input* linguístico e metalinguístico em teorias de aquisição/aprendizagem de línguas não maternas. Implicações pedagógicas. In: *Volume de Homenagem a Maria de Fátima Sousa e Silva*. Coimbra: Imprensa da Universidade de Coimbra, Instituto de Estudos Clássicos, 2022. p. 635-656.

- \_\_\_\_\_. 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, vol. 1, n.1, p. 26-51, 2015.
- MASTROPAVLOU, M.; TSIMPLI, I. The Role of Suffixes in Grammatical Gender Assignment in Modern Greek: A Psycholinguistic Study. **Journal of Greek Linguistics**, v. 11, n.1, p. 27-55, 2011.
- MATTOSO CÁMARA, J. **Estrutura da Língua Portuguesa**. Petrópolis: Vozes, 1970.
- \_\_\_\_\_. **História e Estrutura da Língua Portuguesa**. Rio de Janeiro: Padrão, 1975.
- MOITA, M. *et al.* Affix Selection and Deadjectival Nouns: a data-driven approach. In: HUMPHRIES, C. C.; KOSSEK, J.; GOMOLA, A. (Eds.). **English language, literature and culture: new directions in research**. Bielsko-Biala: Wydawnictwo Akademii Techniczno-Humanistycznej, 2010. p. 118-133.
- MOTA, M. A. Morfologia do nome e do adjetivo. In: RAPOSO, Eduardo *et al.* (Orgs.). **Gramática do português**. Vol. III. Lisboa: Fundação Calouste Gulbenkian, 2020. p. 2835-2930.
- NUNES, J. J. **Compêndio de Gramática Histórica Portuguesa**. Lisboa: Clássica Editora, 1989.
- POPLACK, S.; POUSADA, A.; SANKOFF, D. Competing influences of gender assignment: variable process, stable outcome. **Lingua**, v. 57, n.1, p. 1-28, 1982.
- RIO-TORTO, G. Flexão e derivação: simetrias e assimetrias. **Revista Portuguesa de Filologia (FLUC)**, vol. XXIV, p. 253-289, 2002.
- RIO-TORTO, G. (Org.) *et al.* **Gramática derivacional do português**. 2. ed. Coimbra: Imprensa da Universidade de Coimbra, 2016.
- RIO-TORTO, G.; RODRIGUES, A. S. (2016). Formação de nomes. In: \_\_\_\_\_. **Gramática derivacional do português**. 2. ed. Coimbra: Imprensa da Universidade de Coimbra, 2016. p. 135-240.
- RODRIGUES, A.S. Noções basilares sobre a morfologia e o léxico. In: RIO-TORTO, G. *et al.* **Gramática derivacional do português**. 2. ed. Coimbra: Imprensa da Universidade de Coimbra, 2016. P. 35-133.
- TAFT, M; MEUNIER, F. Lexical representation of gender: A quasiregular domain. **Journal of Psycholinguistic Research**, vol. 27, n. 1, p. 23-45, 1998.
- TUCKER, R *et al.* **The French Speaker's Skill with Grammatical Gender: An Example of Rule-governed Behavior**. The Hague: Mouton, 1977.
- WHITE, L. *et al.* Gender and number agreement in nonnative Spanish. **Applied Psycholinguistics**, vol. 25, n.1, p. 105-133, mar. 2004.