

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/348920000>

Emprego dos pronomes clíticos do português europeu em corpora de aprendizagem de português língua estrangeira

Conference Paper · December 2020

CITATIONS

0

READS

121

9 authors, including:



Jorge Baptista

Universidade do Algarve

186 PUBLICATIONS 656 CITATIONS

[SEE PROFILE](#)



Joaquim Guerra

Universidade do Algarve

23 PUBLICATIONS 29 CITATIONS

[SEE PROFILE](#)



Reis, Sónia

Universidade do Algarve

41 PUBLICATIONS 53 CITATIONS

[SEE PROFILE](#)



Fátima Noronha

Universidade do Algarve

4 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Linguistic Aspects of Developmental Education [View project](#)



Portuguese Language as Foreign Language Teaching: difficulties and challenges [View project](#)

EMPREGO DOS PRONOMES CLÍTICOS DO PORTUGUÊS EUROPEU EM CORPORA DE APRENDIZAGEM DE PORTUGUÊS LÍNGUA ESTRANGEIRA

**Jorge Baptista^{1,2}, Joaquim Guerra¹, Sónia Reis¹, Fátima Noronha¹, Rui
Talhadas^{1,2}, Graça Fernandes¹, José Marques¹, João Marques¹ & Alexandra
Mariano¹**

¹Faculdade de Ciências Humanas e Sociais da Universidade do Algarve

²Spoken Language Lab – INESC-ID

INTRODUÇÃO

Como se sabe, o emprego dos pronomes clíticos não é um aspeto do conhecimento da língua de natureza estritamente morfossintática, já que nele intervêm igualmente aspetos de ordem lexical: tanto a escolha do caso gramatical para um dado pronome como a possibilidade de construir uma passiva pronominal (e fazer corretamente a concordância verbal) implicam o conhecimento da construção verbal em jogo. Por outro lado, os aspetos morfotáticos, envolvidos na ênclise e mesóclise (terminação verbal e forma do pronome), e os aspetos sintáticos, relacionados com a posição do clítico relativamente ao verbo de que depende, são conhecimento linguístico (Paiva Raposo et al., 2013) que é objeto de instrução explícita em contexto de aprendizagem.

Este estudo tem como ponto de partida uma observação de Leiria (2006, p. 338), a propósito de “exercícios de gramática” em sala de aula, isto é, exercícios “que incidem, em boa parte, na morfossintaxe”:

No que respeita a ordem das palavras, a colocação dos clíticos costuma ser um dos aspetos sintáticos que ocupa mais tempo na sala de aula (até mesmo com aprendentes que ainda estão muito longe de um estágio de aquisição em que virão a usá-los espontaneamente...). Porque a modalidade escrita lhes permite aplicar conhecimento explícito repetidamente fornecido pelo professor e porque, nesta fase, o recurso a processos anafóricos é muito limitado, o número de desvios no que respeita à colocação de pronomes de acusativo e dativo é reduzido e, por outro lado, as omissões do pronome de flexão reflexiva são muitas. (Leiria, 2006, p. 338)

A autora fundamenta estas reflexões nos dados da tabela em que reporta “Desvios Sintáticos, Morfossintáticos e outros” (Leiria, 2006, p. 339). Entre outras observações, indicam-se desvios quanto à “ordem das palavras” e quanto à “colocação do pronome pessoal”, distribuídos pelas línguas maternas dos aprendentes. Ora, não é evidente no documento se o problema da colocação dos pronomes (89 casos no total) está incluído nos desvios de ordem das palavras (221 casos) - parece-nos que não. Embora a autora tenha constituído “um sistema de marcação de desvios relativamente impressionista mas muito económico” (p. 209) e apesar de algumas das situações relacionadas com o emprego dos pronomes estarem previstas neste sistema, não é possível relacionar os dados da tabela com os diferentes fenómenos abrangidos nessa tipologia (p. 210), em particular “as omissões do pronome de flexão reflexiva” (correspondentes ao caso marcado como ‘X6’: “este item não foi usado; o falante nativo tê-lo-ia usado” (p. 210). Por outro lado, ainda não dispomos de um referencial de emprego de pronomes clíticos que nos permita dizer que, num *corpus* particular, a sua frequência é superior ou inferior à que se espera de falantes nativos. Insatisfeitos com a forma como foram apresentados estes dados, quisemos caracterizar de forma mais aprofundada e assim compreender melhor o comportamento linguístico dos aprendentes de PLE.

EMPREGO DOS PRONOMES CLÍTICOS POR ESTUDANTES DE PLE/PLNM

Corpora de aprendizagem e constituição do corpus de trabalho

Para este estudo, foi constituído um *corpus* de trabalho (Talhadas, Mamede, & Baptista, no prelo) a partir da reunião de três *corpora* de aprendizagem, publicamente disponíveis, totalizando cerca de 368 mil palavras:

- COPLE2: *Corpus de Português Língua Estrangeira/Língua Segunda* (Mendes *et al.*, 2014; Mendes *et al.* 2016);
- RePLE (Recolha de dados de aprendizagem de português língua estrangeira);
- PEAPL2: *Corpus de Produções Escritas de Aprendentes de PL2*.

A Tabela 1 apresenta o número de textos e de palavras por *corpus* e por nível de proficiência do *Quadro Europeu Comum de Referência para as Línguas* (Conselho da Europa 2011), considerando apenas três níveis (A, B e C). A Tabela 2 apresenta os mesmos valores em percentagem dos totais. Globalmente, o COPLE2 é o *corpus* mais extenso, com 44% das palavras (48% dos textos), seguido pelo PEAPL2 (36% das palavras e 29% dos textos) e pelo RePLE (20% em número de palavras e 23% quanto ao número de textos). Os diferentes níveis de proficiência apresentam-se distribuídos de forma diferente, com o nível A representar 58% do total de palavras (47% dos textos), seguido do nível B (33% das palavras, 45% dos textos). O nível C está pouco representado, com apenas 9% das palavras e 8% dos textos.

Tabela 1. *Corpora* de aprendizagem utilizados

Níveis QECRL	COPLE2		RePLE		PEAPL2		Total	
	Textos	Palavras	Textos	Palavras	Textos	Palavras	Textos	Palavras
A	451	58.928	235	30.246	225	33.032	911	122.206
B	487	95.386	162	29.843	292	87.993	941	213.222
C	26	7.460	72	12.067	59	13.006	157	32.533
Total	964	161.774	469	72.156	576	134.031	2009	367.961

Tabela 2. *Corpora* de aprendizagem utilizados (em percentagem)

QECRL	Textos	Palavras	Corpus	Textos	Palavras
A	45%	33%	COPLE2	48%	44%
B	47%	58%	RePLE	23%	20%
C	8%	9%	PEAPL2	29%	36%

Constituição do *corpus* de trabalho: extração de frases

Os textos dos três *corpora* foram processados, separando as frases com base na pontuação, uma frase por linha. Em seguida, foram extraídas as frases contendo as formas correspondentes (a) aos pronomes pessoais clíticos *me, te, se, nos, vos, lhe* e *lhes*; (b) às formas contraídas *ma, mas, mo, mos; ta, tas, to, tos; lha, lhas, lho, lhos*; e (c) às combinatórias *no-la, no-las, no-lo, no-los; vo-la, vo-las, vo-lo, vo-los*. Ao todo, foram extraídas 7.226 frases. A distribuição das frases extraídas do *corpus* pelos tipos de pronomes (formas) encontra-se na Tabela 3.

Tabela 3. Distribuição das frases extraídas do *corpus* por pronome (forma) e por níveis de proficiência

Pronome	A	B	C	Total	%
ME	646	720	15	1381	0,19
TE	102	213	18	333	0,05
LHE	56	100	17	173	0,02
SE	456	1022	202	1680	0,23
NOS	320	643	110	1073	0,15
VOS	4	6	2	12	0,00
MO	905	1453	208	2566	0,36
LHO	5	3	0	8	0,001
Total	2494	4160	572	7226	
%	0,35	0,58	0,08		

Como se pode verificar, a distribuição das frases extraídas pelos níveis de proficiência é bastante proporcional à representatividade destes níveis no conjunto dos *corpora*, sobretudo considerando-os em termos de número de palavras. Quanto à distribuição das formas, importa considerar que algumas destas são ambíguas. Assim, por exemplo, não se distinguem, na Tabela 3 os diferentes valores de *se* (pronome/conjunção) ou de *nos* (pronome/contração de *em+os*).

Uma vez extraídas as frases, indicou-se para cada uma o nome e a linha do ficheiro de origem e o nível de proficiência em que o estudante que o produziu foi classificado pelos compiladores dos *corpora* de origem. As frases extraídas constituem o *corpus* de trabalho, que será o objeto deste estudo. As frases foram ordenadas aleatoriamente e associadas a um código alfanumérico convencional que identifica de forma unívoca o pronome-alvo. A linha abaixo apresenta um exemplo do *corpus* (campos separados por '|'):

SE-0022|A|CATALAO.ER.A2.105.69.3Q.utf8.txt:1|*Eu sou uma pessoa que pode definir-se como urbanita.*

As frases foram distribuídas aleatoriamente por uma equipa de 7 linguistas e professores de PLE, cabendo a cada um 985 ou 986 frases. A estas soma-se ainda um conjunto de 328 frases, que foram anotadas por todos, a fim de medir o acordo entre anotadores.

Instruções de anotação

Foi elaborado um documento com instruções de anotação, com exemplos ilustrativos das diferentes situações, critérios linguísticos a aplicar e formato da anotação a fazer. A anotação consistiu, basicamente, em indicar, para o pronome-alvo em cada frase do *corpus*:

- o respetivo caso (nominativo, oblíquo, acusativo, dativo e reflexo);
- para o pronome '-se', se se tratava de um pronome indefinido (e.g. *Dizia-se muitas coisas sobre ele*); ou
- no caso dos pronomes reflexos, se se tratava de uma construção passiva pronominal (e.g. *Diziam-se muitas coisas sobre ele*).

Todas as anotações deste conjunto são mutuamente exclusivas. Além disto, para cada pronome observado, indica-se também se este apresenta algum dos seguintes tipos de erro, nomeadamente,

- d) a escolha do caso gramatical (e.g. **Quando cheguei cá, embora eu não conhecesse, outras pessoas aproximavam-me e fizeram beijinhos para mim*);
- e) a forma do pronome resultante dos processos morfológicos dependentes da terminação verbal e da forma do pronome (e.g. *Por isso, sentimos-nos menos "real" sobre [isso]*);
- f) os erros na posição do pronome clítico relativamente ao verbo de que sintaticamente depende (próclise/mesóclise/ênclise, e.g. ; *deveria-se encontrar uma solução; [...] é por isso que as companhias aproveitam-se da situação [...]*);
- g) considerou-se ainda uma categoria 'outro' para todas as outras situações em que os erros não correspondem às três situações anteriores (e.g. **Ele comprou-se um carro novo*; emprego da construção verbal com pronome espúrio).

As anotações quanto ao erro podem combinar-se entre si de diferentes formas, consoante o caso. Por exemplo, em frases como: *Tenho dois amigos irlandeses e no verão gostaria de visitar-lhes em Irlanda porque nunca tinha ido lá*, verifica-se simultaneamente um erro na escolha do caso do pronome (dativo por acusativo) e na posição (ênclise em vez de próclise): *gostaria de os visitar*.

Na atribuição do caso, considerou-se o contexto sintático e semântico e a intenção comunicativa do estudante, ainda que pudesse haver erro na escolha do caso correto naquela instância. Por exemplo, na frase *Tenho dois amigos irlandeses e no verão gostaria de visitar-lhes em Irlanda porque nunca tinha ido lá*, indicou-se que o pronome-alvo deveria estar no caso acusativo (*visitá-los*) e marcou-se o erro na escolha do caso.

Para a distinção entre os casos acusativo e dativo na primeira e segunda pessoa (v.g. *me, te, nos, vos*), foram dadas instruções no sentido de se parafrasear a construção usando uma forma da terceira pessoa (*o/a, lhe*). As formas não pronominais (por exemplo, a contração *nos* ou as conjunções *se* e *mas*) são igualmente marcadas para serem descartadas mais tarde. Encontraram-se muitos casos dos pronomes *nós* e *vós* sem acento gráfico, o que nos levou a assinalá-los com o caso nominativo ou oblíquo, consoante a situação.

No caso de pronome *-se*, indicou-se se se tratava do pronome indefinido (e.g. *Diz-se muitos disparates em futebol*) ou de uma construção passiva pronominal (e.g. *Dizem-se muitos disparates em futebol*), considerando-se a concordância verbal como critério distintivo: no primeiro caso, o verbo encontra-se obrigatoriamente na terceira pessoa do singular a concordar com o pronome indefinido *se*; ao passo que, no segundo caso, o verbo concorda com o objeto da construção. Nos casos em que a concordância não permite fazer essa distinção (*Diz-se muito disparate em futebol*), optou-se por dar preferência à construção com pronome indefinido.

Processo de anotação

A anotação individual das frases do *corpus* foi preparada com várias sessões coletivas de trabalho, em que participaram todos os anotadores, a fim de aferir critérios, detetar incorreções ou insuficiências nas diretivas de anotação, e treinar o processo propriamente dito. Das frases a anotar em conjunto, foram retiradas cerca de 100, que serviram de base de trabalho para essas sessões.

RESULTADOS

Ao todo, o *corpus* anotado acabou por ficar constituído por 4.999 frases, das quais 328 (6,6%) foram anotadas por todos os participantes no estudo, a fim de determinar a taxa

de acordo entre anotadores (ing. *inter-annotator agreement*). Dos 7 anotadores iniciais, um não respondeu a tempo de podermos integrar os seus dados neste estudo; e outro alterou inadvertidamente o formato do ficheiro de anotação pelo que não foi possível igualmente aproveitar os seus dados.

Acordo entre anotadores

Consideraram-se dois conjuntos de dados, para os quais se convencionou o conjunto de etiquetas seguinte:

- a escolha de caso (N-nominativo, O-oblíquo, A-acusativo, D-dativo e R-reflexo), o pronome indefinido *-se* (I) e a passiva pronominal (P);
- o tipo de erro observado (C-caso, F-forma, P-posição e O-outro).

Para determinar a taxa de acordo entre anotadores, utilizou-se a ferramenta *online* de Geertzen (2012) e as métricas Fleiss-kappa e Krippendorff-alpha (variáveis com valores nominais e múltiplos anotadores). Os resultados deste estudo de acordo entre anotadores quanto à marcação de caso encontram-se na Tabela 4.

Tabela 4. Concordância entre anotadores quanto ao caso

	Fleiss-kappa			Krippendorff-alpha		
	F-obs	F-exp	F-kappa	K-obs	K-exp	K-alpha
Caso-Total	0,811	0,320	0,825	0,122	0,694	0,825
Caso-A1	0,867	0,318	0,805	0,132	0,695	0,810
Caso-A2	0,870	0,311	0,811	0,129	0,695	0,814
Case-A3	0,903	0,310	0,859	0,096	0,693	0,862
Case-A4	0,896	0,327	0,846	0,111	0,693	0,840
Case-A5	0,867	0,319	0,804	0,138	0,693	0,801

Como se pode verificar, a taxa de concordância quanto ao caso é bastante elevada (*Fleiss-kappa* = 0,825; *Krippendorff-alpha* = 0,825), não se verificando grande disparidade de resultados quando se retira sucessivamente os dados de cada um dos anotadores (desvio-padrão de 0,016 e 0,025, respetivamente). Estes resultados permitem, por isso, considerar que, globalmente, a anotação individual das restantes frases (4.671) quanto ao caso deverá ser bastante consistente.

Apenas se registaram 86 frases com anotações de erro (36% das 328 frases). Uma vez que, para cada instância de um mesmo pronome, era possível indicar a existência de diferentes tipos de erro (C-caso, F-forma, P-posição e O-outro, por esta ordem), a etiqueta de cada anotador para cada instância, poderia ser constituída *a priori* por qualquer combinação destes valores (C, CF, CP, CO, CFP, CFO, CPO, etc.), a que se junta a possibilidade de não ter sido feita qualquer anotação. Ao todo, foram assinalados 22 erros de caso (C), 21 de forma (F), 39 de posição (P) e 47 de outro tipo (O). A Tabela 5 apresenta os casos em que as frases apresentam apenas a anotação de um tipo de erro. Estes casos (57) representam 64% de todas as frases com anotações de erro.

Tabela 5. Frases anotadas apenas com um tipo de erro

n.º anotadores	C	F	P	O	total
----------------	---	---	---	---	-------

5	0	0	3	5	8
4	0	0	4	1	5
3	0	0	4	4	8
2	0	0	8	4	12
1	0	1	9	14	24
<hr/>					
Total	0	1	28	28	57

Como se pode verificar, apenas para 8 frases (14%) houve acordo total quanto ao tipo de erro, 3 quanto à posição (P) e 5 quanto a “outro” tipo de erro (O). Em muitos casos (24, 42%), o erro apenas foi marcado por uma das 5 pessoas envolvidas. Esta disparidade nas anotações recomenda, pois, a maior cautela na interpretação dos resultados relativos à anotação do tipo de erro. Trata-se, efetivamente, de uma tarefa muito complexa, que poderá não ter sido inteiramente bem compreendida pelos anotadores ou para a qual as diretivas não eram suficientemente claras.

A distribuição dos casos pronominais nas restantes frases (anotadas apenas uma vez e individualmente por cada anotador) é a que se apresenta na Tabela 6.

Tabela 6. Distribuição dos casos pronominais nas frases anotadas individualmente por cada um dos anotadores

Caso	#	% (total)	% (anotados)
-	2214	0,32	-
N	135	0,02	0,03
A	266	0,04	0,06
D	620	0,09	0,13
O	22	0,003	0,005
R	1100	0,16	0,23
I	207	0,03	0,04
P	39	0,01	0,01
X	2295	0,33	0,49
<hr/>			
Total	6898	6898	4684

Na primeira linha da tabela encontram-se os casos ‘-’, que não foram objeto de anotação e que representam 30% das frases inicialmente recolhidas do *corpus* (7.226). Os casos ‘X’ são frases em que a forma-alvo não representava um pronome (33%). Verifica-se que o caso pronominal mais frequente é o pronome reflexo, representado por ‘R’: 16%), seguido do caso dativo (‘D’: 9%) e do acusativo (‘A’: 4%). Os mesmos valores foram calculados (coluna da direita) considerando apenas as frases anotadas (4.684). Neste sentido, os pronomes reflexos constituem 23% dos pronomes, os dativos 13% e os acusativos 6%.

Considerando, então, todo o *corpus*, tanto as formas anotadas em conjunto como as anotadas por cada participante individualmente, verifica-se a seguinte distribuição dos casos pronominais (Tabela 7).

Tabela 7. Distribuição dos casos pronominais nos diferentes recortes do *corpus*

Caso	%R	% I	% R+I ^a
N	0,030	0,029	0,029
A	0,076	0,057	0,058
D	0,104	0,132	0,131
O	-	0,005	0,004
R	0,235	0,235	0,235
I	0,040	0,044	0,044
P	0,012	0,008	0,009
X	0,463	0,490	0,489
Total	328	4684	4999

^a R – recorte do corpus anotado em conjunto; I – recorte do corpus anotado individualmente

Como se pode ver, os resultados não sofrem grandes alterações pela reunião (R+I) dos dados das frases anotadas em conjunto (R) com as frases anotadas individualmente (I). O desvio-padrão na proporção dos diferentes casos é em média 0,5% e a maior variação é de 1,5%. Pela distribuição homóloga dos casos pronominais através dos diferentes recortes do *corpus* podemos confirmar, de forma independente, a consistência da anotação para todo o *corpus*.

Distribuição dos pronomes pelos níveis de proficiência (após a anotação)

Considerando agora a distribuição dos diferentes pronomes pelos níveis de proficiência, após o processo de anotação e eliminando-se os exemplos espúrios (ambíguos), obtiveram-se os seguintes resultados (Tabela 8), para cada tipo de pronome.

Tabela 8. Distribuição dos pronomes pelos níveis de proficiência (em valores absolutos e em percentagem do total de cada tipo de pronome)

PRO	A	B	C	A%	B%	C%	Total
ME	449	470	9	0,484	0,506	0,010	928
TE	71	155	13	0,297	0,649	0,054	239
LHE	39	73	11	0,317	0,593	0,089	123
LHO	4	1	0	0,800	0,200	0,000	5
MO	2	7	0	0,222	0,778	0,000	9
NOS	121	238	44	0,300	0,591	0,109	403
VOS	2	2	0	0,500	0,500	0,000	4
SE	243	486	113	0,289	0,577	0,134	842
Total	931	1.432	190	0,401	0,549	0,050	2.553

Os valores de percentagem desta tabela interpretam-se do seguinte modo: há 928 casos de emprego de *me*, dos quais 449 foram produzidos por estudantes de nível A, 470 por estudante de nível B, e apenas 9 por estudantes de nível C. Como se pode verificar, o emprego das contrações MO e LHO é bastante raro (não foram encontradas formas TO), como é, de resto, pouco frequente em textos de falantes nativos. O mesmo acontece com o emprego da

forma VOS. Apesar de o número de frases por nível ser diferente, o número relativo de ocorrências dos diferentes pronomes mantém globalmente essa proporção, exceto para ME (mais frequente em A e B, raro em C) e para SE (mais frequente em C).

DISCUSSÃO

Referência de distribuição de pronomes

A fim de determinar uma referência que permitisse comparar o emprego efetivo dos pronomes por parte de estudantes de PLE com o que se esperaria de um falante nativo (com um grau escolarização médio), considerámos que um corpus como o CETEMPúblico, pelas suas dimensões e carácter generalista (Hawkins & Filipovic, 2012), permitiria estabelecer alguma base de comparação com os dados aqui produzidos. Verifica-se, de facto, alguma assimetria quanto à percentagem relativa de cada pronome em cada corpus (Tabela 9).

Tabela 9. Distribuição dos pronomes (comparação entre corpora da frequência de cada pronome)

Pronome	CetemPúblico	Corpus	%CP	%Corpus	dif%(C,CP)
me	51.007	928	4,74	36,55	32
Te	2.363	239	0,22	9,41	9
lhe	114.775	123	10,66	4,84	-6
nos	47.631	403	4,42	15,87	11
vos	1.425	4	0,13	0,16	0
Se	859.331	842	79,82	33,16	-47
Total	107.6532	2.539	-	-	-

Em concreto, observa-se: (i) que há um emprego proporcionalmente muito alto (quase 80%) do pronome reflexo *-se* no CETEMPúblico, enquanto no corpus de trabalho esse valor é alto (33%) mas mais equilibrado com outros pronomes; (ii) que o emprego da 1ª pessoa (*me/nos*) é reduzido (como se esperaria) no *corpus* de texto jornalístico – e ainda mais o da 2ª pessoa (*te/vos*). Esta distribuição assimétrica justifica, pois, o grau de correlação (índice de Pearson) relativamente baixo, de apenas Pearson=0,544, que se encontra para os valores da Tabela 9. Não sendo estes dois corpora comparáveis, é, no entanto, assinalável que se verifiquem tais assimetrias, o que pode levar a pensar que o comportamento dos estudantes de PLE, relativamente ao emprego dos pronomes clíticos no conjunto do corpus, apresenta aspetos de sub- e sobre-representação, talvez nem todos fruto das diferenças nos géneros textuais e cada corpus.

Análise de erros

Procedeu-se a uma análise manual dos erros assinalados pelos anotadores. Relativamente aos erros de posição do clítico, das 38 frases assinaladas, 13 não correspondiam de facto a erros. Por exemplo, na frase *Sempre te dizia que tens talento*, em que a atração do clítico pelo advérbio foi incorretamente assinalada como um erro de posição. Em alguns casos o erro de posição assinalado pode ser considerado como resultado de uma variação existente na língua e a preferência do anotador por uma das construções sintáticas, e.g. *Eu vou me aproveitar de uma das desvantagens do uso do telemóvel* (compare-se: *Eu vou aproveitar-me disso*). Dos 20 erros de posição confirmados, há, ainda assim, que considerar 4 casos, em que poderá estar em jogo o uso da variante do português

do Brasil, já que não há informação disponível quanto ao contexto de aprendizagem; *Me entendo muito bem com ela*.

Entre os erros de posição efetivamente encontrados, observa-se situações de não conformidade às regras de próclise, nomeadamente (i) com sujeitos indefinidos, e.g. *Alguém disse-me que parecia uma mulher doméstica*; (ii) em orações subordinadas, e.g. *Porque agora sinto-me mal, É a história dum jovem indiano, que [...] torna-se milionário*; e (iii) dois erros de incorreta construção da mesóclise: *Por tanto, [...] deveria-se encontrar uma solução*; *Se [...], o mundo transformaria-se mais bonito*. No exemplo seguinte, ao erro de caso acrescenta-se um erro de posição sintática: *Tenho dois amigos irlandeses e no verão gostaria de visitar-lhes em Irlanda* (comparar: *gostaria de os visitar*).

Relativamente aos restantes erros de caso, regista-se (i) o problema já referido da falta de acentuação gráfica do pronome nós (6 casos), e.g. *Nos temos de provar*; (ii) bem como o emprego da forma dativa/acusativa pelas formas oblíquas: *outras pessoas aproximavam-me*; *Usei a palavra crescer num sentido diferente para me* (compare-se: *outras pessoas aproximavam-se de mim*; *num sentido diferente para mim*); ou (iii) o uso do caso dativo pelo acusativo: *nós levámos-lhe ao hospital rapidamente* (compare-se: *nós levámo-lo ao hospital*); (iv) ou, ainda, o emprego de um pronome reflexo mas com a escolha da pessoa-número errada: *Eu levanto-se às oito da manhã* (compare-se: *Eu levanto-me às oito da manhã*).

Finalmente, registámos dois casos marcados como “outros erros” por se tratar de empregos espúrios de pronomes que não são pedidos pela construção verbal; nestes casos, poderá haver uma influência da língua materna (espanhol ou francês) ou de interferência de uma língua segunda na aprendizagem do português: *Não falo frequentemente porque não me gosta muito*; [*no centro comercial não veio nada que me gosta tão de comprar-lo*]. (comparar ES: *no me gusta* ou FR: *ne me plait*).

O exemplo seguinte também é um caso de um emprego espúrio mas não é evidente qual é o fator que terá determinado este erro: *O meu desejo é de comprar-me uma casa perto da aldeia*.

CONCLUSÃO

Neste estudo, procurou-se fazer uma caracterização da distribuição de pronomes clíticos em corpora de aprendizagem, depois de os pronomes terem sido anotados manualmente quanto ao caso e tipo de erro encontrado. Foram também anotadas as construções impessoais e as passivas pronominais. O processo de anotação de 7.200 frases por vários linguistas/professores de PLE/PLNM foi avaliado quanto ao grau de concordância entre anotadores, o que revelou, relativamente à marcação de caso, elevada concordância, mas, quanto à identificação do tipo de erro, apenas uma concordância moderada. De notar, nos resultados, que a frequência de emprego dos pronomes clíticos é proporcionalmente muito mais baixa do que num corpus de textos jornalísticos. A escassez de dados, contudo, não permite tirar conclusões seguras sobre uma esperada progressão no domínio destas estruturas gramaticais, ao longo dos diversos níveis de proficiência. De futuro, pretendemos cruzar estas informações com as orientações oficiais para o ensino de PLE/PLNM (Grosso et al., 2011a, 2011b; Instituto Camões, 2017), bem como caracterizar a forma como este tópico é tratado em manuais escolares.

AGRADECIMENTOS

Parte desta investigação foi financiada por fundos públicos através da Fundação para a Ciência e a Tecnologia (ref. UID/CEC/50021/2013).

REFERÊNCIAS

- Conselho da Europa (2011). *Quadro Europeu Comum de Referência para as Línguas: Aprendizagem, ensino, avaliação*. Texto: Conselho da Europa.
- Geertzen, J. (2012). Inter-Rater Agreement with multiple raters and variables. Recuperado de <https://nlp-ml.io/jg/software/ira/>.
- Grosso, M. (Coord.), Soares, A., Sousa, F., & Pascoal, J. (2011a). *QuaREPE: Quadro de Referência para o Ensino Português no Estrangeiro - Documento Orientador*. Lisboa: Ministério da Educação - Direção Geral de Inovação e de Desenvolvimento Curricular.
- Grosso, M.J. (Coord.), Soares, A., Sousa, F., & Pascoal, J. (2011b). *QuaREPE: Quadro de Referência para o Ensino Português no Estrangeiro - Tarefas, Actividades, Exercícios e Recursos para a Avaliação*. Lisboa: Ministério da Educação - Direção Geral de Inovação e de Desenvolvimento Curricular.
- Hawkins, J., & Filipovic, L. (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. Cambridge: University Press.
- Instituto Camões (2017). *Referencial Camões Português Língua Estrangeira*. Lisboa: Camões, Instituto da Cooperação e da Língua I.P. - Direção de Serviços de Língua e Cultura.
- Leiria, I. (2006). *Léxico, aquisição e ensino do Português Europeu língua não materna*. Lisboa: FCG/FCT.
- Mendes, A., Antunes, S., Alexandre, N., Avelar, A., Castelo, A., Duarte, I., Freitas, M., Gonçalves, A., Pascoal, J., Pinto, J., & Janssen, M. (2014). Corpus de português língua estrangeira/língua segunda – COPLE2. *XXX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL.
- Mendes, A., Antunes, S., Janssen M., & Gonçalves, A. (2016). The COPLE2 Corpus: A Learner Corpus for Portuguese. *Proceedings of the Tenth Language Resources and Evaluation Conference – LREC'16*, Portoroz, Slovenia, 3207-3214.
- Paiva Raposo, E., Nascimento, M., Mota, M., Segura, M., & Mendes, A. (2013). *Gramática do Português*. Lisboa: Fundação Calouste Gulbenkian.
- Talhadas, R., Mamede, N., Baptista, J. (no prelo). Mapeamento de estruturas gramaticais do Português nos níveis de proficiência do QECRL. *VI Simpósio Mundial De Estudos da Língua Portuguesa*, Santarém.