

O LÉXICO DOS ROMANCES CAROLÍNGEOS DA TRADIÇÃO ORAL MODERNA PORTUGUESA EDITADOS ENTRE 1828 E 1960: UMA AMOSTRA

NATÁLIA ALBINO PIRES

Escola Superior de Educação de Coimbra

1 – Considerações Preliminares

No seguimento do trabalho de investigação que estou a desenvolver no âmbito do estudo do léxico do romanceiro português da tradição oral moderna, mais especificamente das versões editadas entre 1828 e 1960, tem vindo a ser notório que, por um lado, existe um léxico comum a todo o género romancístico e que, por outro, determinados grupos de romances, determinadas áreas geográficas e determinados editores apresentam lemas/lexemas específicos.

O estudo, que a seguir se apresenta, tem como base a análise estatística comparativa de um *corpus* constituído por 174 textos, o número total de versões dos Romances Carolíngios editadas entre 1828 e 1960 e catalogadas por Ferré (2000)¹. No entanto, apresentar, em poucas páginas, o léxico total dos Romances Carolíngios revela-se uma tarefa impossível de concretizar dada a vastidão dos dados manejados: 4.261 palavras únicas e 28.950 ocorrências.

Assim, cingimo-nos apenas a alguns dos dados mais relevantes do *corpus* dos romances carolíngios e que, indubitavelmente, o caracterizam face a outros *sub-corpora* do romanceiro.

Escolhendo como amostra dos dados do *sub-corpora* romancístico aqui em análise, os Romances Carolíngios, três romances²: “Conde Preso” (BRTOM XIII/IGR 0118); “Morte de D. Beltrão” (BRTOM XV/IGR 0150) e “D. Gaifeiros” (BRTOM XVI/IGR 0151), num total de 44 textos, analisaremos, para os três romances, as principais classes gramaticais Nome, Adjectivo e Verbo e a sua distribuição: por romance, por área geográfica e por editor.

Conscientes de que todos os textos do *corpus* aqui em análise foram, em maior ou menor grau, retocados pelos seus editores, para a distribuição de lexemas por editor, optamos pela comparação dos dados entre dois grupos de editores: Garrett — Outros Editores (Braga, Leite, Lopo, Luís Chaves, Martins, Sepúlveda, Tavares e Thomás). É de salientar, contudo, que a nossa opção pretende apenas exemplificar uma linha de investigação possível para a análise dos dados linguísticos dos textos de um editor que retoca bastante as suas versões frente a outros mais respeitadores da tradição.

¹ Ferré, Pere, *Romanceiro Português da Tradição Oral Moderna*, Vol. 1, Lisboa, Fundação Calouste Gulbenkian, 2000, pp. 192-409.

² Os romances serão referenciados com o número romano atribuído em Ferré, Pere e Carinhas, Cristina, *Bibliografia do Romanceiro Português da Tradição Oral Moderna*, Madrid, Seminario Menéndez Pidal/Universidad Complutense de Madrid, 2000 e com o número atribuído em Catalán, Diego, *El Romancero Panhispánico — Catálogo General Descriptivo*, vol. II e III, Madrid, Seminario Menéndez Pidal/Universidad Complutense de Madrid, 1982 e 1983, respectivamente. A partir de agora a primeira referência será indicada com a sigla BRTOM e a segunda com IGR, ambas seguidas do respectivo número identificativo do romance.

Ainda que sem qualquer pretensão de me alargar, uma vez que o objectivo primordial deste estudo é a apresentação de dados e não das metodologias de trabalho que permitiram a obtenção dos mesmos, há algumas notas metodológicas e terminológicas que importa ressaltar antes de proceder à sua apresentação.

Em termos metodológicos, antes do processamento informático dos textos que compõem o *corpus*, procedeu-se à filtragem de todas as indicações numéricas de verso, à substituição de todas as abreviaturas pelas palavras plenas, à eliminação de todos os parêntesis curvos e de todos os parêntesis rectos que indicassem texto introduzido pelo reeditor, sendo, nesses casos, as palavras contabilizadas como frequências/ocorrências. Mantiveram-se, todavia, todos os parêntesis rectos que indicam a irregularidade métrica de um verso por forma a não invalidar o resultado dos totais de ocorrências e frequências em fim de verso.

Antes do processamento informático dos dados, atribuíram-se campos-código a todas as informações identificativas dos romances, nomeadamente aos códigos do IGR³ pelos quais os romances são identificados; aos nomes dos editores e à data da primeira edição da versão; aos números atribuídos a cada romance em BRTOM⁴; à localização geográfica de cada versão, especificada por Concelho e Distrito ou, em caso disso, com a indicação de geografia atribuída ou de falta de informação.

Para os dados aqui apresentados não foram tidas em conta as variantes dos textos, embora estejam contabilizadas no estudo mais alargado que continuo a desenvolver.

No que se refere a questões de terminologia linguística, usaremos maioritariamente termos da Gramática Tradicional, recorrendo, sempre que necessário e com vista ao rigor, a termos de outras correntes de análise gramatical. Não é, no entanto, aqui o momento para a defesa de toda a terminologia adoptada.

Uma última nota antes da apresentação dos dados vai para o *tagger* utilizado para a etiquetagem dos textos. Após análise e testagem dos pouquíssimos etiquetadores morfológicos de português (in)disponíveis que, para o caso do romancista, revelaram margens de acerto baixíssimas, optou-se pelo *Ulisses*, especificamente desenvolvido por uma empresa portuguesa para a etiquetagem do *corpus* romancístico.

2 – Dados Estatísticos

2.1 – Da Introdução aos quadros de dados

Os dados estatísticos, sintetizados nos quadros que seguidamente apresentamos dividem-se em dois grupos distintos: dados gerais (Quadro I e II) do *corpus* constituído pelos romances carolíngios e dados específicos (Quadro III a XIII) do *corpus* de amostra constituído por três romances (“Conde Preso”, “Morte de D. Beltrão” e “D. Gaifeiros”).

2.2 – Da apresentação e análise dos dados

2.2.1 – Dados Gerais (*Corpus* dos Romances Carolíngios)

Os quadros I e II pretendem ser apenas exemplificativos dos dados do *corpus* constituído pelos Romances Carolíngios editados entre 1828 e 1960. Dão conta do total de

³ Cf. Nota 2, Catalán, Diego.

⁴ Cf. Nota 2, Ferré, Pere e Carinhas, Cristina.

ocorrências de cada palavra no *corpus*, cujo total de palavras é de 28.950, independentemente da sua classe gramatical.

Entre os dados apresentados no Quadro I e os apresentados no Quadro II há um número significativo de palavras cujo índice de ocorrência varia entre ≤ 199 e ≥ 2 .

O Quadro II dá, apenas, conta de uma ínfima parte das palavras com índice de ocorrência = 1, uma vez que o seu total é de 1.864, ou seja, 6,44% do léxico dos Romances Carolíngeos.

QUADRO I

Palavras com maior índice de ocorrência no *corpus*
(Romances Carolíngeos editados entre 1828 e 1960).

≥ 199

palavra	total	%	palavra	total	%
a	2511	8,67	dar	304	1,05
que	2162	7,47	há-de	299	1,03
de	1465	5,06	vai	296	1,02
não	1199	4,14	os	294	1,02
o	1176	4,06	Deus	293	1,01
se	987	3,41	nem	291	1,01
me	707	2,44	dom	286	0,99
para	557	1,92	seu	286	0,99
e	549	1,90	uma	281	0,97
por	513	1,77	conde	273	0,94
eu	510	1,76	menina	269	0,93
meu	457	1,58	mais	253	0,87
com	443	1,53	quem	248	0,86
do	442	1,53	lhe	244	0,84
um	435	1,50	três	227	0,78
te	406	1,40	à	226	0,78
ó	384	1,33	ele	222	0,77
ao	346	1,20	queimar	216	0,75
minha	332	1,15	as	211	0,73
no	325	1,12	foi	206	0,71
é	325	1,12	já	204	0,70
lá	320	1,11	aqui	199	0,69
da	313	1,08	TOTAL	14113	48,75

233

QUADRO II

Exemplo de palavras com menor índice de ocorrência no *corpus*
(Romances Carolíngeos editados entre 1828 e 1960).

= 1

palavra	total	palavra	total	palavra	total	palavra	total
momento	1	plantada	1	assomou-se	1	ceiar	1
rouquinha	1	há-de-se	1	igualada	1	mandai-a	1
acompanhai-me	1	quedar	1	vejam	1	reparava	1
vigário	1	dá-se-lhe	1	lonjura	1	pudessem	1
praticar	1	dormiste	1	começa-se	1	capela	1
acção	1	confiança	1	ponham-se	1	olhinhos	1
preparado	1	perdeu	1	tremia	1	atopou-o	1
alvores	1	geada	1	soluços	1	deixasse-lo	1
considerar	1	fossem	1	corriam	1	achasse	1
dormiríeis	1	marão	1	viro-me	1	abancar	1
deixasse	1	entregasse	1	aula	1	de-la-mar	1
tola	1	refego	1	apenas	1	aventou	1
deitava	1	alevantada	1	companhia	1	TOTAL	52

Em relação aos totais das palavras <a>, <que>, <o> e <se>, há que ter em conta que, no Quadro I, dizem respeito ao cômputo de todas as categorias gramaticais a que podem pertencer na língua portuguesa e não ao total de ocorrências por categoria. Depois de etiquetadas estas palavras, os totais sofrem algumas alterações, mantendo-se, no entanto, entre as palavras com maior índice de ocorrência.

No Quadro I, é de salientar que nenhuma das 20 palavras com maior índice de ocorrência pertence às classes gramaticais aqui em análise (Nome, Adjectivo e Verbo). Quanto à forma verbal <queimar> e aos nomes <conde> e <menina>, tendo em conta que o número de versões do romance “Conde Claros em Hábito de Frade” (BRTOM XVII/IGR 0159) é significativamente maior do que o número de versões dos outros romances carolíngios, havia toda a probabilidade, tal como se verifica nos dados do Quadro I, de se encontrarem alguns dos lexemas pertencentes ao léxico particular desse romance entre os cinquenta lexemas com maior índice de ocorrência no *corpus*.

Da análise dos dados do Quadro I, verificamos, então, que os lemas verbais com maior índice de ocorrência no *corpus* são <ser>, <dar>, <haver de> e <ir>, verificamos que os lemas nominais com maior índice de ocorrência são <Deus> e <Dom> e que não existe qualquer lema ou lexema adjectival entre as formas com total de ocorrência \geq a 199.

No que se refere ao Quadro II, que lembro ser apenas exemplificativo dos dados do *corpus*, constata-se que entre as palavras com menor índice de ocorrência se encontram determinadas formas populares como <alevantada>, <lonjura>, <ceiar>, <atopou>, <abancar> e <aventou>, algumas delas caídas em desuso ou consideradas regionalismos, como é o caso de <atopar> ou <aventar>.

2.2.2 – Dados Específicos (*Corpus* de amostra)

QUADRO III

Distribuição de totais das categorias gramaticais Nome, Adjectivo e Verbo por Romance no *corpus* de amostra.
Lemas por BRTOM

Romance BRTOM	Total de Palavras	Nome	%	Adjectivo	%	Verbo	%	% Total das 3 Classes
XIII	2.491	370	14,85	58	2,33	390	15,66	32,84
XV	4.091	776	18,97	129	3,15	653	15,96	38,08
XVI	3.249	395	12,16	59	1,82	357	10,99	24,96
TOTAL	9.831	1.541	15,67	246	2,50	1.400	14,24	32,42

Da análise do Quadro III (Distribuição de Lemas por BRTOM) verifica-se que o número de lemas nominais, adjectivais e verbais presentes quer no romance “Conde Preso” (BRTOM XIII) quer no romance “D. Gaifeiros” (BRTOM XVI) é bastante próximo, destacando-se o romance “Morte de D. Beltrão” (BRTOM XV) que apresenta o dobro de lemas adjectivais e praticamente o dobro de lemas nominais e verbais.

Quando comparado o Quadro III com o Quadro VII (Distribuição de Lexemas por BRTOM), verificamos que Nomes, Adjectivos e Verbos constituem apenas 51% do vocabulário dos romances, pelo que, obviamente, os restantes 49% são constituídos por todas as restantes classes de palavras da língua portuguesa. Relevante, ainda, dos dados deste quadro (Quadro VII) é o facto de se verificar que os romances “Conde Preso” (BRTOM

XIII) e “D. Gaifeiros” (BRTOM XVI) apresentam mais formas verbais do que nominais ao passo que o romance “Morte de D. Beltrão” (BRTOM XV) apresenta mais formas nominais do que verbais.

Da análise comparativa dos dois quadros (Quadro III e VII), constatamos que as formas mais flexionadas são as nominais e as verbais, ao passo que a flexão das formas adjectivais é pouco significativa. Salientam-se, também, no romance “D. Gaifeiros”, os dados referentes à classe verbo uma vez que o número de lexemas é mais do dobro do número de lemas. Estes dados relativos à flexão permitem-nos afirmar que, pelo menos para este *corpus* de amostra, o género literário romance dá preferência às nominalizações e às verbalizações em detrimento das adjectivalizações, ou seja, temos uma clara opção por modalizações da acção em vez de modalizações estativas.

Nos Quadros IV a VI, que dão conta dos lemas com maior índice de ocorrência no *corpus* de amostra, surgem destacados os lemas que são comuns aos três romances.

QUADRO IV

Lemas nominais com maior índice de ocorrência no *corpus* de amostra.

XIII		XV		XVI		Corpus	
Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.
conde	12	luar	23	senhor	10	Deus	35
Deus	12	sol	23	mar	8	cavalo	33
criado	10	areal	20	mouro	8	arma	27
rei	10	cavaleiro	20	arma	7	el-rei	26
caminho	9	gavião	20	carta	7	cavaleiro	25
cavalo	9	corpo	19	cavalo	7	corpo	25
conselho	9	arma	18	mão	7	homem	24
donzela	9	asa	18	mulher	7	sol	24
homem	9	pai	18	palácio	7	luar	23
amor	8	sorte	18	rua	7	areal	20
cabeça	8	cavalo	17	tabuleiro	7	gavião	20
cabelo	8	Deus	17	ano	6	mouro	20
el-rei	8	el-rei	17	corpo	6	senhor	20
ladrão	8	velho	17	Deus	6	cabeça	19
letrado	8	falsidade	15	direita	6	pai	19
passageiro	8	jantar	15	dom	6	sorte	19
Santiago	8	alto	14	lugar	6	velho	19
recado	7	chaga	14	par	6	asa	18
mal	6	homem	14	primo	6	vez	17
igreja	5	passar	14	entrada	5	lugar	16
mercado	5	senhor	13	terra	5	alto	15

Ainda que no caso do *corpus* de amostra com poucos dados, os Quadros IV a VI revelam-nos que existem lemas comuns a todos os romances (aqueles que surgem destacados em cada um dos quadros) e lemas particulares de cada um dos romances, por exemplo, <desonrar> em BRTOM XIII; <ensanguentar> em BRTOM XV ou <resgatar> em BRTOM XVI. Analisada a distribuição de lemas por versão dentro de cada um dos romances, embora esses dados não sejam aqui apresentados por gerarem quadros analíticos demasiado grandes, é-nos revelado que existem, também, lemas específicos de cada versão.

QUADRO V

Lemas adjetivais com maior índice de ocorrência no *corpus* de amostra

XIII		XV		XVI		Corpus	
Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.
bom	11	mortal	23	real	9	bom	29
sagrado	8	branco	19	bom	5	mortal	23
discreto	6	pequeno	18	carnal	5	branco	21
malogrado	3	bom	13	só	5	pequeno	18
santo	3	menos	10	gentil	2	real	18
direito	2	mau	9	alto	1	mau	12
mau	2	tremedal	9	branco	1	menos	10
secreto	2	real	8	cansado	1	tremedal	9
verde	2	pequenino	4	cativo	1	sagrado	8
averdegado	1	triste	4	chã	1	discreto	6
branco	1	igual	2	cheio	1	só	6
constipado	1	brando	1	choroso	1	carnal	5
contrito	1	cansado	1	cortês	1	triste	5
desgraçado	1	cativo	1	covarde	1	pequenino	4
divino	1	estremedal	1	cru	1	malogrado	3

QUADRO VI

Lemas verbais com maior índice de ocorrência no *corpus* de amostra

XIII		XV		XVI		Corpus	
Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.	Lema	nº de ocor.
casar	12	ter	23	estar	9	ter	40
dizer	12	ver	23	ir	9	ir	39
ir	12	entrar	22	ter	9	ver	35
prender	12	morrer	21	encontrar	8	dar	32
ser	12	achar	20	jogar	8	estar	32
dar	11	estar	20	poder	8	morrer	31
degolar	11	voar	20	querer	8	querer	31
matar	11	passar	19	sentar	8	ser	29
perdoar	10	faltar	18	ser	8	entrar	25
queixar	10	ir	18	ver	8	passar	25
deixar	9	mandar	18	acompanhar	7	matar	24
morrer	9	quedar	17	buscar	7	falar	23
querer	9	buscar	15	falar	7	buscar	22
enterrar	8	contar	15	dar	6	mandar	22
fazer	8	dar	15	resgatar	6	vir	22
haver de	8	ensanguentar	15	começar	5	faltar	21
ter	8	guardar	14	dizer	5	guardar	21
zombar	8	querer	14	emprestar	5	voar	21
desgraçar	7	falar	13	guardar	5	achar	20
haverde	7	vir	13	ouvir	5	encontrar	20
viver	7	abrir	12	pentear	5	haver	20
bondar	6	cair	12	vir	5	dizer	18
desonrar	6	lavar	12	voltar	5	haver de	18
dormir	6	tornar	12	haver de	4	quedar	17
entrançar	6	chorar	11	haver de	4	contar	16
haver	6	deitar	11	passar	4	fazer	16

Apesar de os dados apresentados nos quadros acima (Quadro IV a VI) possuírem um carácter exemplificativo, da sua análise é possível inferir que determinados campos semânticos (quer se trate de lemas nominais, adjectivais ou verbais) apresentam maior probabilidade de ocorrência no Romanceiro do que outros. Assim, tendo em conta o nosso conhecimento, não será de estranhar que termos como <cavalo>, verbos como <ir>, <ter> e <dar> e topónimos como <Santiago>, cujos índices de ocorrência no *corpus* de amostra são já significativos, possuam índice de ocorrência elevado no Romanceiro.

Os Quadros VII a XII dão conta dos dados estatísticos de lexemas do *corpus* de amostra e da sua distribuição por BRTOM, por área geográfica e por editor.

QUADRO VII

Distribuição de totais das categorias gramaticais Nome, Adjectivo e Verbo por Romance no *corpus* de amostra.

Romance BRTOM	Total de Palavras	Nome	%	Adjectivo	%	Verbo	%	% Total das 3 Classes
XIII	2.491	533	21,40	77	3,09	711	28,54	53,03
XV	4.091	959	23,44	162	3,96	911	22,27	49,67
XVI	3.249	722	22,22	97	2,99	831	25,58	50,78
TOTAL	9.831	2.214	22,52	336	3,42	2.453	24,95	50,89

O Quadro VII foi analisado em comparação com o Quadro III.

No que se refere aos dados apresentados no Quadro VIII (Distribuição das categorias gramaticais Nome, Adjectivo e Verbo por área Geográfica no *corpus* de amostra), há que ter em consideração o facto de o número de textos do *corpus* de amostra não provir de forma equitativa de todas as regiões do país, havendo maior número de textos cuja proveniência é Trás-os-Montes, nomeadamente o distrito de Bragança. Por outro lado, é importante lembrar que existe um número significativo de romances da tradição oral moderna portuguesa que existem apenas na província de Trás-os-Montes, caso do romance “Morte de D. Beltrão” (BRTOM XV) que incluímos no *corpus* de amostra. Daí a discrepância no que se refere aos valores totais específicos apresentados para cada distrito.

QUADRO VIII

Distribuição das categorias gramaticais Nome, Adjectivo e Verbo por área Geográfica no *corpus* de amostra

Área Geográfica	Nome	%	Adjectivo	%	Verbo	%
Bragança	1.079	48,74	168	50,00	1.181	48,15
Castelo Branco	76	3,43	6	1,79	107	4,36
Vila Real	116	5,24	22	6,55	105	4,28
Loc. Atribuída	298	13,46	41	12,20	316	12,88
Sem informação	68	3,07	3	0,89	101	4,12
Versão Factícia	577	26,06	96	28,57	643	26,21
TOTAL	2.214		336		2.453	

Assim, muito embora os dados referentes à distribuição de Nome, Adjectivo e Verbo por área geográfica sejam mais relevantes no sub-*corpus* dos romances carolíngios do que neste *corpus* de amostra, visto que o número de versões provenientes de outras regiões do país, de versões sem indicação de proveniência geográfica e de versões com localização

atribuída aumenta significativamente, há que ter em atenção que, também no *corpus* constituído pelos romances carolíngios, os textos oriundos do distrito de Bragança superam em número os textos originários de outros distritos.

Os dados apresentados no Quadro IX têm de ser analisados à luz dos dados contidos nos Quadros X a XII.

QUADRO IX

Distribuição de totais das categorias gramaticais Nome, Adjectivo e Verbo por grupo de editor no *corpus* de amostra.

238

a) Nome

Editor	XIII		XV		XVI		TOTAL	%
	Nº	%	Nº	%	Nº	%		
Garrett	101	17,5	97	16,8	379	65,7	577	26,1
Outros	432	26,4	862	52,7	343	21,0	1.637	73,9
Total	533	24,1	959	43,3	722	32,6	2.214	100,0

b) Adjectivo

Editor	XIII		XV		XVI		TOTAL	%
	Nº	%	Nº	%	Nº	%		
Garrett	20	20,8	21	21,9	55	57,3	96	28,6
Outros	57	23,8	141	58,8	42	17,5	240	71,4
Total	77	22,9	162	48,2	97	28,9	336	100,0

c) Verbo

Editor	XIII		XV		XVI		TOTAL	%
	Nº	%	Nº	%	Nº	%		
Garrett	109	17,0	100	15,6	434	67,5	643	26,2
Outros	602	33,3	811	44,8	397	21,9	1.810	73,8
Total	711	29,0	911	37,1	831	33,9	2.453	100,0

QUADRO X

Distribuição de textos e de palavras por editor no *corpus* de amostra

Editor	Textos /autor		Total palavras / autor	
	Nº	%	Nº	%
Braga	1	2,3	380	3,9
Garrett	3	6,8	2.534	25,8
Leite	28	63,6	4.642	47,2
Lopo	1	2,3	171	1,7
Luís Chaves	1	2,3	259	2,6
Martins	4	9,1	663	6,7
Sepúlveda	1	2,3	227	2,3
Tavares	4	9,1	642	6,5
Thomás	1	2,3	313	3,2
TOTAL	44	100,0	9.831	100,0

QUADRO XI

Distribuição de Nome, Adjectivo e Verbo por Editor na *corpus* de amostra

Editor	Nome		Adjectivo		Verbo	
	Nº	%	Nº	%	Nº	%
Braga	76	3,4	6	1,8	107	4,4
Garrett	577	26,1	96	28,6	643	26,2
Leite	1.060	47,9	153	45,5	1.142	46,6
Lopo	40	1,8	11	3,3	33	1,3
Luis Chaves	61	2,8	8	2,4	56	2,3
Martins	134	6,1	24	7,1	163	6,6
Sepúlveda	59	2,7	10	3,0	50	2,0
Tavares	139	6,3	25	7,4	158	6,4
Thomás	68	3,1	3	0,9	101	4,1
TOTAL	2.214	100,0	336	100,0	2.453	100,0

239

QUADRO XII

Total de palavras em cada versão do romance “D. Gaifeiros”

BRTOM	Editor	Ano da 1ª edição	Texto	Total de Palavras
XVI	Garrett II	1851	110	1696
XVI	Leite	1958	111	331
XVI	Leite	1958	112	109
XVI	Leite	1958	113	124
XVI	Leite	1958	114	282
XVI	Martins	1928	115	243
XVI	Martins	1938	116	73
XVI	Tavares	1903	117	187
XVI	Leite	1958	118	204

Da análise dos dados do Quadro IX que dá conta da distribuição dos totais de Nome, Adjectivo e Verbo por grupo de editor na *corpus* de amostra, no que se refere ao romance “D. Gaifeiros” (BRTOM XVI), constata-se que o editor Garrett apresenta mais nomes, mais adjectivos e mais verbos do que o grupo Outros composto por três editores (Leite, Martins e Tavares).

Quando comparado este quadro com o Quadro XII que nos mostra que só a versão garrettiana possui 1696 palavras, isto é, mais palavras do que as restantes versões que, no total das oito versões, possuem 1553 palavras, verificamos que a versão de Garrett do romance “D. Gaifeiros” se distingue não apenas porque uma só versão é bastante maior do que qualquer das outras oito versões, mas também porque uma única versão apresenta mais lexemas nominais, adjectivais e verbais do que o cômputo das restantes oito versões comparadas.

O Quadro XII mostra-nos, ainda, que, à excepção da versão de Martins editada em 1938, todas as restantes versões possuem uma variação pouco significativa ao nível do número total de palavras. Será possível que entre 1851, data de edição da versão de Garrett, e 1903, data de edição da versão de Tavares, se tenha dado um decréscimo tão acentuado no número de palavras?

Tendo, ainda, em consideração a análise comparativa estatística da presença de outras classes gramaticais de palavras nas diferentes versões do romance, nomeadamente os

advérbios, as preposições e as conjunções (dados que, pela sua extensão, não nos é possível sintetizar), no caso da versão garrettiana do romance, parece ser-nos possível afirmar que o seu léxico confirma o que já todos sabíamos: que se trata de uma versão factícia.

Em relação aos dados dos Quadros X e XI é de salientar que, tal como acontece para o *corpus* de romances carolíngios, a Leite se deve o maior número de versões de cada romance pelo que os totais de lexemas nominais, adjectivais e verbais possuem percentagens mais elevadas para este editor.

O último quadro, Quadro XIII, que dá conta da distribuição de tempos verbais em cada um dos romances do *corpus* de amostra, revela-nos que nem todos os tempos verbais possuem igual importância para todos os romances e que nem todos os tempos verbais são utilizados com a mesma acuidade, confirmando-se, deste modo, parte das teorias de Sandmann⁵, de Szertics⁶, de Marínez-Gil⁷, de Mirrer⁸ e de Webber,⁹ no que se refere aos tempos verbais característicos do género romancístico e no que se refere à sua íntima relação com a rima.

QUADRO XIII

Distribuição de tempos verbais por BRTOM no *corpus* de amostra

Tempo Verbal	XIII		XV		XVI		TOTAL	%
	Nº	%	Nº	%	Nº	%		
PRE IND	172	28,9	186	31,2	238	39,9	596	6,1
PERF IND	78	21,3	198	54,0	91	24,8	367	3,7
IMPF IND	42	24,7	74	43,5	54	31,8	170	1,7
MPER IND	16	47,1	9	26,5	9	26,5	34	0,3
FUT IND	12	46,2	3	11,5	11	42,3	26	0,3
PRE CONJ	61	43,9	33	23,7	45	32,4	139	1,4
IMPF CONJ	6	46,2	1	7,7	6	46,2	13	0,1
FUT CONJ	1	20,0	0	0,0	4	80,0	5	0,1
INF FLEX	4	8,3	34	70,8	10	20,8	48	0,5
IMP	23	25,0	43	46,7	26	28,3	92	0,9
COND	1	25,0	3	75,0	0	0,0	4	0,0
INF NFLEX	100	15,0	258	38,7	308	46,2	666	6,8
GER	2	5,3	27	71,1	9	23,7	38	0,4
PCP	42	84,0	2	4,0	6	12,0	50	0,5
PCP-A	151	73,7	40	19,5	14	6,8	205	2,1
TOTAL	711	29,0	911	37,1	831	33,9	2453	25,0

⁵ Sandmann, Manfred, "La 'mezcla de los tiempos narrativos' en el Romancero Viejo", in *Romanistisches Jahrbuch*, XXV, Berlin/New York, 1974, pp. 278-293.

⁶ Szertics, Joseph, "Observaciones sobre algunas funciones estilísticas del pretérito indefinido en el romancero viejo", in *Explicación de Textos Literarios*, vol. II, Sacramento (California), Dep. of Spanish and Portuguese - California State University, 1974, pp. 189-197 e Szertics, Joseph, "Tiempo verbal y asonancia en el Romancero Viejo", in *Homenaje a Don Agapito Rey*, Bloomington, Indiana University, 1980, pp. 179-194.

⁷ Martínez-Gil, Fernando, "Las inversiones del orden de palabras en el Romancero", in *Hispania*, nº 4, 1989, Dezembro, pp. 895-908.

⁸ Mirrer, Louise, "The characteristic patterning of romancero language: some notes on tense and aspect in the romances viejos", in *Hispanic Review*, nº 55, 1987, Autumn, pp. 441-461.

⁹ Webber, Ruth House, *Formulistic Diction in the Spanish Ballad*, Berkeley/Los Angeles, University of California Press, 1951.

Da análise dos dados do Quadro XIII, constatamos que, nestes três romances, existem tempos verbais usados por excelência e tempos verbais com pouquíssima expressão. Os mais utilizados são o presente, o perfeito e o imperfeito do indicativo, o presente do conjuntivo, o imperativo, o infinitivo não flexionado e o particípio passado com valor adjectival e os menos utilizados o mais-que-perfeito, o futuro do indicativo, os imperfeito e futuro do conjuntivo, o condicional, o gerúndio e o particípio passado sem função adjectival.

No entanto, muito embora neste *corpus* de amostra tal não se verifique, é importante referir que o mais-que-perfeito, o condicional e o particípio passado são tempos verbais bastante utilizados no romanceiro, sobretudo algumas das suas pessoas verbais em final de verso, tal como o provam os estudos citados de Szertics, Mirrer, Sandmann e de Martínez-Gil.

No Quadro XIII, verificamos, também, que, apesar de ser possível encontrar qualquer dos tempos verbais da língua portuguesa em qualquer romance, excepto o futuro do conjuntivo em “Morte de D. Beltrão” (BRTOM XV) e o condicional em “D. Gaifeiros” (BRTOM XVI), cada um dos romances do *corpus* de amostra apresenta uma clara opção pela utilização maioritária de determinados tempos verbais em detrimento do uso de outros tempos.

Assim, no “Conde Preso” (BRTOM XIII), cuja rima é *a/o*, os tempos verbais com maior número de ocorrências são o presente do indicativo e o particípio passado com função adjectival, surgindo este último maioritariamente em fim de verso. No caso do romance “Morte de D. Beltrão” (BRTOM XV), cuja rima é *a/e*, os tempos verbais com maior índice de ocorrência são o presente do indicativo, o perfeito do indicativo e o infinitivo não flexionado, surgindo este último maioritariamente em fim de verso. E no caso de “D. Gaifeiros” (BRTOM XVI), cuja rima é *a/e*, os tempos verbais com maior índice de ocorrência são o presente do indicativo e o infinitivo não flexionado, surgindo, também, este último maioritariamente em fim de verso.

Tal como já afirmámos aquando da análise dos Quadros IV a VI, também o Quadro XIII nos revela que no Romanceiro da tradição oral moderna portuguesa, embora seja possível verificar a presença de todos os tempos verbais, se constata uma tendência para o uso maioritário de determinados tempos verbais em detrimento de outros.

4 – Considerações Finais

Com base nos dados estatísticos que já possuímos no seguimento do trabalho de investigação que continuamos a desenvolver, dos quais a análise acima apresentada é apenas uma amostra, cremos poder afirmar que, para o *corpus* base deste estudo e por extensão para outros sub-*corpora* do Romanceiro, o léxico se pode diferenciar em três níveis distintos: o do léxico comum, isto é, o conjunto de palavras presentes em todos os romances de um mesmo tema; o do léxico específico de cada romance, ou seja, aquele que é comum a todas as versões de um mesmo romance e, finalmente, o do léxico particular de cada versão que corresponde ao léxico específico de cada informante. O cômputo dos léxicos de cada um dos sub-*corpora* que constituem o Romanceiro da Tradição Oral Moderna Portuguesa dar-nos-á o léxico do Romanceiro da Tradição Oral Moderna Portuguesa.

Da análise comparativa dos dados, podemos inferir que o Romancero é um género literário que recorre sobretudo a nominalizações e verbalizações e muito raramente a adjectivalizações, e que este facto, que lhe confere especificidade quando comparado com outros géneros literários, advém indubitavelmente do carácter dinâmico das estruturas abertas, em palavras de Diego Catalán, que o compõem, muito mais receptivas a modalizações da acção do que a modalizações estativas.

Para finalizar, da presença notoriamente elevada de categorias gramaticais específicas como os advérbios <não>, <lá>, <já> e <aqui>, das conjunções coordenativas copulativas <e> e <nem>, de formas pronominais e das preposições <a>, <com>, <de>, <para>, <por> e contraídas ou simples, podemos deduzir que, para além de um léxico específico, o género romancístico possui uma sintaxe própria com modalizações temporais e locativas também elas específicas.

BIBLIOGRAFIA

- Catalán, Diego, *El Romancero Pan-hispánico — Catálogo General Descriptivo*, vol. II e III, Madrid, Seminario Menéndez Pidal/Universidad Complutense de Madrid, 1982 e 1983, respectivamente.
- Ferré, Pere, *Romancero Português da Tradição Oral Moderna*, Vol. 1, Lisboa, Fundação Calouste Gulbenkian, 2000.
- Ferré, Pere e Carinhas, Cristina, *Bibliografia do Romancero Português da Tradição Oral Moderna*, Madrid, Seminario Menéndez Pidal/Universidad Complutense de Madrid, 2000.
- Martínez-Gil, Fernando, “Las inversiones del orden de palabras en el *Romancero*”, in *Hispania*, nº 4, 1989, Dezembro, pp. 895-908.
- Mirrer, Louise, “The characteristic patterning of *romancero* language: some notes on tense and aspect in the *romances viejos*”, in *Hispanic Review*, nº 55, 1987, Autumn, pp. 441-461.
- Sandemann, Manfred, “La ‘mezcla de los tiempos narrativos’ en el *Romancero Viejo*”, in *Romanistisches Jahrbuch*, XXV, Berlin/New York, 1974, pp. 278-293.
- Szertics, Joseph, “Observaciones sobre algunas funciones estilísticas del pretérito indefinido en el *romancero viejo*”, in *Explicación de Textos Literarios*, vol. II, Sacramento (California), Dep. of Spanish and Portuguese – California State University, 1974, pp. 189-197.
- Szertics, Joseph, “Tiempo verbal y asonancia en el *Romancero Viejo*”, in *Homenaje a Don Agapito Rey*, Bloomington, Indiana University, 1980, pp. 179-194.
- Webber, Ruth House, *Formulistic Diction in the Spanish Ballad*, Berkeley/Los Angeles, University of California Press, 1951.