

UNIDADE E DIVERSIDADE: O VOCABULÁRIO DA LÍNGUA ORAL EM PORTUGAL E NO BRASIL

MARIA TERESA BIDERMAN
(UNESP)

MARIA FERNANDA BACELAR DO NASCIMENTO
(CLUL)

Nos últimos dez anos tem-se observado um interesse crescente pela criação de recursos linguísticos, nomeadamente de *corpora* e de léxicos de grandes dimensões que constituem hoje infra-estruturas indispensáveis à realização de projectos de investigação e desenvolvimento sobre as línguas, assim como à criação de aplicações baseadas em dados reais.

No que respeita ao português, tanto em Portugal como no Brasil, como também em alguns países africanos de língua oficial portuguesa (como é o caso de Moçambique), estão já disponíveis recursos, em larga escala, que começam a tornar possível a realização de estudos contrastivos entre variedades do português, visando estabelecer factores de unidade e de diversidade entre elas, assentes em dados autênticos e caracterizados quantitativa e qualitativamente.

Pretendemos hoje descrever aqui sucintamente dois grandes *corpora* existentes em Portugal e no Brasil, dar conta de algumas das explorações e análises lexicais que sobre elas vimos fazendo e apresentar dados contrastivos entre o português europeu (PE) e o português do Brasil (PB), extraídos de sub-*corpora* orais que deixam entrever as grandes potencialidades que oferecem os *corpora*, em associação com ferramentas informáticas de extracção de dados e de conhecimentos linguísticos.

No Quadro 1A apresentamos um descrição sucinta do *corpus* PB.

Quadro 1A
CORPUS DO PORTUGUÊS CONTEMPORÂNEO DO BRASIL
UNESP, Campus de Araquara, SP

A) língua escrita (de 1950 a 1998) - 30 milhões de palavras
composição do *corpus*:

1. *linguagem literária*: romances, contos, peças de teatro, roteiros de telenovelas e de filmes, poesia.

2. *linguagem técnico-científica*: principais áreas do conhecimento - obras de divulgação - nível: estudantes universitários e público em geral e não de pesquisa acadêmica.

3. *linguagem jornalística*:

- principais jornais das capitais brasileiras de circulação em todo o país;
- revistas de informação geral de circulação nacional como *Veja*, *Isto é*, *Manchete*, etc.;
- revistas temáticas como *Placar* (futebol e esportes), *Pais & Filhos* (educação e psicologia), etc.;

4. *linguagem oratória*:

- discursos presidenciais e do parlamento brasileiro (Senado e Câmara Federal,
- sermões religiosos.

5. *traduções*

B) língua falada: 963.535.palavras

(falantes de idades de 20 a 70 anos e de todos os níveis educacionais - do analfabeto ao formado em universidade)

Vários projectos de língua oral de diferentes regiões do Brasil:

- projecto NURC (norma urbana culta) das principais capitais do país (Recife, Salvador, Rio de Janeiro, São Paulo, Porto Alegre) - década de 70;
- projecto PEUL do Rio de Janeiro (informantes com escolaridade primária e/ou secundária) - década de 80;
- projecto da fala de Maceió (Alagoas);
- a fala de estudantes universitários de Curitiba (Paraná);
- banco de dados da PUC (Pontifícia Universidade Católica de S. Paulo), S. Paulo (entrevistas com executivos e empresários; sessões administrativas da PUC; entrevistas da televisão);
- projecto da linguagem rural da UNESP, Campus de Assis (entrevistas com falantes analfabetos - bóias-frias - do interior do estado de S. Paulo).

O *Corpus* de Referência do Português contemporâneo (CRPC) do Centro de Linguística da Universidade de Lisboa é um *corpus* linguístico que contém actualmente 83,3 milhões de palavras, sendo constituído por amostragens de diversos tipos de textos de discurso escrito (literários, jornalísticos, técnicos, científicos, didácticos, económicos, jurídicos, parlamentares, etc.) e de discurso oral (elocuições informais e formais).

Estas amostragens dizem respeito a variedades nacionais e regionais do português; de facto, estão representados no *corpus* o português europeu, o português do Brasil, o português dos cinco países africanos de língua oficial portuguesa e o português de Macau. Do ponto de vista cronológico, o *corpus* contém textos que vão desde a 2ª metade do século XIX até 1998, sendo, na sua maior parte, posteriores a 1970.

No Quadro 1B apresentamos dados sobre a constituição e dimensão do *corpus* português - PE¹.

Quadro 1B

Dimensão total: 83,3 milhões de palavras																																					
<table border="1"> <tr> <td>ESCRITO 81617671</td> <td>→</td> <td>PORTUGAL 78460676</td> </tr> <tr> <td>ORAL 1725240</td> <td></td> <td>ANGOLA 60976</td> </tr> <tr> <td></td> <td></td> <td>BRASIL 4009953</td> </tr> <tr> <td></td> <td></td> <td>CABO VERDE 534013</td> </tr> <tr> <td></td> <td></td> <td>GUINÉ 46984</td> </tr> </table>	ESCRITO 81617671	→	PORTUGAL 78460676	ORAL 1725240		ANGOLA 60976			BRASIL 4009953			CABO VERDE 534013			GUINÉ 46984		<table border="1"> <tr> <td colspan="2">LIVROS 20328252</td> </tr> <tr> <td>DIDÁCTICO</td> <td>1968143</td> </tr> <tr> <td>LITERÁRIO</td> <td>8593249</td> </tr> <tr> <td>TÉCNICO E CIENTÍFICO</td> <td>9766860</td> </tr> <tr> <td colspan="2">JORNAIS 48052196 (5044298 - SEMANÁRIO)</td> </tr> <tr> <td colspan="2">REVISTAS 5982879</td> </tr> <tr> <td colspan="2">DOCUMENTOS 2999799</td> </tr> <tr> <td>JURÍDICOS</td> <td>1437667</td> </tr> <tr> <td>PARLAMENTARES</td> <td>1482132</td> </tr> <tr> <td colspan="2">FOLHETOS 322538</td> </tr> </table>	LIVROS 20328252		DIDÁCTICO	1968143	LITERÁRIO	8593249	TÉCNICO E CIENTÍFICO	9766860	JORNAIS 48052196 (5044298 - SEMANÁRIO)		REVISTAS 5982879		DOCUMENTOS 2999799		JURÍDICOS	1437667	PARLAMENTARES	1482132	FOLHETOS 322538	
	ESCRITO 81617671	→	PORTUGAL 78460676																																		
	ORAL 1725240		ANGOLA 60976																																		
			BRASIL 4009953																																		
			CABO VERDE 534013																																		
		GUINÉ 46984																																			
LIVROS 20328252																																					
DIDÁCTICO	1968143																																				
LITERÁRIO	8593249																																				
TÉCNICO E CIENTÍFICO	9766860																																				
JORNAIS 48052196 (5044298 - SEMANÁRIO)																																					
REVISTAS 5982879																																					
DOCUMENTOS 2999799																																					
JURÍDICOS	1437667																																				
PARLAMENTARES	1482132																																				
FOLHETOS 322538																																					
<table border="1"> <tr> <td>ANTES DE 1900 1000000</td> <td></td> <td></td> </tr> </table>	ANTES DE 1900 1000000																																				
ANTES DE 1900 1000000																																					

1901-1970 2600000	→	MACAU 7997	CORRESPONDÊNCIA 163192
DEPOIS DE 1970 79730911		MOÇAMBIQUE 210321	
		S. TOMÉ 12000	
			VARIA 3848815

O exame dos dados quer ordenados alfabeticamente, quer por Frequências, quer em listas de concordâncias, é muito esclarecedor sobre os empregos, no *corpus*, das unidades lexicais que vimos estudando, pois tornam patentes certas regularidades que a simples leitura cursiva dos textos, naturalmente, não evidencia.

Por outro lado, a análise dos contextos e o estudo quantitativo dos factos linguísticos comparáveis permite descrever com objectividade a variação dos fenómenos evidenciados cujos contornos sintácticos e semânticos são difíceis de delimitar de forma apriorística (Haber, B., 1997).

A análise extensiva de concordâncias de palavras ou de associações de palavras permite, ainda, o estabelecimento de tipologias de variação do português respeitantes quer a unidades lexicais isoladas e seus paradigmas, quer a recorrências sintagmáticas: co-ocorrentes, combinatórias, fraseologias.

Os exemplos que vamos apresentar resultaram da comparação entre os seguintes conjuntos de materiais:

1º conjunto de materiais (C1):

1.179 vocábulos com frequência maior ou igual a 40, extraídos de um *corpus* de português europeu de 700.000 palavras, o *corpus* oral do *Português Fundamental*. Estes vocábulos foram comparados com um repertório de 1.572 vocábulos que ocorreram com frequência maior ou igual a 40 num *corpus* oral do português brasileiro de 963.535 palavras (cf. quadro anterior).

2º conjunto de materiais (C2):

Dados extraídos de dois sub-*corpora* orais do *Corpus de Referência do Português Contemporâneo*, cada um deles com uma dimensão da ordem das 550.000 palavras, sendo um do português europeu e outro do português do Brasil.

Estabelecidos estes dois conjuntos de materiais, encontramos-nos, presentemente, numa fase exploratória dos léxicos e dos *corpora*, com vista à elaboração de um primeiro plano de estudos lexicais contrastivos. Assim, apresentaremos hoje, apenas, alguns exemplos, aleatórios, obtidos aquando da extracção de dados provenientes dos dois conjuntos de materiais acima referidos; são exemplos ilustrativos de aspectos lexicais que pretendemos venham a constar dos estudos a fazer, de forma sistemática, sobre estas duas variedades do Português.

No Quadro 2 apresentamos um excerto das listas de vocábulos com Frequência igual ou superior a 40, extraídos de C1. Esta lista está ordenada alfabeticamente de forma a permitir a observação contrastiva dos dados de Frequência e da ocorrência ou não-ocorrência dos vocábulos nos *corpora* orais comparáveis das duas variantes do português.

Quadro 2

Excerto da lista de vocabulário comparado (PE/PB) com frequência igual ou superior a 40 em *corpora* orais

PE		PB	
	Frequência		Frequência
CÁ	1580	CÁ	408
CABEÇA	97	CABEÇA	404
CABELO	56	CABELO	170
CABO	108		
CAÇA	41		
		CACHORRO	179
CADA	346	CADA	700
CADEIRA(S)	68	CADEIRA	86
		CADERNO	45
CAFÉ	195	CAFÉ	294
CAIR	150	CAIR	317
CAIXA, s.f.	173	CAIXA	49
CALÇAS	43	CALÇA	68
CALHAR	110		
		CALMO	54
CALOR	63	CALOR	57
CAMA	120	CAMA	92
CAMADA	62		

ACTAS DO XIV ENCONTRO NACIONAL DA APL

		CAMARADA	107
		CAMINHÃO	116
CÂMARA	124		
CAMINHO	62	CAMINHO	102
CAMIONETE	69		
CAMPO	255	CAMPO	314
		CANA	191
		CANAL	42
CANÇÃO	42		
		CANSADO	82
		CANSAR	41
CANTAR	89	CANTAR	152
		CANTO	61
		CANTOR	42
CÃO	67		
		CAPACIDADE	43
CAPAZ	206	CAPAZ	90

Como já foi dito, só estão representados, nestas listas, os vocábulos com frequência igual ou superior a 40. Assim, o facto de haver vocábulos não indexados numa das variantes — PE ou PB — não significa, necessariamente, que eles não tenham ocorrido no *corpus*; significa, sim, que não atingiram a frequência 40. Por exemplo, no *corpus* PE, os vocábulos não incluídos, e que constam da lista PB, atingiram as seguintes frequências: cachorro-4, caderno-35, calmo-14, camarada-22, caminhão-0, cana-12, canal-8, canto-36, cantor-2, capacidade-27. E no PB, vocábulos não incluídos por não atingirem a frequência 40 ocorreram, porém, e com os seguintes quantitativos: cabo-36, caça-22, calhar-2, camada-25, câmara-24, camionete-2, canção-2 e cão-22.

Por outro lado, as altas frequências de alguns vocábulos podem dever-se ao facto de eles terem ocorrido como elementos de formação de locuções e expressões cristalizadas cujo uso difere em Portugal e no Brasil, pelo que se impõe fazer uma análise contrastiva das associações lexicais.

Estas listagens constituem, pois, uma fonte de indícios que iremos seguir, para que à informação quantitativa se venha juntar a caracterização morfossintáctica, semântica e pragmática. A simples observação destas listas começa a apontar caminhos de pesquisa como, por exemplo, a do uso, nas duas variantes, de formas sinónimas cujas frequências se apresentam em distribuição complementar. É o caso de *inclusivamente* e *inclusive* apresentado no Quadro 3.

Quadro 3
Vocábulo em distribuição complementar

	PE	PB
INCLUSIVAMENTE	117	1
INCLUSIVE	28	534

Outro aspecto a estudar diz respeito à variação observada nos advérbios em *-mente*.

O 2º. conjunto de materiais atrás citado (C2) consta de um *corpus* oral do português europeu contendo 565.639 palavras e de um *corpus* oral do português do Brasil contendo 540.815 palavras. Trata-se, pois, também de dois *corpora* comparáveis dos quais extraímos todas as formas adverbiais em *-mente*. Constatámos que a ocorrência e variedade de advérbios em *-mente* em PE é muito superior à obtida em PB. Assim, em PE, os advérbios em *-mente* atingiram a Frequência de 5.108 com 354 formas diferentes, enquanto em PB atingiram a Frequência de 2.706 com 245 formas diferentes.

Das listas comparativas apresentamos, no Quadro 4, excertos retirados de vários pontos da lista alfabética.

Quadro 4
Excertos da lista comparativa de advérbios em *-mente* extraída de *corpora* do português europeu e do português do Brasil

	PE		PB
abertamente	1		
		abruptamente	1
absolutamente	132	absolutamente	40
abundantemente	1	abundantemente	1
acentuadamente	1		
activamente	2		
actualmente	48	atualmente	62
adequadamente	1		
		acidentalmente	1
admiravelmente	3	admiravelmente	3
agradavelmente	1		
airosamente	2		
		adoidadamente	1
		agudamente	1

ACTAS DO XIV ENCONTRO NACIONAL DA APL

altamente	13	altamente	9
		alternadamente	1
		altissimamente	1
		amorosamente	1
amplamente	2	amplamente	1
anarquicamente	1		
		analogamente	1
anteriormente	5	anteriormente	9
automaticamente	18	automaticamente	11
basicamente	3	basicamente	26
bestialmente	20		
botanicamente	1		
brevemente	3		
brilhantemente	1		
		biologicamente	2
		brabamente	1
brutalmente	3	brutalmente	1
calculadamente	1		
calmamente	5	calmamente	2
capazmente	1		
geralmente	111	geralmente	117
globalmente	1	globalmente	3
		gradativamente	2
gradualmente	2	gradualmente	1
grandemente	8		
		graficamente	1
		gramaticamente	1
gratuitamente	4	gratuitamente	2
habilmente	1		
habitualmente	2	habitualmente	4
harmoniosamente	3		
		hierarquicamente	1
historicamente	1	historicamente	4
honestamente	3	honestamente	11
horriavelmente	1	horriavelmente	1
humanamente	1	humanamente	3
igualmente	5		
imediatamente	54	imediatamente	17

UNIDADE E DIVERSIDADE: O VOCABULÁRIO DA LÍNGUA ORAL EM PORTUGAL E NO BRASIL

imensamente	1	imensamente	3
implicitamente	3	implicitamente	1
imprevistamente	1		
incessantemente	2		
inclusivamente	114		
incompreensivelmente	1		
inconscientemente	2	inconscientemente	5
incrivelmente	1	incrivelmente	1
indefinidamente	1		
intensamente	2		
intensivamente	1		
interiormente	4		
internacionalmente	1		
internamente	1		
		intencionalmente	1
intimamente	1	intimamente	3
invariavelmente	1		
isoladamente	3		
judicialmente	1		
juntamente	9	juntamente	1
justamente	37	justamente	109
lamentavelmente	1		
lateralmente	1		
legalmente	3		
		largamente	1
legitimamente	5	legitimamente	1
naturalmente	89	naturalmente	83
necessariamente	11	necessariamente	4
negativamente	1		
neuroticamente	1		
nitidamente	13	nitidamente	1
nomeadamente	32		
normalíssimamente	1		
normalmente	184	normalmente	161
		notadamente	2
novamente	50	novamente	18
simplesmente	109	simplesmente	37
simultaneamente	9		

ACTAS DO XIV ENCONTRO NACIONAL DA APL

sinceramente	15	sinceramente	16
sistematicamente	3	sistematicamente	5
socialmente	3	socialmente	4
solenemente	2		
solidariamente	1		
		sociologicamente	1
somente	18	somente	13
subitamente	1		
substancialmente	1		
sucessivamente	12	sucessivamente	4
sucintamente	1		
suficientemente	10	suficientemente	2
superficialmente	1		
superiormente	1		
surpreendentemente	2		
tacitamente	2		
tardiamente	2		
		supostamente	2
tecnicamente	7	tecnicamente	4
telefonicamente	1		
temporariamente	1		
teoricamente	2	teoricamente	7
terminantemente	1	terminantemente	2
terrivelmente	5	terrivelmente	6
textualmente	3		
tipicamente	2	tipicamente	5
torrencialmente	2		
totalmente	42	totalmente	50
tradicionalmente	3	tradicionalmente	1
tranquilamente	2	tranquilamente	1

Os valores obtidos relativamente à frequência destes advérbios suscitam também a observação dos seus aspectos funcionais. Nessa intenção observámos o comportamento colocacional dos advérbios em *-mente* como modificadores de adjectivos. Este processo, muito produtivo em português, parece comprovar a observação de Bolinger (1972) que considera que virtualmente todo o advérbio que modifica um adjectivo tem tendência a ter ou a desenvolver um sentido intensificador. Nos casos que observamos, a

colocação de um advérbio como modificador de um adjetivo induz uma interpretação de gradação ou de intensificação.

Esses advérbios apresentam diferenciações semânticas importantes tornando muito menos restritivas do que seria de prever as seleções combinatórias. Por exemplo, a co-ocorrência dos elementos da combinação "redondamente enganado" só se justifica pelo valor semântico de intensificador adquirido pelo advérbio.

Uma abordagem baseada apenas na competência rejeitaria combinações de advérbios com adjetivos que os *corpora* registam e que, em alguns casos, o uso cristalizou. O advérbio *totalmente*, que pelo seu semantismo próprio atribui uma quantificação máxima ao adjetivo, detém uma frequência semelhante nos dois *corpora*, assim como um uso muito variado de co-ocorrentes adjetivais sendo em ambos o adjetivo *diferente* o mais frequente, conforme se pode observar no Quadro 5.

Quadro 5
Totalmente + Adjectivo

PE	Freq.	PB	Freq.
TOTALMENTE	42	TOTALMENTE	50
Totalmente + adjectivo	41	Totalmente + adjectivo	38
abandonada	1	alterado	1
aberto	1	dependente	1
acompanhado	1	desvinculada	1
agressivo	1	determinados	1
desactualizado	1	diferante	11
desorganizado	1	disperso	2
desprotegido	1	dissolvido	1
destruída	2	errônea	1
detalhado	1	fechada	1
diferente	21	grande	1
enganado	1	ignorante	1
inventada	1	inferior a	1
negativo	2	inoperante	1
oposto	1	insolúveis	3
realizada	1	isolados	1
realizado	2	livre(-s)	2
reformada	1	marginalizada	1
ultrapassado	1	oposta(-o,-s)	3

ACTAS DO XIV ENCONTRO NACIONAL DA APL

		preto	1
		subordinado	1
		surda	1
		vazio	1

Um outro aspecto lexical em observação diz respeito ao uso de diminutivos que ocorreram em C2. No caso dos diminutivos em *-inho* e diferentemente do que se observou para os advérbios em *-mente*, há uma maior convergência de resultados. Assim, no *corpus* PE registaram-se 1.642 ocorrências de diminutivos *-inho*, das quais 707 formas diferentes. No *corpus* PB registaram-se 1.847 ocorrências e 772 formas diferentes.

O Quadro 6A constitui um excerto da lista comparativa dos diminutivos em *-inho*, até à Frequência 10.

Quadro 6A

Excerto da lista comparativa dos diminutivos mais frequentes (até à Freq. 10) extraídos de *corpora* orais do português europeu e do português do Brasil

PE	Freq.		PB	Freq.	
BAIXINHO	10		BAIXINHO	16	
baixinha	3		baixinha	3	
baixinhas	2		baixinho	13	
baixinho	5				
BARRINHA	10	(PB - 0)			
			BICHINHO	12	(PE - 2)
			bichinho	10	
			bichinhos	2	
BOCADINHO	398		BOCADINHO	12	
bocadinho	384		bocadinho	11	
bocadinhos	14		bocadinhos	1	
BOCADITO	24	(PB - 0)			
bocadito	22				
bocaditos	2				

UNIDADE E DIVERSIDADE: O VOCABULÁRIO DA LÍNGUA ORAL EM PORTUGAL E NO BRASIL

			BONITINHO	21	(PE - 3)
			bonitinha	7	
			bonitinhas	1	
			bonitinho	12	
			bonitinhos	1	
CANTINHO	13	(PB - 3)			
			CARRINHO	24	(PE - 3)
			carrinho	22	
			carrinhos	2	
CASINHA	12	(PB - 8)			
casinha	11				
casinhas	1				
			CERTINHO	15	(PE - 2)
			certinha	3	
			certinho	11	
			certinhos	1	
COISINHA	36		COISINHA	26	
coisinha	19		coisinha	22	
coisinhas	17		coisinhas	4	
COITADINHO	26	(PB - 9)			
coitadinha	5				
coitadinho	15				
coitadinhos	6				
			DIREITINHO	19	(PE - 0)
			FESTINHA	20	(PE - 4)
			festinha	19	
			festinhas	1	
			FRESQUINHO	21	(PE - 5)
			fresquinha	6	

ACTAS DO XIV ENCONTRO NACIONAL DA APL

			fresquinho	15	
			GRACINHA	13	(PE - 2)
			gracinha	11	
			gracinhas	2	
HOMENZINHO	17	(PB - 0)			
homenzinho	13				
homenzinhos	4				
LUISINHO	17	(PB - 1)			
MÃEZINHA	11	(PB - 2)			
MOCINHO	12		MOCINHO	19	
mocinha	6		mocinha	15	
mocinhas	3		mocinhas	1	
mocinhos	3		mocinhos	3	
MULHERZINHA	12	(PB - 0)			
mulherzinha	9				
mulherzinhas	3				
			PASSARINHO	87	(PE - 4)
			passarinho	64	
			passarinhos	23	
			PAULINHO	14	(PE - 0)
			PEQUENININHO	42	(PE - 5)
			pequeninha	16	
			pequeninhas	5	
			pequeninho	19	
			pequeninhos	2	
			PERTINHO	10	(PE - 7)
PONTINHA	10	(PB - 2)			

POUCOCHINHO	10	(PB - 0)		
poucochino	9			
poucochinhas	1			
			POUQUINHO	127 (PE - 5)
			pouquinho	126
			pouquinhos	1
			QUADRINHO	17 (PE - 0)
			quadrinho	14
			quadrinhos	3
VELHINHO	10		VELHINHO	18
velhinha	2		velhinhas	2
velhinhas	2		velhinho	16
velhinho	5			
velhinhas	1			

Quanto aos diminutivos em *-ito*, o Quadro 6B exemplifica a sua abundância em PE e a sua total ausência em PB.

Quadro 6B

Excertos da lista comparativa de diminutivos em *-ito* extraída de *corpora* orais do português europeu e do português do Brasil

PE	Freq.	PB	Freq.
	54		0
bacoritos	1		
barquitos	1		
bercitos	1		
bocaditos	2		
botanitos	1		
calçõezitos	1		
casitas	2		
coiritos	1		
coisitas	4		
coitaditas	1		
copitos	1		

criaditas	2		
criticazitas	1		
curvitas	1		
descontozitos	1		
dinheirozitos	1		
escuditos	1		
estimaditas	1		
festitas	1		
fraquitos	1		
grupitos	1		
horitas	1		
hortaliçazitas	1		
massitas	1		
miuditas	3		
mulherzitas	2		
novitas	1		
novitos	1		
pauzitos	3		
pequenitas	1		
pequenitos	4		
problemazitos	2		
rapariguitas	1		
subsídiozitos	1		
tempitos	1		
tinturazitas	1		
veiozitos	1		
verbenazitas	1		

Os *corpora* permitem observar também aspectos sintáctico-semânticos de uso em cada uma das variedades do português que ficam patentes em alinhamento de concordâncias.

A título de exemplo, apresentamos, alinhadas, concordâncias extraídas do *corpus* PE e do *corpus* PB que exibem resultados contrastantes.

O verbo *acabar* como auxiliar aspectual e modal apresenta semelhanças nos usos de *acabar de + infinitivo*; as diferenças registaram-se nos usos de *acabar por + infinitivo* em PE aos quais corresponde *acabar + gerúndio* em PB conforme se pode ver no Quadro 7 em que damos exemplos de concordâncias destas duas estruturas.

Quadro 7
Concordâncias de ACABAR POR + INFINITIVO no *Corpus* Oral do Português Europeu

PF:1248

as pessoas lá tinham mais dinheiro, **acabavam por**, muitas vezes, **dar** boas gorjetas, e essas gorjetas **acabavam por compensar** bastante o trabalho

PF: 699

se não aproveitar um filme melhor que venha cá, **acaba por ficar** estúpida.

PF: 336

depois começámos a falar de (lisboa) e ele **acabou por me contar** toda a vida dele.

PF: 505

eles casam muito entre si e eles **acabam por ser** todos do mesmo ramo familiar

PF: 540

e foi assim que eles se conheceram. e **acabaram por casar**.

Concordâncias de ACABAR + GERÚNDIO no *Corpus* Oral do Português do Brasil

NPA2 : PROJETO NURC - PORTO ALEGRE

... então o pessoal **acaba adquirindo** gosto, **acaba gostando** da comida apimentada ...

NSP3 : PROJETO NURC - SÃO PAULO

... o cinema::...ahn já **acaba sendo** uma forma de diversão extremamente rara na nossa vida...

RIO18 : PROJETO NURC - RIO

... **acaba** a criança **ficando** sem vergonha....

SEEG1 - PUC - SP

... todo o mês a gente **acaba fazendo** um estudo diferente.

SEEG1 - PUC - SP

... ele **acaba tendo** um postura muito parecida ...

RIO31 : PROJETO PEUL - RIO

... e o ladrão **acabou devolvendo** o papagaio.

Nesta comunicação quisemos, tão só, dar a conhecer o início de uma investigação conjunta baseada em recursos linguísticos suficientemente

amplos e fiáveis para a realização de estudos contrastivos, cujo particular valor lhes será conferido pela autenticidade dos usos observados.

Como se pode ver a partir dos dados apresentados, as duas variedades do português, PE e PB exibem identidades e diferenças que exigem uma análise mais profunda dos *corpora*. Podemos anunciar que, entre outros estudos, pretendemos realizar, de forma sistemática, o confronto entre combinatórias lexicais em *corpora* orais e escritos cujas dimensões se situam nos 12.000.000 de palavras.

Dos resultados obtidos destes confrontos sistemáticos poderemos, com objectividade, avaliar da unidade e diversidade lexical das variantes europeia e brasileira do português (vocabulário fundamental, realidades extra-linguísticas que determinam variações, associações lexicais, estruturas léxico-sintácticas, etc.).

É um programa de trabalho ambicioso que nos propomos realizar e no qual, mais ambiciosamente ainda, pretendemos vir a incluir variantes africanas da língua portuguesa.

Notas

1 **Fornecedores de dados para o CRPC:** Academia das Ciências de Lisboa; Agência Lusa; Assembleia da República; Caixa Geral de Depósitos; Centro de Informática do Ministério da Justiça; Coimbra Editora; DECO; Editora Colibri; Editora Nova Fronteira - Brasil; Editorial Verbo; Estação de Rádio TSF; Fundação Calouste Gulbenkian - Serviço de Bibliotecas e Apoio à Leitura; Instituto de Apoio à Criança; Instituto do Consumidor; Jornais portugueses: Expresso, O Público, Diário de Notícias, Diário Económico, Jornal de Notícias, A Bola, A Capital, O Independente, Diário do Minho; Jornais de Cabo Verde: Correio Quinze, Novo Jornal e A Semana; Procuradoria-Geral da República; Corpus do Português Contemporâneo do Brasil (Universidade Estadual Paulista - UNESP); Projecto NURC-BR (São Paulo e Rio de Janeiro); Projecto PEUL (Rio de Janeiro); Revistas: Grande Reportagem, Ingenium, ProTeste, Máxima, Visão; Selecções do Reader's Digest; Sociedade Bíblica Portugal.

Bibliografia

- BACELAR DO NASCIMENTO, M. F. (1998), "Quelques ressources linguistiques de base pour l'étude des variétés du portugais contemporain", *Revue Française de Linguistique Appliquée*, vol. III-1, *Aspects de la diversité linguistique*, pg. 81-86.
- BACELAR DO NASCIMENTO, M. F. (1998), "Resultados do Projecto 'Dicionário de Combinatórias do Português' ", Revista *ALFA*, Universidade de S. Paulo (no prelo).
- BIDERMAN, M. T. C. (1997), Relatório CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) relativo à pesquisa *Dicionário de Frequências do Português Brasileiro Contemporâneo*.

UNIDADE E DIVERSIDADE: O VOCABULÁRIO DA LÍNGUA ORAL EM PORTUGAL E NO BRASIL

BOLINGER, Dwight (1972), *Degree Words*, Mouton, The Hague.

HABER, B. et alii (1997), *Les Linguistiques de Corpus*, Paris, Armand Colin, pg.183 e seg.