

CARACTERÍSTICAS DISTRIBUTIVAS DO ESPANHOL DE CUBA A PARTIR DA ANÁLISE DE DOIS ESTUDOS LINGÜÍSTICOS¹

Lucía Marconi², Leonel Ruiz Miyares³, Paola Cutugno⁴

Tradução: Cyrano da Rosa Silva⁵

Supervisão e revisão: Cleci Regina Bevilacqua⁶

1. Introdução

Desde 1998, existe uma frutífera colaboração acadêmica entre o grupo de pesquisas do Instituto di Linguistica Computazionale “Antonio Zampolli” (ILC, Gênova), pertencente ao *Consiglio Nazionale delle Ricerche* (Itália), e o Centro de Linguística Aplicada (CLA), do Ministério de Ciência, Tecnologia e Meio Ambiente, em Santiago de Cuba. Essa colaboração, em particular com a Dra. Eloína Miyares Bermúdez, com o Dr. Julio Vitelio Ruiz Hernández, com o Dr. Leonel Ruiz Miyares e com outros pesquisadores do CLA, ocorreu por meio da construção conjunta de duas obras editoriais: o *Diccionario Ortográfico de Español* (DOE) (MARCONI, MIYARES BERMÚDEZ, RATTI, ROLANDO e RUIZ MIYARES, 2000a) e o *Vocabulário Inverso e Anagramas do Espanhol* (VIAE) (MARCONI, MIYARES BERMÚDEZ, RATTI, ROLANDO e RUIZ MIYARES, 2000b).

As entradas do DOE e do VIAE foram obtidas a partir dos dados da obra científica *Léxico Activo-Funcional del Escolar Cubano* (LAFEC) (MIYARES BERMÚDEZ, 2006) realizada por pesquisadores do CLA, de Santiago de Cuba, sob a direção de Eloína Miyares.

Conforme destacam Ruiz Hernández e Miyares Bermúdez (1997, p. 67):

Os pesquisadores do nosso Centro, desde a sua fundação, há exatamente vinte e cinco anos, se dedicaram a realizar pesquisas

¹ A presente tradução foi autorizada para ser publicada em português por Leonel Ruiz. Referência bibliográfica completa do artigo original: MARCONI, L. RUIZ MIYARES, L. CUTUGNO, P. Características distributivas del español de Cuba a partir del análisis de dos estudios lingüísticos. In: MIYARES, L. R. (ed.) *Estudios de Lexicología y Lexicografía. Homenaje a Eloína Miyares Bermúdez*. Santiago de Cuba, Ediciones Centro de Linguística Aplicada, 2018, p. 87-110.

² Instituto di Linguistica Computazionale “Antonio Zampolli”, Consiglio Nazionale della Ricerche Gênova, Italia.

³ Centro de Linguística Aplicada, Ministério de Ciência, Tecnologia e Meio Ambiente, Santiago de Cuba, Cuba.

⁴ Instituto di Linguistica Computazionale “Antonio Zampolli”, Consiglio Nazionale Dell e Ricerche Gênova, Italia.

⁵ Aluno do Curso de Bacharelado em Letras – Tradução Português/Espanhol, UFRGS.

⁶ Professora do Departamento de Línguas Modernas e do Programa de Pós-Graduação em Letras, Instituto de Letras, UFRGS.

aplicadas com a finalidade de proporcionar aos professores do ensino básico, aos professores universitários, aos metodólogos e a outros trabalhadores da educação básica, bem como aos meios de difusão, especialmente o rádio e a televisão, alguns resultados científicos que permitissem o conhecimento das características fundamentais do espanhol de Cuba, especialmente no plano fonético e no nível léxico da língua. Para isso, realizamos algumas pesquisas de caráter descritivo da pronúncia de nossos falantes e, posteriormente, do vocabulário ativo dos estudantes do país, sendo concluídas em 1975 e 1995, respectivamente. Em ambas as vertentes de pesquisa, depois de coletar os resultados referidos, tivemos a tarefa de realizar trabalhos de aplicação e generalização dos dois níveis mencionados anteriormente – plano fonético e o nível léxico – incluindo o importante aspecto ortográfico.

O LAFEC é um dos resultados das pesquisas do CLA que permitiu o conhecimento de algumas características fundamentais do espanhol de Cuba. Essa obra foi feita através do processamento de textos de estudantes de ambos os sexos, da 2ª a 6ª série, dos setores urbanos e rurais em todo o país. Na pesquisa, foram coletadas 5.800 composições escritas e 1.200 amostras gravadas, e, ao final, os dados coletados constituíram um *corpus* de mais de 700.000 palavras, das quais mais de 25.000 são diferentes. Essas composições foram escritas a partir de diversas temáticas que cobriam esferas como a natureza, família, escola, entretenimento, entre outras, que permitiram uma aproximação à realidade do vocabulário das crianças da escola primária cubana e, além disso, a identificação das principais dificuldades de ortografia apresentadas por elas.

Com as palavras diferentes do LAFEC e suas diversas categorias gramaticais, foi iniciado o trabalho para a elaboração do DOE, formado por mais de 8.400 entradas de vários tipos: verbos, substantivos, adjetivos, entre outras categorias.

Na parte inicial do livro, são reproduzidas as atuais regras ortográficas, as quais devem ser seguidas para se escrever corretamente. Em seguida, constam algumas informações sobre as irregularidades gráficas de alguns verbos e outras palavras. Logo após, há uma pequena descrição do dicionário, seguida pelos modelos de conjugação verbal e, finalmente, o dicionário com suas entradas. Cada entrada contém informações relativas à categoria gramatical, à flexão da palavra para substantivos e adjetivos e, para os verbos, indica-se o número que remete ao modelo de conjugação correspondente. Com os resultados do LAFEC, elaborou-se também, como já mencionado, o *Vocabulário Inverso y Anagramas del Español* (VIAE).

O DOE e o VIAE se complementam a partir dos elementos de flexão das categorias como substantivos, adjetivos, etc., incluídos no primeiro, e mais de 80.000 palavras flexionadas em ordem reversa e mais de 10.000 anagramas que constam no segundo. Essas

duas obras são, atualmente, instrumentos muito úteis de trabalho e consulta para estudantes e professores em todo o país.

No VII Simpósio Internacional Comunicação Social, realizado em Santiago de Cuba, em 2001, foi apresentado o estudo *Rimas y anagramas en el Léxico del Escolar, Cubano* (MARCONI, MIYARES BERMUDEZ, RATTI, ROLANDO, 2001), em colaboração com o CLA, sobre as terminações dos lemas e das palavras flexionadas e sobre os anagramas dos espanhol de Cuba.

Em 1995, para o italiano, foi realizado o trabalho *Radiografia dell'italiano* (RATTI, MARCONI, BURANI, 1995), baseado em dados de dicionário *Zingarelli minore* (ZINGARELLI, ZANICHELLI, 1987) e nas flexões geradas a partir das entradas do dicionário *Flessioni-Rime-Anagrammi* (RATTI, MARCONI, MORGAVI, ROLANDO, 1988). Nesse estudo, foram apresentados alguns dados sobre a distribuição dos lemas e das palavras flexionadas.

Na introdução da obra *Radiografia dell'italiano* especifica-se:

Frequentemente, no decorrer de nosso trabalho, nos perguntamos sobre os componentes formais da linguagem, seja do tipo qualitativo ou quantitativo, e percebemos que, em muitas áreas da ciência, onde a linguagem é objeto ou sujeito, há um crescente interesse nas propriedades de distribuição de unidades linguísticas pertencentes a um idioma específico.

As "propriedades de distribuição" expressam um conjunto de medidas empíricas que levam em conta o número de palavras de um tipo particular ou de uma categoria gramatical específica, pertencentes a um dado idioma e as proporções relativas, a frequência de ocorrência em comparação com todo o conjunto, o número absoluto e a frequência de certas unidades sublexicais, tais como sequências de letras, sílabas e morfemas.

A importância atribuída às propriedades distributivas nasce do interesse de pesquisar quais os componentes, nos processos de acesso a representação mental do léxico e nos processos linguísticos em geral, são atribuídos às propriedades universais da linguagem e ao sistema cognitivo e quais podem ser considerados dependentes de linguagem específica a qual um leitor está exposto⁷.

Analisar as propriedades distributivas significa, portanto, realizar uma série de medidas empíricas de uma dada língua, considerando, por exemplo, o número de palavras de um tipo específico ou de uma categoria gramatical de um determinado conjunto, seja um dicionário ou um léxico de frequências. Isso significa calcular as proporções, a

frequência de ocorrência, relacionando-a ao conjunto e ao grupo específico, bem como calcular a frequência de certas unidades sublexicais, como sequências de letras.

O objetivo deste artigo é realizar um estudo sobre as características distributivas de um subconjunto do espanhol de Cuba. Referimo-nos à informação contida no DOE – excluindo as lexias complexas e as locuções –, ou seja, aos 7.927 lemas gerados a partir das 93.759 palavras flexionadas com suas categorias gramaticais. Os dados do LAFEC também serão analisados para mapeamento das palavras utilizadas pelos alunos. A partir do LAFEC, foram extraídas as frequências de certas unidades sublexicais e, neste caso, foi feita referência apenas aos termos usados pelas crianças em suas produções e, posteriormente, às formas utilizadas.

O artigo também contém uma série de comparações de algumas propriedades distributivas derivadas do tratamento dos dados do DOE com os resultados obtidos da análise do LAFEC.

Quando se fala de propriedades distributivas de uma língua, essas podem ser consideradas em relação aos lemas, às várias palavras flexionadas ou ao número de ocorrências⁸ em um *corpus* específico. Neste artigo, faremos referência a uma amostra de lemas e flexões derivadas do DOE e somente às flexões do LAFEC.

2. A informação linguística e a estrutura quantitativa do léxico

Decidiu-se realizar uma série de cálculos sobre o espanhol de Cuba com os dados do DOE, para os quais foram consideradas todas as formas possíveis ou palavras flexionadas geradas a partir dos lemas. Na Tabela 1, encontra-se a distribuição dos lemas e das formas das diferentes categorias gramaticais, incluindo as palavras homógrafas.

<i>categoria gramatical</i>	<i>lema</i>	<i>forma</i>
adjetivo	1831	6486
advérbio	265	265
artigo	4	7
conjunção	14	14
contração	2	2
interjeição	27	27
preposição	17	17
pronome	59	132
sustantivo	4460	9471

⁸ Ocorrência do lema ou palavra flexionada (forma ou flexão) em um *corpus*, ou seja, quantas vezes é encontrada a palavra como lema ou na sua forma flexionada.

verbo	1248	77336
Total	7927	93759

Tabela 1 – Distribuição dos lemas e das formas das diferentes categorias gramaticais a partir de dados do DOE.

Nesse estudo de distribuição de lemas e de palavras flexionadas, as palavras homógrafas foram analisadas uma só vez; portanto, de 7.927 lemas foram analisados 7.535 lemas diferentes e de 93.759 formas, 80.336 diferentes.

A tabela 2 contempla a informação relacionada com a quantidade de lemas e palavras flexionadas presentes no DOE, considerando a quantidade de letras (longitude) da palavra.

<i>longitud e</i>	<i>lemas</i>	<i>formas</i>	<i>longitude</i>	<i>lemas</i>	<i>formas</i>	<i>longitude</i>	<i>lemas</i>	<i>formas</i>
1	5	5	9	958	13505	17	3	41
2	34	48	10	630	12858	18	3	8
3	95	184	11	371	9781	19	0	1
4	431	1085	12	196	6423	20	0	0
5	884	3041	13	94	3617	21	1	2
6	1150	5619	14	61	1623	22	0	2
7	1301	9344	15	28	669			
8	1279	12281	16	11	199			

Tabela 2 – Quantidade de lemas e palavras flexionadas presentes no DOE, considerando a quantidade de letras (longitude) da palavra.

Como é possível observar, tanto para os lemas quanto para as palavras flexionadas, há picos, ou seja, a presença de um maior número de palavras para uma determinada longitude. Para os lemas, o pico é atingido com as entradas de 7 letras, enquanto que para as formas, com palavras que possuem 9 letras.

Se representarmos os dados da Tabela 2 na Figura 1, podemos ver que ambas as curvas têm um comportamento tipo Gaussiano: existem valores muito pequenos nas fases inicial e final; tais valores aumentam para a longitude máxima de 7 letras para os lemas e de 9 para as formas e, a partir dessas longitudes, ambas as curvas diminuem novamente.

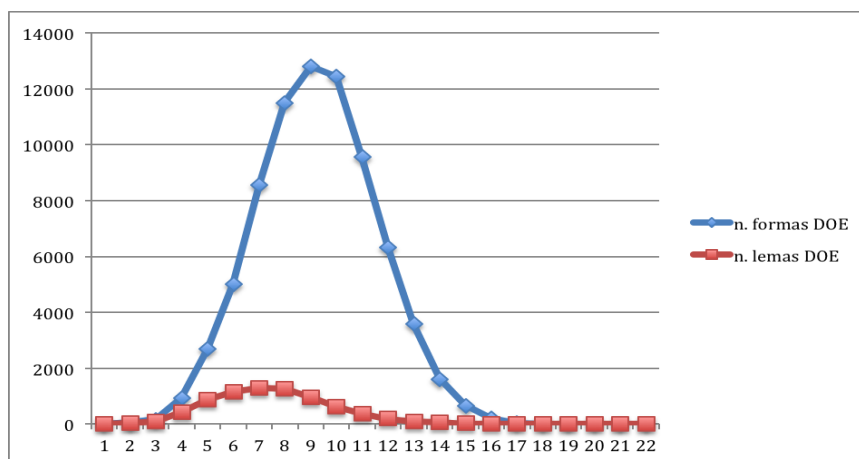


Figura 1– Variação dos lemas e das flexões segundo o número de letras.

A Tabela 3 fornece informações sobre a frequência das letras do alfabeto nos lemas/formas em qualquer posição da palavra. As tabelas 4 e 5 trazem informações relacionadas à distribuição das letras do alfabeto na posição inicial e final da palavra, respectivamente.

<i>letras</i>	<i>lemas</i>	<i>formas</i>	<i>letras</i>	<i>lemas</i>	<i>formas</i>
a/á	7929	116995	Ñ	133	1336
b	988	14262	o/ó	5842	51215
c	3132	31542	P	1585	17648
d	2433	27662	Q	196	1891
e/é	5553	85370	R	5575	73066
f	613	7026	S	2289	73813
g	882	9052	T	3243	31091
h	472	4297	u/ú/ü	1751	17940
i/í	4697	62492	V	645	7238
j	341	3422	W	1	1
k	19	33	X	101	1470
l	2654	22299	Y	116	1211
m	1852	26678	Z	381	4182
n	3470	43036			

Tabela 3– Frequência das letras do alfabeto nos lemas/formas em qualquer posição da palavra.

<i>Posição inicial</i>			
<i>letras</i>	<i>formas</i>	<i>letras</i>	<i>formas</i>
a/á	11320	n	769
b	2120	ñ	5
c	9593	o/ó	1825
d	7651	p	6323
e/é	9035	q	328
f	2516	r	6819
g	1421	s	4331
h	1159	t	3436
i/í	3614	u/ú/ü	395
j	412	v	1710
k	12	x	0
l	1948	y	43
m	3328	z	223

Tabela 4 – Distribuição das letras do alfabeto em posição inicial da palavra.

<i>Posição final</i>			
<i>letras</i>	<i>formas</i>	<i>letras</i>	<i>formas</i>
a/á	10028	ñ	0
b	1	o/ó	7000
c	0	p	1
d	1354	q	0
e/é	7912	r	1502
f	2	s	40186
g	5	t	12
h	4	u/ú/ü	15
i/í	284	v	1
j	1	w	1
k	2	x	0
l	248	y	23
m	2	z	49
n	11703	z	49

Tabela 5– Distribuição das letras do alfabeto em posição final da palavra.

A partir da distribuição das letras do alfabeto na posição final da palavra (Tabela 5), foram selecionadas aquelas com frequência maior que 1.000 (a, d, e, n, o, r, s) e analisadas as sequências constituídas por duas letras. Na Tabela 6, pode-se observar a produtividade de algumas sequências de duas letras.

<i>a</i>	<i>forma</i>	<i>e</i>	<i>forma</i>	<i>o</i>	<i>forma</i>	<i>d</i>	<i>forma</i>	<i>s</i>	<i>forma</i>
ba	1006	ae	8	ao	6	ad	1051	as	10010
ca	286	be	58	bo	65	ed	153	es	5317
da	1601	ce	187	co	335	Id	139	is	11234
ea	75	de	167	do	2885	ud	11	ls	1
fa	11	ee	104	eo	136			os	13601
ga	125	Fe	14	fo	12	<i>n</i>	<i>forma</i>	ps	1
ha	55	ge	17	go	167	an	6253	rs	7
Ia	1845	he	62	ho	78	en	3772	ss	1
ja	85	Ie	95	io	488	in	28	ts	6
ka	1	Je	110	jo	117	on	1641	us	8
la	257	ke	2	lo	286	un	9		
ma	137	Le	226	mo	177				
ña	49	me	94	ño	57	<i>r</i>	<i>forma</i>		
na	309	ñe	37	no	319	ar	991		
oa	2	ne	203	oo	1	er	176		
pa	48	oe	1	po	59	ir	143		
ra	3029	pe	50	ro	475	or	189		
sa	233	Re	2787	so	263	ur	3		
ta	605	se	1377	to	713				
ua	29	Te	1979	uo	24				
va	80	üe	3	vo	106				
ya	31	ue	246	yo	49				
za	128	ve	64	zo	179				
		ye	20						

Tabela 6 – Frequência da sequência de duas letras em palavras flexionadas.

A Tabela 7 mostra as sequências de terminações de 3 letras com uma frequência superior a 100, escolhidas com base na tabela anterior (Tabela 6) com uma frequência maior que 100.

Analizamos as sequências de duas letras, com as informações da Tabela 4, escolhidas dentre aquelas que tiveram frequência superior a 1000, ou seja, igual a 97% do número de formas do DOE. A Tabela 8 mostra a produtividade de alguns grupos de duas sequências de letras iniciais.

<i>A</i>	<i>forma</i>	<i>D</i>	<i>forma</i>	<i>N</i>	<i>forma</i>	<i>o</i>	<i>forma</i>	<i>s</i>	<i>forma</i>	<i>s</i>	<i>forma</i>
Ara	1937	Tad	172	Ran	2627	ndo	1317	mos	10220	nas	310
Ria	1383	dad	143	Ian	1583	ado	1137	ais	4983	cas	284
ada	1145	rad	112	Ren	1378	ido	321	eis	3745	las	253
aba	961			Sen	1291	ero	178	ras	3027	ros	251
era	590	<i>e</i>	<i>forma</i>	Ron	1268	ito	169	éis	2348	sas	231
ida	307	are	1928	Ban	985	nto	167	ias	1833	des	203
Ira	149	ste	1311	Ión	249	ico	164	res	1670	ios	193
ora	144	ase	980	Ten	210	rio	120	das	1602	sos	179
ica	140	ere	452	Tan	208	oso	117	dos	1502	cos	171

cia	136	nte	375	Uen	117	ino	108	ses	1319	nos	164
Ita	133	ese	323	Cen	115	cio	103	bas	1005	ces	162
sta	130	ire	159	Can	109			tas	602	los	144
osa	122	que	146	Nen	108	<i>r</i>	<i>forma</i>	nes	548	mas	141
lla	108					tar	168	tes	411	ues	139
ina	104					dor	112	tos	390	zas	129
						rar	112	les	385	gas	127

Tabela 7 – Frequência de seqüências de três letras em palavras flexionadas.

<i>a/á</i>	<i>forma</i>	<i>c</i>	<i>forma</i>	<i>f</i>	<i>forma</i>	<i>l</i>	<i>forma</i>	<i>p</i>	<i>forma</i>
ab	849	ca	1890	fa	477	la	604	pa	1108
ac	1555	ce	438	fe	222	le	244	pe	1267
ad	790	ch	630	fi	399	li	488	pi	636
ae	12	ci	210	fl	155	ll	372	pl	359
af	412	cl	106	fo	356	lo	103	po	562
ag	773	co	5123	fr	409	lu	137	pr	2127
ah	167	cr	490	fu	498			pu	264
ai	10	cu	706			<i>m</i>	<i>forma</i>		
aj	71			<i>g</i>	<i>forma</i>	ma	1437	<i>r</i>	<i>forma</i>
al	1055	<i>d</i>	<i>forma</i>	ga	318	me	705	ra	468
am	537	da	150	ge	41	mi	253	re	5473
an	666	de	5353	gi	75	mo	662	ri	169
añ	2	di	1545	gl	16	mu	271	ro	602
ap	1240	do	416	go	211			ru	107
aq	10	dr	68	gr	334	<i>o/ó</i>	<i>forma</i>		
ar	939	du	119	gu	426	ob	302	<i>s</i>	<i>forma</i>
as	942					oc	229	sa	983
at	716	<i>e/é</i>	<i>forma</i>	<i>h</i>	<i>forma</i>	od	55	se	853
au	164	eb	2	ha	446	oe	1	sh	6
av	147	ec	67	he	139	of	174	si	421
ay	60	ed	70	hi	172	oh	1	so	1148
az	202	ef	69	ho	206	oi	33	sp	1
		eg	2	hu	196	oj	3	st	1
<i>b</i>	<i>forma</i>	eh	1			ol	176	su	918
ba	746	ej	175	<i>i/í</i>	<i>forma</i>	om	1		
be	261	el	318	ib	5	op	238	<i>t</i>	<i>forma</i>
bi	67	em	1253	ic	6	or	339	ta	474
bl	125	en	3318	id	123	os	67	te	479
bo	396	ep	6	if	1	ot	58	ti	244
br	347	eq	117	ig	19	ou	1	to	516
bu	178	er	171	il	66	ov	12	tr	1583
		es	2011	im	643	ox	1	tu	140
		et	7	in	2672	oy	21		
		eu	6	ir	22	oz	1	<i>v</i>	<i>forma</i>
		ev	224	is	3			va	367
		ex	1216	it	2			ve	439

		ey	1	iz	52			vi	583
								vo	297
								vu	24

Tabela 8 – Frequência da combinação de duas letras iniciais em palavras flexionadas ou formas.

A partir de algumas sequências de duas letras iniciais referentes às formas da Tabela 8, selecionadas entre aquelas com frequência superior a 1.000, foram gerados grupos de três letras no início da palavra. Na Tabela 9, foram incluídos aqueles grupos com frequências superiores a 100.

Nos grupos obtidos a partir da sequência de três letras destacam-se:

- 6 grupos com uma frequência superior a 300 (ver tabela 9): *arr; imp; sal; agr; pla; ama;*
- 39 grupos com uma frequência entre 150 e 300: *ase; for; tor; atr; ven; gua; fun; chi; abo; que; cha; cri; gra; bri; mer; lib; asi; cur; lim; sus; abu; eje; cre; vac; afi; vol; fal; ado; ade; lle; eva; aho; hal; ocu; sac; ata; sen; fre; hab;*
- 72 grupos com uma frequência entre 100 e 150: *cer; sup; bal; pin; san; men; fra; pic; ate; bat; flo; tem; dom; val; agu; cul; jug; pos; cho; gui; uni; bas; fin; sem; adm; bor; ser; llo; ori; ani; lab; qui; cru; aso; pis; ras; ano; ele; equ; aba; lam; ord; sel; mol; ide; tal; adi; agi; tap; abr; obs; sum; azo; ond; sur; afe; ofr; aga; bar; fum; luc; abs; mor; asa; mon; mov; rod; sab; err; pol; her; ara; ten.*

ac/ác	forma	de/dé	forma	ex	forma	re	forma
aca	178	deb	114	Exh	108	rea	117
ace	217	dec	356	Exi	119	reb	163
acl	110	def	178	Exp	529	rec	1091
aco	539	dem	122	Ext	314	red	232
acr	110	den	114			ref	451
acu	193	dep	177	in	forma	reg	282
		der	279	Inc	439	rei	136
al/ál	forma	des	3508	Ind	108	rel	302
alb	119	det	125	Inf	261	rem	125
ale	125	dev	165	Ins	377	ren	202
ali	242			Int	737	rep	473
alm	130	di/dí	forma	Inv	348	res	877
		dif	117			ret	388
ap/áp	forma	dis	777	ma	forma	rev	405
apa	233	div	286	mal	109		
apl	168			man	458	so	forma
apo	270	em/ém	forma	mar	344	sob	402
apr	442	emb	296	mat	145	sol	143

apu	116	emp	780			son	117
				pa	forma	sop	112
ca/cá	forma	en/én	forma	Par	387	sos	114
cal	397	enc	775	Pas	302		
car	247	end	108			tr	forma
cas	195	enf	336	pe	forma	tra	1131
cam	178	eng	270	Pel	157	tri	251
can	177	eno	111	Pen	135	tro	130
cap	140	enr	179	Per	627		
cau	110	ens	283	Pes	134		
		ent	731				
co/có	forma	env	223	pr	forma	Outros grupos	
coa	108			Pre	749	Frec> 300	
col	495	es/és	forma	Pro	1251	arr	631
com	1174	esc	791			imp	525
con	2353	esp	264			sal	370
coo	112	est	809			agr	363
cor	407					pla	333
cos	179					ama	301

Tabela 9 – Frequência de combinações de três letras iniciais em palavras flexionadas.

3. Comparação entre o *Léxico Ativo-Funcional del Escolar Cubano* e o *Diccionario Ortográfico del Español*

Ao analisarmos as palavras do DOE, observamos que estas não incluem algumas formas presentes no LAFEC. Existem 1.804 formas do LAFEC que não foram incluídas entre os lemas do dicionário; a maior parte consiste em formas verbais com enclíticos (terminado em: *el, los, la, las, te, nos, se, me*), alguns tipos e números de substantivos e alguns diminutivos de adjetivos (terminados em: *-ito, -itos, -ita, -itas*).

A entrada do DOE é formada pelo lexema, seguido pelas siglas que indicam seu gênero e número e categoria gramatical. No caso dos verbos, eles sempre aparecem no infinitivo como é lógico, seguidos pela categoria gramatical do verbo e um número escrito entre colchetes que se refere ao seu modelo de conjugação. Entre parênteses, é apresentada a divisão em segmentos para indicar onde se encontra a mudança em relação à norma, ou seja, em relação aos verbos regulares.

Exemplos:

perr -o, -os, -a, -as; sust. m. y f.
volver [38]; v. (v -o -lv -er)

Como pode ser visto na Tabela 10, o número de formas do DOE (80.336) é muito maior do que as formas do LAFEC (17.967); no entanto, as entradas no DOE totalizaram apenas 7.535.

Além disso, na tabela 10, mostra-se a distribuição do número de palavras flexionadas diferentemente, tanto para o DOE quanto para o LAFEC, ao variar a longitude (número de letras); o número de lemas para o DOE também é apresentado na mesma tabela quando a longitude varia.

Nas três últimas colunas da Tabela 10, foram acrescentados alguns dados mais significativos que mostram o número de formas diferentes no LAFEC, tanto em relação à longitude da palavra quanto à frequência no léxico, ou seja, sua ocorrência na produção dos estudantes. Para sermos mais precisos, as informações, nas três últimas colunas, referem-se ao léxico com frequência ≥ 3 , ≥ 5 e ≥ 10 entre palavras de três a catorze letras.

A figura 2 indica o número de lemas e palavras flexionadas referentes ao DOE e o número de palavras flexionadas para o LAFEC, de acordo com a longitude das palavras.

Long.	Formas DOE	Lemas DOE	Formas LAFEC	Formas LAFEC (f. ≥ 3)	Formas LAFEC (f. ≥ 5)	Formas LAFEC (f. ≥ 10)
1	5	5	5			
2	48	34	45			
3	184	95	160	130	117	102
4	1085	431	789	564	468	360
5	3041	884	1732	1135	922	647
6	5619	1150	2418	1335	983	641
7	9344	1301	3125	1482	1060	654
8	12281	1279	3009	1306	897	512
9	13505	958	2518	935	617	309
10	12858	630	1941	658	404	201
11	9781	371	1119	348	220	101
12	6423	196	600	167	96	47
13	3617	94	275	72	40	15
14	1623	61	129	35	20	8
15	669	28	63			
16	199	11	22			
17	41	3	8			
18	8	3	5			
19	1	0	1			
20	0	0	0			
21	2	1	1			
22	2	0	2			

Tabela 10 – Número de lemas/formas no DOE e número de formas no LAFEC para diferentes longitudes.

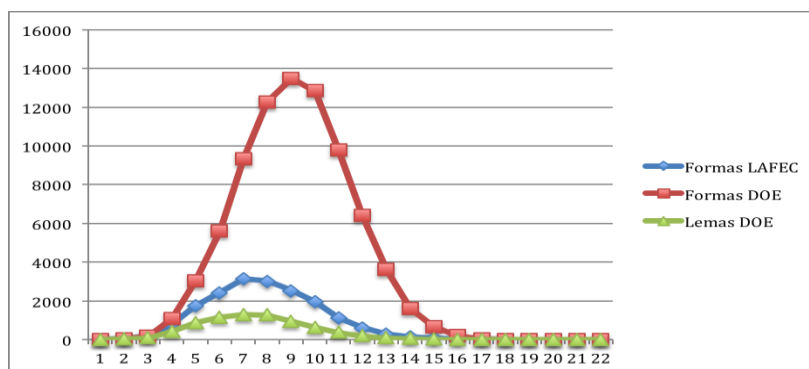


Figura 2 – Número de lemas e palavras flexionadas (DOE) e número de palavras flexionadas (LAFEC), de acordo com o número de letras das palavras.

A figura 2 mostra que o número de formas diferentes do LAFEC e do DOE é variável. No caso do LAFEC, 99% das palavras contêm entre três e catorze letras e, no DOE, o índice é de 98% nessa faixa de longitude.

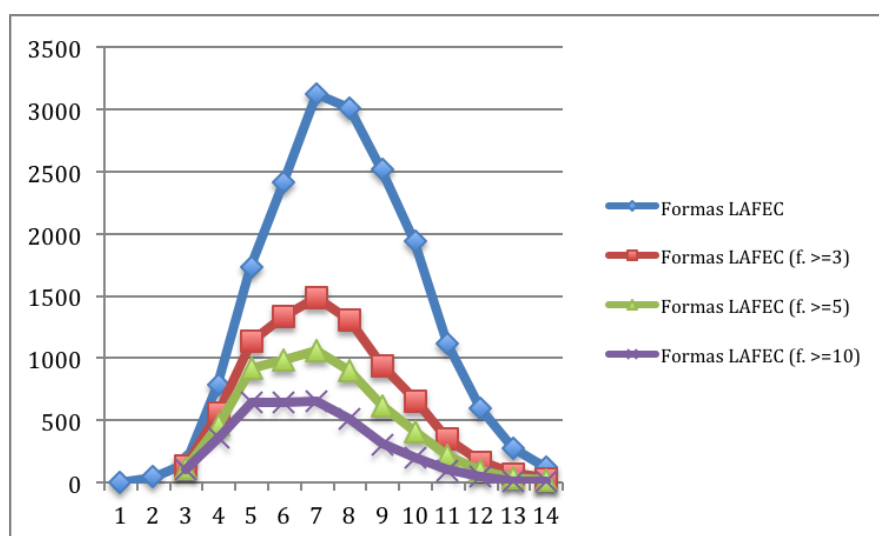


Figura 3 – Número de palavras flexionadas de acordo com o número de letras das palavras e frequências (LAFEC).

A Figura 3 mostra o número de palavras flexionadas no LAFEC, considerando a longitude das palavras entre três e catorze letras, para frequências ≥ 3 , ≥ 5 , ≥ 10 . Ao analisar a Figura 3, fica evidente que, para o LAFEC, as palavras mais utilizadas pelas crianças são aquelas com 7 letras, enquanto que para as palavras geradas com DOE (Fig. 2), o pico ocorre nas palavras de 9 letras. Não é difícil entender que a obtenção de palavras do DOE é mais produtiva do que a dos alunos do ensino fundamental.

Considerando a decisão de selecionar formas com a longitude variável entre 3 e 14 caracteres pertencentes ao LAFEC, um subconjunto de palavras com uma extensão entre 3 e 14 letras (LAFEC-P) foi extraído e, em seguida, um subconjunto de palavras com uma frequência ≥ 3 (LAFEC-PF3).

Há quatro conjuntos de dados que usaremos nesta seção dedicada à comparação entre o LAFEC e DOE: as formas geradas com o DOE, a informação do LAFEC e seus subconjuntos (LAFEC-P e LAFEC-PF3). Com esses grupos, combinações possíveis de 2, 3 e 4 letras foram geradas. Esses resultados foram comparados entre si para tentar identificar as sequências relevantes de 2, 3 e 4 letras a partir do menor subconjunto (LAFEC-PF3), a fim de verificar se essas sequências eram relevantes também para as do DOE, ou seja, para o maior conjunto.

A Figura 4 mostra os valores das diferentes sequências de duas letras nos diferentes conjuntos e para as sequências com frequência $\Rightarrow 3$, isto é, para sequências que aparecem em três ou mais palavras diferentes. Como é possível observar, o número de diferentes sequências nos conjuntos varia pouco, oscila entre 344 sequências para a DOE, 355 para o LAFEC e 329 para o LAFEC-P, chegando a 291 para o LAFEC-PF3, o menor conjunto.

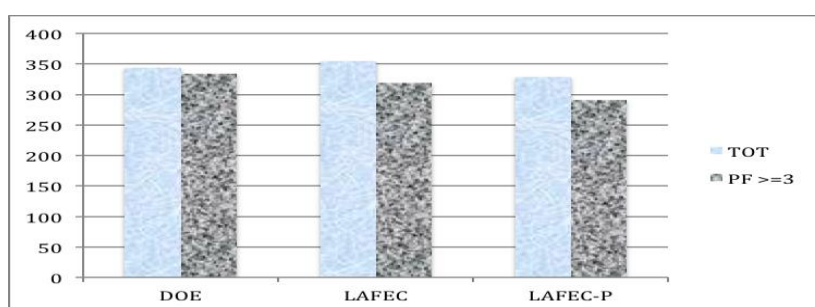


Figura 4– Valores das diferentes sequências de duas letras nos diversos conjuntos.

Quando as combinações de duas letras do DOE e do LAFEC-PF3 são comparadas, obtém-se 344 combinações para o DOE, entre as quais 21 não aparecem no conjunto do LAFEC-PF3. Essas combinações são: *bn*: 57, *dj*: 2, *dv*: 54, *dy*: 54, *ih*: 2, *ik*: 8, *ix*: 54, *iy*: 2, *jt*: 2, *lh*: 2, *nh*: 58, *nn*: 2, *ow*: 1, *pc*: 8, *ps*: 3, *ry*: 1, *sk*: 2, *uq*: 12, *xh*: 108, *xq*: 4, *zn*: 2⁹.

Das 21 combinações de duas letras, analisaremos as seis que possuem frequência superior a 50 em diferentes palavras: a sequência *xh* que aparece em 108 palavras diferentes, refere-se às conjugações dos verbos *exhibir* e *exhortar*, cujo uso no LAFEC teve uma frequência muito baixa. O mesmo acontece com a sequência *nh* que se refere às formas do verbo *anhelar* (ansiar¹⁰) e às formas do adjetivo *inhumano* (desumano); a combinação *ix* refere-se às formas do verbo *asfixiar*; o *dy* ao verbo *coadyuvar* (auxiliar); *dv* refere-se ao verbo *advertir* a combinação *bn*, ao verbo *abnegar*.

Ao analisar o conjunto LAFEC-P, 329 combinações de duas letras foram obtidas. Apenas 5 combinações do LAFEC-P não estão presentes no DOE: *ek*, *kw*, *wo*, *rk*, *yi*. As combinações *ek*, *kw*, *wo*, *rk* aparecem em uma única palavra. A baixa frequência é lógica

⁹ Os números representam a frequência da combinação de duas letras em diferentes palavras.

¹⁰ N.T.: Foram indicados os equivalentes em português somente para as palavras cujo significado era menos transparente.

devido ao fato de serem estrangeirismos pouco utilizados. No caso de *ek*, *kw* e *wo*, as combinações referem-se à palavra *taekwondo* e, no caso de *rk*, *asnorkel*.

Para a combinação *yi*, a sequência foi encontrada em palavras diferentes, tal como nos diminutivos *cayito*, *rayito*, *hoyito*, *rayita*, *hoyitos*, *playita*, *rayitas*, *rayitos*, *arroyito*¹¹.

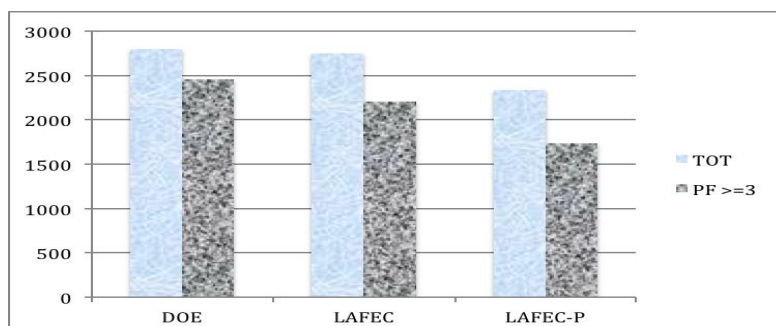


Figura 5 – Valores das diferentes sequências de três letras nos diferentes conjuntos.

Em relação à sequência de três letras, na Figura 5, pode-se observar que há mudanças significativas nos três conjuntos (DOE, LAFEC e LAFEC-P). Para o DOE existem 2.800 sequências diferentes, 2.747 para o LAFEC e 2.333 para o LAFEC-P. As diferentes sequências para o conjunto LAFEC-PF3 totalizaram 1.735 (sequências envolvendo três ou mais palavras diferentes).

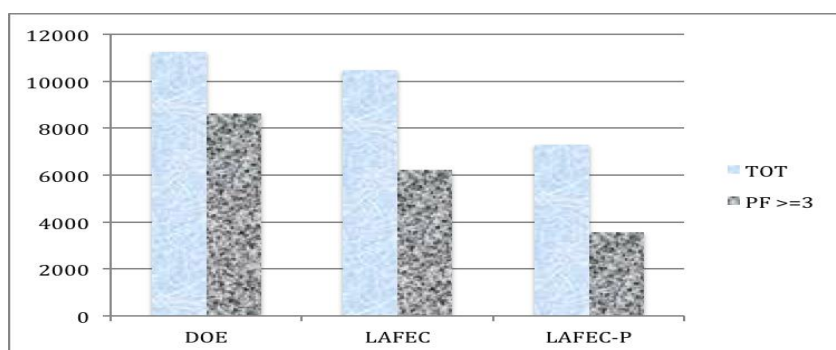


Figura 6 – Valores das diferentes sequências de quatro letras nos diversos conjuntos

A figura 6 mostra a sequência de quatro letras. As variações são maiores nos três conjuntos. Para o DOE existem 11.245 sequências diferentes, 10.478 para o LAFEC e 7.295 para o LAFEC-P. As diferentes sequências para o conjunto LAFEC-PF3 totalizaram 3.567 (sequências envolvendo três ou mais palavras diferentes).

¹¹ N.T.: ilhazinha, raiozinho, burquinho, linhazinha, buraquinhos, praiazinhas, linhazinhas, raiozinhos, riozinho.

Com base nos cálculos feitos, percebe-se que muitas sequências não são comuns aos três conjuntos; algumas pertencem apenas ao DOE, outras apenas ao LAFEC-P e outras ao LAFEC-PF3. Por exemplo, se consideramos um conjunto de 4 letras que pertencem apenas ao LAFEC-PF3, onde a frequência das sequências aparece em pelo menos dez palavras, obteremos 32 grupos (ver tabela 11).

Depreende-se da Tabela 11 que esses grupos de 4 letras estão presentes principalmente na parte interna das palavras; apenas quatro sequências encontram-se no final das palavras (*irse, selo, erlo, rlos*), três na parte interna e final (*ogía, selo, erlo*) e três no início e no meio da palavra (*tiri, irno, darl*).

<i>total</i>	<i>princípio</i>	<i>interior</i>	<i>final</i>	<i>sequência</i>
10	0	10	0	derl
10	0	10	0	eint
10	0	10	0	iarl
10	0	10	0	jarl
10	0	10	0	zarl
11	0	11	0	gerl
11	0	11	0	ueli
11	1	10	0	tiri
12	0	12	0	sarl
13	0	13	0	cerl
13	0	13	0	larl
13	0	13	0	varl
13	0	13	0	rarm
14	0	5	9	ogía
14	0	14	0	rarn
14	0	14	0	tars
15	0	15	0	nerl
16	0	16	0	narl
17	0	0	16	irse
18	0	8	10	selo
19	0	19	0	rars
20	0	20	0	tarn
21	1	20	0	irno
23	2	21	0	darl
26	0	26	0	ndom
27	0	27	0	carl
28	0	28	0	rsel
41	0	18	23	erlo
42	0	42	0	rarl
48	0	48	0	tarl
62	0	0	62	rlos
87	0	87	0	arno

Tabela 11 – Grupo de 4 letras pertencentes ao LAFEC-PF3 e com frequência de sequências igual ou superior a 10.

Nesse caso, por exemplo, a sequência de quatro letras **arno**, com frequência 87 em posição interna da palavra, corresponde às formas: *ayudarnos, buscarnos, darnos, saludarnos, etc.*, ou seja, todas as formas verbais com enclíticos. Inclusive, nos casos da combinação **rlos**, com frequência 62 no final da palavra, também estamos diante de formas verbais com enclíticos: *asarlos, curarlos, dejarlos, sacarlos, etc.*; no caso de **erlo**, refere-se a: *conocerlo, esconderlo, protegerlo, etc.*; no caso de **selo** a: *comérselo, entregárselo, llevárselo, etc.*

Para a combinação **ueli**, no interior da palavra e com frequência 11, temos a presença de diminutivos como *abuelito, abuelita, escuela, vuelitos*¹². Com frequência inferior a 10, destaca-se o uso de diminutivos com as seguintes sequências **brit**: *librito*, **ayit**: *rayito*, **ceci**: *lucecita*, **jari**: *pajarito*, **reri**: *sombrerito*, **añit**: *montañitas*, **yito**: *arroyito, hoyitos*¹³. Todos os casos acima estão em correspondência correta com o vocabulário dos estudantes de nível de ensino analisado. Exemplos de palavras que terminam em **ogía** são: *ideología, mitología, arqueología, ginecología, etc.*

No grupo das 32 sequências, há algumas, como **erlo, rlos, darl, arno, rarn, nerl**, que estão na mesma palavra, como por exemplo, *ponerlos*, constituída de duas sequências no interior da palavra (**nerl, erlo**) e uma no final (**rlos**).

Analisaremos, abaixo, um grupo de sequências de quatro letras pertencentes ao DOE. Trata-se de um pequeno subconjunto em que as sequências têm uma frequência igual ou superior a 100, ou seja, aparecem em pelo menos cem palavras. Os grupos pertencentes ao DOE, com frequência igual ou superior a 100, correspondem a um total de 1.145 sequências, das quais trinta apenas se encontram no DOE (ver Tab.12) e não nos demais conjuntos (LAFEC, LAFEC-P e LAFEC-PF3).

<i>total</i>	<i>príncipi o</i>	<i>Interior</i>	<i>final</i>	<i>sequência</i>
100	0	0	100	néis
102	0	102	0	east
105	0	0	105	céis
107	0	0	107	cáis
108	0	0	108	uéis
116	0	87	29	jase
118	0	118	0	zast
145	0	145	0	jarí
155	0	155	0	tirí
163	0	123	40	mase
168	0	126	42	iase
180	0	135	45	sase
196	0	147	49	dase

¹² N.T.: vovozinho, vovozinha, escolinha, voozinho.

¹³ N.T.: livrinho, raiozinho, luzinha passarinho, chapeuzinho, montanhazinhas, riozinho, buraquinho.

202	0	0	202	táis
204	0	153	51	ease
236	0	177	59	zase
294	0	294	0	esei
295	1	294	0	erai
295	0	295	0	erei
320	0	240	80	nase
952	0	952	0	asei
954	0	954	0	arai
954	2	952	0	arei
955	0	955	0	abai
956	0	0	956	bais
1246	0	0	1246	teis
1247	0	0	1247	rais
1247	0	0	1247	reis
1267	0	1267	0	ríaí
1537	0	0	1537	íais

Tabela 12 – Grupo de sequências de quatro letras com frequência igual ou superior a 100 (DOE).

Focalizaremos nas sequências de terminações *iais*, *reis*, *rais*, *teis*, *bais*, *táis*, *uis*, *cáis*, *céis*, *néis*. No caso de *íais*, presente em 1.537 palavras diferentes, a sequência refere-se à forma verbal do futuro do pretérito do indicativo da segunda pessoa do plural (vosotros/vosotras); para *reis*, as palavras correspondem à forma verbal do futuro do subjuntivo na segunda pessoa do plural; para *teis*, à forma verbal do pretérito perfeito do indicativo na segunda pessoa do plural; para *bais*, à forma verbal do pretérito imperfeito do indicativo na segunda pessoa do plural; para *táis*, *cáis*, *céis* às formas verbais do presente do indicativo na segunda pessoa do plural; para *uéis*, à forma verbal do presente do subjuntivo na segunda pessoa do plural e para *néis*, à forma verbal do presente do indicativo na segunda pessoa do plural.

Como se sabe, na América Hispânica, a segunda pessoa do plural *vosotros/vosotras* não é usada, razão pela qual, evidentemente, as formas verbais mencionadas anteriormente da segunda pessoa do plural não foram utilizadas nos escritos – e muito menos nas exposições orais – das crianças cubanas. A análise comparativa do DOE e do LAFEC produziu uma valiosa informação sobre o uso de palavras pelos estudantes. Os resultados obtidos relacionados às diferentes sequências são uma fonte rica de dados úteis para estudos de reconhecimento de fala e estudos de escrita de textos, bem como para pesquisas psicolinguísticas e para a preparação de exercícios didáticos dirigidos aos estudantes.

4. Os anagramas

Estudos sobre anagramas são frequentemente muito úteis na educação e reabilitação. A partir de algumas pesquisas neurolinguísticas¹⁴, sabe-se que as informações sobre as sequências de letras são muito importantes para o estudo dos problemas apresentados por sujeitos disgráficos e afásicos. Com a extensa base de dados do DOE foram gerados todos os códigos alfabéticos ou alfa-code (a-code¹⁵) tanto para os lemas quanto para as formas.

Destaca-se que alguns a-code representam apenas uma palavra, outros podem gerar mais de uma. Foram calculados todos os possíveis anagramas com seu a-code¹⁶ que podem gerar mais de uma solução na língua real, ou seja, palavras que existem no espanhol. No que se refere aos lemas, o número dos possíveis a-code, no DOE, chega a 218 com uma produção de 457 anagramas. Para as formas, o número dos possíveis a-code atinge 4.541, e o número dos anagramas chega a 10.182.

Nas tabelas 13 e 14, são apresentados o número de lemas e formas e as porcentagens de palavras com e sem anagramas, subdivididos de acordo com a quantidade de letras.

nr. let.	nr. lem.	% semanag.	% anag.	nr. let.	nr. lem.	% sem. anag.	% anag.
1	5	100,0	0,0	12	196	100,0	0,0
2	34	82,4	17,6	13	94	100,0	0,0
3	95	87,4	12,6	14	61	100,0	0,0
4	431	78,4	21,6	15	28	100,0	0,0
5	884	86,0	14,0	16	11	100,0	0,0
6	1150	92,3	7,7	17	3	100,0	0,0
7	1301	94,7	5,3	18	3	100,0	0,0
8	1279	96,8	3,2	19	0	-	-
9	958	98,1	1,9	20	0	-	-
10	630	99,0	1,0	21	1	100,0	0,0
11	371	100,0	0,0	22	0	-	-

Tabela 13 – Porcentagem de lemas com e sem anagramas de acordo com a quantidade de letras.

nr. let.	nr. form.	% semanag.	% anag.	nr. let.	nr. form.	% sem. anag.	% anag.
1	5	100,0	0,0	12	6423	97,9	2,1
2	48	75,0	25,0	13	3617	98,5	1,5
3	184	82,1	17,9	14	1623	99,6	0,4
4	1085	68,2	31,8	15	669	100,0	0,0

¹⁴ As análises estatísticas realizadas em Deloche, Debili, Andreewsky (1980) referem-se às listas de palavras inglesas e francesas extraídas de um *corpus* de referência e de listas de frequências.

¹⁵ Sequências de letras em ordem alfabética de uma palavra.

¹⁶ Por exemplo, o a-code **aglo** pode gerar os anagramas: **algo**, **lago**.

5	3041	73,0	27,0	16	199	100,0	0,0
6	5619	77,7	22,3	17	41	100,0	0,0
7	9344	76,7	23,3	18	8	100,0	0,0
8	12281	83,6	16,4	19	1	100,0	0,0
9	13505	86,5	13,5	20	0	-	-
10	12858	92,3	7,7	21	2	100,0	0,0
11	9781	94,7	5,3	22	2	100,0	0,0

Tabela 14 – Porcentagem de formas com e sem anagramas de acordo com a quantidade de letras.

Nas duas tabelas seguintes, podemos ver os cálculos em relação ao número de letras nos lemas (Tabela 15) e nas formas (Tabela 16). Da distribuição dos a-code podem-se originar 1 palavra, 2 palavras, 3 palavras, etc.

<i>l. lemas</i>	<i>n. lemas</i>	<i>a-code 1</i>	<i>a-code 2</i>	<i>a-code 3</i>	<i>a-code 4</i>
1	5	5	0	0	0
2	34	28	3	0	0
3	95	83	6	0	0
4	431	338	36	7	0
5	884	760	49	6	2
6	1150	1062	41	2	0
7	1301	1232	33	1	0
8	1279	1238	19	1	0
9	958	940	9	0	0
10	630	624	3	0	0
11	371	371	0	0	0
12	196	196	0	0	0
13	94	94	0	0	0
14	61	61	0	0	0
15	28	28	0	0	0
16	11	11	0	0	0
17	3	3	0	0	0
18	3	3	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	1	1	0	0	0
22	0	0	0	0	0

Tabela 15 – Distribuição dos a-code nos lemas de acordo com a quantidade de letras.

<i>l. formas</i>	<i>n. formas</i>	<i>a-code 1</i>	<i>a-code 2</i>	<i>a-code 3</i>	<i>a-code 4</i>	<i>a-code 5</i>	<i>a-code 6</i>	<i>a-code 7</i>	<i>a-code 8</i>
1	5	5	0	0	0	0	0	0	0
2	48	36	6	0	0	0	0	0	0
3	184	151	15	1	0	0	0	0	0

4	1085	740	107	34	6	1	0	0	0
5	3041	2220	274	57	15	6	2	0	0
6	5619	4365	428	84	19	9	3	1	0
7	9344	7163	691	159	46	11	8	5	0
8	12281	10268	774	107	25	6	1	0	1
9	13505	11681	687	105	25	7	0	0	0
10	12858	11870	427	36	4	2	0	0	0
11	9781	9264	236	12	1	1	0	0	0
12	6423	6289	64	2	0	0	0	0	0
13	3617	3563	27	0	0	0	0	0	0
14	1623	1617	3	0	0	0	0	0	0
15	669	669	0	0	0	0	0	0	0
16	199	199	0	0	0	0	0	0	0
17	41	41	0	0	0	0	0	0	0
18	8	8	0	0	0	0	0	0	0
19	1	1	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
21	2	2	0	0	0	0	0	0	0
22	2	2	0	0	0	0	0	0	0

Tabela 16 – Distribuição dos a-code nas formas de acordo com a quantidade de letras.

As duas tabelas anteriores mostram que, no DOE, a maioria dos a-code geram palavras únicas, tanto para os lemas como para as formas. Nos lemas, a porcentagem de palavras únicas é de 93,9% em comparação com 6,1% nos anagramas; nas formas, o percentual de palavras únicas é de 87,3%, em contraste com 12,7% nos anagramas.

A concentração do maior número de anagramas, tanto para os lemas como para as formas, encontra-se no a-code= 2, ou seja, quando um a-code origina duas palavras diferentes (anagramas).

Para os lemas, a concentração do maior número de anagramas encontra-se nas palavras de 4 a 7 letras e, para as formas, em palavras com uma longitude entre 5 e 10 letras.

É interessante notar que, para o a-code = 8 *acenorst*, foram obtidos os seguintes 8 anagramas: *canteros* (canteiros), *cartones* (papelões), *contares* (segunda pessoa do singular do verbo *contar* no futuro do subjuntivo), *contraes* (segunda pessoa do singular do verbo *contraer* no presente do indicativo), *cornetas*, *cortasen* (terceira pessoa do plural do verbo *cortar* no pretérito imperfeito do subjuntivo), *costaren* (terceira pessoa do plural do verbo *costar* no futuro do subjuntivo).

Estudos futuros sobre a variante do espanhol cubano – por exemplo, o projeto em curso *Corpus del español de Cuba* (CECu, 1995-2014), realizado entre o Centro de Linguística Aplicada, a Empresa de Desarrollo de Aplicaciones, Tecnología y Sistemas (DATYS), ambos de Santiago de Cuba, e o Instituto de Literatura e Linguística "José Antonio Portuondo", de Havana –, poderiam ser comparados com os resultados alcançados com o DOE para confirmar as características na formação de anagramas, ou

seja, se há um aumento de sua porcentagem, ao aumentar o número de formas, e se a concentração do maior número de anagramas permanece no a-code= 2.

5. Conclusões

Neste artigo, foi feita uma análise detalhada das características de um subconjunto do espanhol de Cuba com base no LAFEC e no DOE. O estudo de lemas, palavras flexionadas, sequências de letras e extensão de palavras, com o auxílio de tabelas, figuras e estatísticas, ofereceu uma visão mais completa das características distributivas da variante cubana do espanhol presente nos estudantes cubanos.

Finalmente, gostaríamos de indicar os endereços da Internet para *download* dos seguintes estudos:

- *Diccionario Ortográfico del Español*
<http://www.ge.ilc.cnr.it/page.php?ID=dic-cubano&lingua=it>
- *Vocabulario Inverso y Anagramas del Español*
<http://www.ge.ilc.cnr.it/page.php?ID=inverso&lingua=it>

Referências

DELOCHE, G., DEBILI, F., ANDREEWSKY, E. Order Information Redundancy of Verbal Codes in French and English: Neurolinguistic Implications, *Journal of Verbal Learning and Verbal Behavior*, Nr. 19, 1980, p. 525 – 530.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C. Rimas y anagramas en el léxico del escolar cubano. Actas-I, VII Simposio Internacional de Comunicación Social, Santiago de Cuba 22-26 de enero 2001. Ediciones Centro de Lingüística Aplicada y Universidad de Málaga, 2001, p. 101-105.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C., RUIZ MIYARES, L. *Diccionario Ortográfico del Español basado en el léxico del escolar cubano*, CNR – Istituto per i Circuito Elettronici, Genova, Italia, Centro de Lingüística Aplicada, Santiago de Cuba, Cuba. Genova: Ed. CNR – Istituto per i Circuito Elettronici, 2000a.

MARCONI, L., MIYARES BERMÚDEZ, E., RATTI, D., ROLANDO, C., RUIZ MIYARES, L. *Vocabulario Inverso y Anagramas del Español basado en el léxico del escolar cubano*; CNR – Istituto per i circuito Elettronici, Genova, Italia, Centro de Lingüística Aplicada, Santiago de Cuba, Cuba; Genova: Ed. CNR – Istituto per i Circuito Elettronici, 2000b.

MIYARES BERMÚDEZ, E. (DIR.). *Léxico Activo-Funcional del Escolar Cubano*, Santiago de Cuba: Ediciones Centro de Lingüística Aplicada, 2006.

MIYARES BERMÚDEZ, E. (DIR.). *Diccionario básico escolar*. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada y Editorial Oriente, 2014.

RATTI, D., MARCONI, L., BURANI, C. *Radiografia dell'italiano*, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università degli Studi di Roma, *La Sapienza*, N. 2, 1995.

RATTI, D., MARCONI, L., MORGAVI, G., ROLANDO, C. *Flessioni-Rime-Anagrammi*. Bolonha: Ed. Zanichelli, 1988.

RUIZ HERNÁNDEZ, V., MIYARES BERMÚDEZ, E. Aplicaciones metodológicas para elevar la cultura del idioma. In: *Estudios de Comunicación Social, 25 años en defensa del idioma*. La Habana: Editorial Academia, 1997, p. 66–73.

ZINGARELLI, N., ZANICHELLI. *Il nuovo Zingarelli minore*. Bologna: Zanichelli, 1987.

Como citar este texto (ABNT):

MARCONI, L.; MIYARES, L.R.; CUTUGNO, P. Tradução de Cyrano da Rosa Silva. Características distributivas do Espanhol de Cuba a partir da análise de dois estudos linguísticos. **Cadernos de Tradução**, Porto Alegre, n. 43, jul/dez, p. 126-148, 2018.