

Maria Olinda Rodrigues Santana

Departamento de Letras da Universidade de Trás-os-Montes e Alto Douro;
osantana@utad.pt

Crónicas de Alçada Baptista, Inês Pedrosa e Júlio Machado Vaz: uma análise lexicométrica

1. Introdução

A leitura de crónicas jornalísticas de alguns escritores portugueses contemporâneos nas mais variadas revistas levaram-nos a colocar a seguinte questão:

– a escrita jornalística de escritores actuais anuncia a escrita literária dos mesmos ou não?

Para responder à questão levantada, resolvemos efectuar uma análise lexicométrica¹ de um corpus composto por vinte e duas crónicas: oito crónicas de Alçada Baptista publicadas na revista *Máxima*, sete da Inês Pedrosa editadas na *Única* e sete de Júlio Machado Vaz publicadas na revista *Notícias Magazine*.

A seguir, apresentam-se os resultados obtidos.

2. Análise lexicométrica de algumas crónicas de Alçada Baptista, Inês Pedrosa e Júlio Machado Vaz

2.1. A constituição do corpus

As crónicas seleccionadas para o tratamento lexicométrico foram as seguintes:

1. Histórias Inéditas: Você aí nem vai...

BAPTISTA, Alçada – Histórias Inéditas. Do Brasil ao bairro de Campolide, em Lisboa, o mundo é pequeno e as pessoas são grandes. *Máxima*. (Maio de 1998), 37.

2. O Tempo nas Palavras: A gente tem alma?

BAPTISTA, Alçada – O Tempo nas Palavras: A gente tem alma. *Máxima*. (Setembro de 1998), 35.

¹ Conjunto de métodos que permitem operar, a partir de análises estatísticas, reorganizações formais do vocabulário (conjunto de formas actualizadas no discurso, atestadas num texto ou num *corpus* de textos)” CARVALHO, Dulce, MARQUES, Maria Emília Ricardo, SILVA, Maria Fátima, 1999, “Discurso: Práticas lexicométricas” In *Linguística Computacional: Investigação Fundamental e Aplicações*, Lisboa, Edições Colibri / Associação Portuguesa de Linguística, p. 255.

3. O Tempo nas Palavras: O grande avô
BAPTISTA, Alçada – O Tempo nas Palavras: O grande avô. Máxima. (Dezembro de 2000), 31.
 4. O Tempo nas Palavras: uma história de sentimentos
BAPTISTA, Alçada – O tempo nas palavras. Uma história de sentimentos. Máxima. (Agosto de 2002), 57.
 5. O tempo nas palavras: Os novos tempos
BAPTISTA, Alçada – O tempo nas palavras. Os novos tempos. Máxima. (Outubro de 2002), 101.
 6. O tempo nas palavras: A guerra, outra vez
BAPTISTA, Alçada – O tempo nas palavras. A guerra, outra vez. Máxima. (Abril de 2003), 37.
 7. O tempo nas palavras: o valor das coisas
BAPTISTA, Alçada – O tempo nas palavras. O valor das coisas. Máxima. (Agosto de 2003), 33.
 8. O tempo nas palavras: O tempo
BAPTISTA, Alçada – O tempo nas palavras. O tempo. Máxima. (Setembro de 2003), 33.
1. Crónica Feminina: A “péssima conduta” das ditaduras
PEDROSA, Inês – Crónica Feminina: A “péssima conduta” das ditaduras. Única. Expresso. (18 de Abril 2003), 14.
 2. Crónica Feminina: O Presente agora a cores
PEDROSA, Inês – Crónica Feminina: O Presente agora a cores. Única. Expresso. (26 de Julho 2003), 10.
 3. Crónica Feminina: As mulheres exemplares
PEDROSA, Inês – Crónica Feminina: As mulheres exemplares. Única. Expresso. (2 de Agosto de 2003), 12.
 4. Crónica Feminina: Jovens rigorosamente orientados
PEDROSA, Inês – Crónica Feminina: Jovens rigorosamente orientados. Única. Expresso. (9 de Agosto de 2003), 14.
 5. Crónica Feminina: Se te portas mal, vais para o jornal.
PEDROSA, Inês – Crónica Feminina: Se te portas mal, vais para o jornal. Única. Expresso. (15 de Agosto de 2003), 12.
 6. Crónica Feminina: Se, numa noite de Verão, um pacifista
PEDROSA, Inês – Crónica Feminina: Se, numa noite de Verão, um pacifista. Única. Expresso. (23 de Agosto de 2003), 12.
 7. Crónica Feminina: As almas mortas
PEDROSA, Inês – Crónica Feminina: As almas mortas. Única. Expresso. (6 de Setembro de 2003), 12.
1. O cambaleio das palavras: O vampiro triste
VAZ, Júlio Machado – O cambaleio das palavras: o vampiro triste. Notícias Magazine. (16 de Fevereiro de 2003), 7.
 2. O cambaleio das palavras: Lapsos
VAZ, Júlio Machado – O cambaleio das palavras: Lapsos. Notícias Magazine. (30 de Março de 2003), 3.

3. O cambaleio das palavras: Ditadoras
VAZ, Júlio Machado – O cambaleio das palavras: Ditadoras. Notícias Magazine. (6 de Julho de 2003), 7.
4. O cambaleio das palavras: O conselheiro nostálgico
VAZ, Júlio Machado – O cambaleio das palavras: O conselheiro nostálgico. Notícias Magazine. (3 de Agosto de 2003), 7.
5. O cambaleio das palavras: O milagre
VAZ, Júlio Machado – O cambaleio das palavras: O milagre. Notícias Magazine. (17 de Agosto de 2003), 3.
6. O cambaleio das palavras: O jogo
VAZ, Júlio Machado – O cambaleio das palavras: O jogo. Notícias Magazine. (31 Agosto de 2003), 3.
7. O cambaleio das palavras: Variações sobre a vida
VAZ, Júlio Machado – O cambaleio das palavras: Variações sobre a vida. Notícias Magazine. (Outubro de 2003), 7.

Fornece-se abaixo a definição linguística de *crónica jornalística* ou *artigo jornalístico*.

2.2. Definição linguística de crónica jornalística

A *crónica jornalística* é um texto com características de *texto utilitário* (“não-literário”) que pode ser classificado como um texto monológico que apresenta uma temática “estabelecida mais ou menos com precisão” (Vilela, 1999: 487) ou ainda a *crónica* é um “género do discurso jornalístico que, autorizando a expressão do redactor (jornalista ou figura pública), revela a importância, mais ou menos polémica, de um facto ou acontecimento tido como significativo (ou que se quer fazer valer como tal)” (Coutinho, 2001: 50)

Depois de termos citado duas definições linguísticas de *crónica jornalística*, passámos à análise estatístico-descritiva do corpus escolhido.

2.3. Análise estatístico-descritiva do corpus

A execução de um tratamento lexicométrico de um corpus exige a realização de uma edição uniformizada de todas as formas que compõem o referido corpus, com a finalidade de tornar essas formas lexicalmente comparáveis. Para tal, foi necessário aplicar normas específicas à mencionada edição.

As normas usadas na uniformização das formas do corpus foram, essencialmente, as seguintes:

- ligação dos nomes próprios: “Walter-Moreira-Sales. Eu, cá para mim, às vezes até lhe chamo o Walter-Ego”.
- Ligação de topónimos: “Uma vez, no Rio-de-Janeiro”.
- Ligação dos nomes de Instituições: “Colégio-António-Vieira”.
- Ligação de datas: “11-anos”.
- Ligação de expressões temporais “Uma-vez estava a visitar a casa de um amigo que já não via há-muito-tempo.” (*Crónica*: “Histórias Inéditas: Você aí nem vai...” – Alçada Baptista).

- Junção de numerais e nomes aos quais se referem: “Com Raul-Rivero foram presos mais 77 cubanos, entre os quais 37-jornalistas – todos eles por delito-de-opinião” Inês Pedrosa: “A péssima conduta das ditaduras”.
- Junção dos complementos determinativos: “parzinho-de-namorados, chapéu-de-abas-largas, pernas-de-garotas” “O Vampiro Triste” de Júlio Machado Vaz.
- Junção dos nomes e adjectivos, qualificadores dos nomes: “cabelos-longos”, “O vampiro-triste” de Júlio Machado Vaz.

Após a realização da edição uniformizada dos textos do corpus, estes foram analisados pelo programa Stablex². Este programa combinado com macros do Excel construiu uma tabela de distribuição das frequências, uma tabela de desvios reduzidos e vários índices vocabulares por ordem alfabética e decrescente ou hierárquica.

O caminho adoptado no desvelar dos discursos dos três autores passou pelas leituras horizontal e vertical da tabela de distribuição das frequências e da tabela de desvios reduzidos, bem como pela pesquisa do vocabulário nos índices vocabulares construídos por ordem hierárquica e alfabética, tornando possível gizar uma análise contrastiva e comparativa das três variáveis do corpus. Foi possível aferir o vocabulário preferido de cada autor na escrita das suas crónicas entre muitas outras possibilidades proporcionadas por este tipo de metodologia.

2.4. Os dados quantitativos do corpus

Os dados quantitativos do corpus foram reunidos na tabela de distribuição das frequências. Esta tabela recenseou a totalidade do vocabulário do corpus. Assim, o vocabulário dos três grupos de textos organizou-se numa tabela composta por 64 linhas, integrando o vocabulário da frequência mais elevada (652) à mais baixa (1).

O corpus é composto por 14.209 formas. O texto 1 (crónicas de Alçada Baptista) é formado por 4.085 formas, o texto 2 (crónicas de Inês Pedrosa) é constituído por 5.957 formas e o texto 3 (crónicas de Júlio Machado Vaz) por 4.167 formas. O texto que apresenta uma massa vocabular quantitativamente mais significativa é o da escritora Inês Pedrosa.

² Programa e manual da autoria de CAMLONG. A., 1991, *Stablex Pratique: indexation des textes, traitement statistique des lexiques, extraction des séquences, création des dictionnaires*, Toulouse, Teknea.

Tabela 1: Distribuição de Frequências do corpus

		14209	4085	5957	4167
63	p		0,287493842	0,419241326	0,293264832
	q		0,712506158	0,580758674	0,706735168

Ordem	Ocorr	N.º pal	Freq.	T1	T2	T3
1	652	1	652	176	296	180
2	569	1	569	222	225	122
3	539	1	539	168	196	175
4	466	1	466	126	187	153
5	442	1	442	87	230	125
6	222	1	222	76	81	65
7	166	1	166	47	90	29
8	163	1	163	41	92	30
9	159	1	159	49	55	55
10	158	1	158	52	61	45
11	153	1	153	58	57	38
12	152	1	152	43	61	48
13	145	1	145	52	74	19
14	135	1	135	34	55	46
15	116	1	116	22	54	40
16	115	1	115	33	40	42
17	95	1	95	23	41	31
18	87	1	87	29	24	34
19	168	2	84	42	81	45
20	81	1	81	19	46	16
21	74	1	74	25	24	25
22	67	1	67	23	30	14
23	114	2	57	38	34	42
24	110	2	55	11	64	35
25	47	1	47	21	18	8
26	44	1	44	25	4	15
27	42	1	42	18	18	6
28	41	1	41	7	23	11
29	40	1	40	14	8	18
30	78	2	39	26	25	27
31	74	2	37	33	31	10
32	70	2	35	30	26	14
33	33	1	33	15	10	8
34	64	2	32	18	26	20
35	31	1	31	13	9	9
36	60	2	30	25	20	15
37	29	1	29	7	7	15
38	28	1	28	6	16	6
39	135	5	27	51	50	34
40	104	4	26	25	41	38

Ordem	Ocorr	N.º pal	Freq.	T1	T2	T3
41	72	3	24	43	17	12
42	23	1	23	13	8	2
43	88	4	22	36	25	27
44	105	5	21	34	39	32
45	40	2	20	14	16	10
46	133	7	19	60	45	28
47	72	4	18	30	27	15
48	68	4	17	21	37	10
49	112	7	16	40	46	26
50	30	2	15	21	7	2
51	84	6	14	35	29	20
52	65	5	13	19	25	21
53	120	10	12	61	28	31
54	121	11	11	53	43	25
55	100	10	10	29	51	20
56	180	20	9	70	76	34
57	200	25	8	66	92	42
58	224	32	7	72	98	54
59	276	46	6	100	128	48
60	320	64	5	113	135	72
61	428	107	4	135	205	88
62	561	187	3	164	269	128
63	1084	542	2	294	444	346
64	3635	3635	1	732	1537	1366

Antes de avançar na análise lexical, convém distinguir palavra gramatical e palavra nocional ou plena. As palavras gramaticais são, essencialmente, os numerais, os advérbios, os determinantes, os pronomes, as preposições, as conjunções e as locuções adverbiais, prepositivas e conjuncionais. As palavras nocionais ou plenas são os substantivos, os nomes próprios, os verbos, os adjectivos e os advérbios em *-mente* (CAMLONG, 1986: 22).

A leitura da tabela de distribuição de frequências permitiu constatar o seguinte:

- as frequências elevadas congregam as palavras gramaticais. A frequência mais elevada do corpus é a 652, a palavra gramatical que lhe corresponde é a palavra (a). A segunda frequência mais elevada surge 569 vezes e é a palavra gramatical (que), a terceira aparece 539 vezes [palavra gramatical (o)], a quarta assoma 466 vezes, [palavra gramatical (e)], a quinta 442 vezes [palavra gramatical (de)], a sexta tem 222 ocorrências [palavra gramatical (não)], a sétima aparece 166 vezes [palavra gramatical (do)], a oitava tem 163 ocorrências [palavra gramatical (da)], a nona apresenta 159 ocorrências [palavra gramatical (com)], décima tem 158 ocorrências [palavra gramatical (um)], etc.

De acordo com as conclusões de Moshe Koppel e dos seus colegas, a categorização automática dos textos escritos tem demonstrado que as escritoras usam com mais frequência a preposição *de* e os artigos e os escritores os pronomes e a preposição *com*. (ARGAMON, KOPPEL e SHIMONI, 2002, 401-412)

No corpus em análise, talvez por mera coincidência que não deixa, contudo, de ser sintomática, verificámos que a escritora Inês Pedrosa usa 230 vezes (5957) a preposição *de*, característica da escrita feminina, de acordo com os autores acima citados. Alçada Baptista emprega a mesma preposição apenas 87 vezes (4085) e Júlio Machado Vaz 125 vezes (4167).

Quanto à preposição *com*, característica da escrita masculina, Alçada utiliza a preposição 49 vezes (4085), Inês 55 (5957) e Júlio também 55 vezes (4167). Tendo em conta a massa vocabular de cada grupo de crónicas, na verdade, os escritores, Alçada Baptista e Júlio Machado Vaz, empregam mais a preposição *com* do que a escritora, Inês Pedrosa.

Foi possível constatar genericamente que as conclusões de Koppel sobre a escrita feminina e masculina se adequavam aos escritores analisados. Pois, em relação aos pronomes mais usados pelos escritores do que pelas escritoras, no nosso corpus, apurou-se uma presença bem mais significativa do pronome pessoal (*ele*) nos textos de Alçada Baptista e Júlio Machado Vaz (f. 44 – T1 – 25, T2 – 4, T3 – 15) do que no de Inês Pedrosa.

As altas frequências são, como dissemos, palavras gramaticais muito repetidas nos discursos, indicando, normalmente, apenas uma palavra. Porém, também acontece ocorrer nas altas frequências várias palavras. Atente-se no exemplo da ocorrência 168. A décima nona frequência mais elevada do corpus refere-se a duas palavras gramaticais (*como* – com 84 frequências e *por* – com o mesmo número de frequências).

As frequências médias, por exemplo, a ocorrência 78 indica duas palavras gramaticais (*sobre* e *eu*) com 39 frequências cada. A ocorrência 74 refere-se a uma palavra plena (a forma verbal – *tem*) e a uma palavra gramatical (*sua*) com 37 frequências cada. A ocorrência 70 indica duas palavras: uma gramatical e uma nocional (*muito* e *ser*) com 35 frequências cada. A ocorrência 60 reporta-se a duas palavras gramaticais (*quando* e *até*) com 30 frequências cada. A ocorrência 29 alude, tão-só, a uma palavra gramatical (*também*), a ocorrência 28 indica, igualmente, uma palavra gramatical (*só*), a 23 refere-se apenas à palavra gramatical (*onde*), etc.

As baixas frequências são aquelas que contam com um maior número de palavras, e, contrariamente, às altas frequências, que congregam palavras gramaticais, a maior parte das palavras incluídas nas baixas frequências são palavras plenas que apontam para determinados temas escolhidos na escrita de cada autor.

Atentemos, agora, nos dados qualitativos fornecidos na tabela de desvios reduzidos.

2.5. Os dados qualitativos do corpus

A tabela dos desvios reduzidos permite passar dos dados quantitativos aos dados qualitativos, através da aplicação da fórmula do desvio reduzido à tabela de distribuição de frequências (CAMLONG, 1991: 44-45).

No cabeçalho da TDR aparece a soma do total dos desvios reduzidos no corpus e nas três variáveis. Assoma a média no corpus e nas três variáveis e ocorre, igualmente, o teste estatístico do Khi^2 , que é um método de avaliação da significância de uma distribuição normal. A distribuição em causa é uma distribuição perfeitamente normal, pois os valores do Khi^2 são valores próximos da média 0.

Tabela 2: Desvios Reduzidos

1,745	Tot	58,566	-16,722	-40,098
0,027	Moy	0,915	-0,261	-0,627
1,298	Khi2	0,837	0,068	0,393
Ecart:	Moy	Max	Min	
	0,027	10,929	-11,472	
Rang	Fréq	T1	T2	T3
1	-0,157	-0,990	1,798	-0,964
2	0,128	5,411	-1,151	-4,132
3	0,227	1,241	-2,616	1,602
4	0,061	-0,816	-0,785	1,663
5	-0,386	-4,211	4,308	-0,483
6	0,148	1,806	-1,642	-0,015
7	-0,270	-0,124	3,210	-3,355
8	-0,321	-1,014	3,756	-3,063
9	0,161	0,576	-1,874	1,458
10	0,078	1,156	-0,845	-0,233
11	0,113	2,503	-1,170	-1,220
12	0,037	-0,125	-0,448	0,610
13	-0,175	1,892	2,223	-4,291
14	0,018	-0,915	-0,279	1,212
15	-0,098	-2,328	1,010	1,220
16	0,130	-0,013	-1,552	1,695
17	-0,026	-0,977	0,244	0,708
18	0,233	0,945	-2,710	1,998
19	-0,145	-1,074	1,652	-0,723
20	-0,234	-1,052	2,711	-1,893
21	0,144	0,957	-1,655	0,842
22	-0,034	1,009	0,473	-1,516
23	0,226	1,081	-2,618	1,763
24	-0,315	-4,345	3,456	0,574
25	0,056	2,413	-0,504	-1,853
26	0,394	4,114	-4,414	0,694
27	0,002	2,020	0,123	-2,141
28	-0,164	-1,652	1,839	-0,351
29	0,241	0,873	-2,810	2,177
30	0,153	0,894	-1,767	1,026
31	0,018	3,012	-0,006	-2,988
32	0,083	2,608	-0,811	-1,714
33	0,126	2,120	-1,353	-0,642
34	0,017	-0,110	-0,211	0,338
35	0,131	1,622	-1,455	-0,036

36	0,126	2,211	-1,349	-0,736
37	0,160	-0,549	-1,941	2,649
38	-0,142	-0,856	1,632	-0,918
39	0,110	2,318	-1,151	-1,057
40	0,037	-1,061	-0,517	1,616
41	0,298	5,807	-3,149	-2,360
42	0,075	2,943	-0,694	-2,173
43	0,230	2,520	-2,569	0,279
44	0,088	0,822	-0,993	0,259
45	0,026	0,873	-0,247	-0,601
46	0,183	4,170	-1,891	-2,096
47	0,078	2,422	-0,761	-1,583
48	-0,173	0,389	2,087	-2,648
49	0,025	1,629	-0,183	-1,421
50	0,202	4,992	-2,064	-2,726
51	0,131	2,616	-1,375	-1,111
52	0,048	0,086	-0,566	0,528
53	0,377	5,345	-4,127	-0,841
54	0,141	3,658	-1,424	-2,094
55	-0,154	0,055	1,839	-2,049
56	0,011	3,006	0,081	-3,076
57	-0,090	1,328	1,168	-2,587
58	-0,040	1,122	0,554	-1,716
59	-0,110	2,747	1,499	-4,355
60	0,007	2,594	0,095	-2,682
61	-0,203	1,277	2,504	-3,983
62	-0,241	0,253	2,893	-3,387
63	0,047	-1,184	-0,644	1,875
64	-0,104	-11,472	0,439	10,929

A passagem da TDF à TDR foi efectuada através da utilização da função Macrostab (*Excel*). Todos os cálculos foram executados automaticamente. Os dados quantitativos transformaram-se em dados qualitativos e analisáveis.

A TDR proporciona dois tipos de leitura: uma horizontal, outra vertical. A leitura horizontal manifesta as flutuações do emprego de cada item vocabular, nas variáveis, do corpus. A leitura vertical salienta as opções de escrita, as preferências na estruturação de cada texto.

3. Análise lexical comparativa e contrastiva dos três grupos de crónicas

A metodologia seguida na análise lexical comparativa e contrastiva do corpus baseou-se numa leitura horizontal da tabela de desvios reduzidos. Num primeiro momento, foi executada uma leitura horizontal das três variáveis do corpus. Procedeu-se ao levantamento do vocabulário centrado, ou seja, que apresenta valores algébricos à volta da média 0, que é um vocabulário comum aos três textos.

O vocabulário comum pertence a um “tronco comum” dos textos do corpus, qualquer que seja o grau de significação expresso pelo valor algébrico (CAMLONG, 1991: 122-123). Num segundo momento, efectuou-se uma leitura vertical de cada variável. Foi feito o repertório do vocabulário preferencial e diferencial de cada autor. O vocabulário preferencial ou privilegiado é o vocabulário que apresenta valores “significativos positivos”, ou seja, um valor algébrico igual ou superior a **1,96**. O vocabulário diferencial ou rejeitado, pelo contrário, apresenta valores “significativos negativos”, isto é, algebricamente iguais ou inferiores a **-1,96** (CAMLONG, 1991: 122-123).

Num terceiro momento, levantou-se o vocabulário comum aos três textos, posteriormente, de dois em dois e, por último, o vocabulário exclusivo, particular de cada escritor, o que quer dizer que só aparece na escrita, no discurso desse escritor.

3.1. Vocabulário centrado nos três textos

O vocabulário banalizado ou centrado é o vocabulário que se situa no intervalo de confiança entre os valores de **-1,96 e +1,96**, é o vocabulário constituído, sobretudo, por palavras nocionais, características de um determinado tipo de texto.

Nos três grupos de *crônicas*, as frequências comuns são as f. 652, f. 466, f. 222, f. 159, f. 158, f. 152, f. 135, f. 115, f. 95, f. 84, f. 74, f. 67, f. 41, f. 39, f. 32, f. 31, f. 28, f. 26, f. 21, f. 20, f. 16, f. 13, f. 7, f. 2. São 24 as frequências banalizadas ou centradas.

Na primeira triagem efectuada, registou-se o vocabulário comum aos três textos em carácter normal, o vocabulário comum a dois em itálico e o vocabulário exclusivo a um dele em carregado.

– F. 652 (T1 – 0,99, T2 – 1,79, T3 – -0,96) – a, f. 466 (T1—0,061, T2 – -0,78, T3 – 1,66 – e, f. 222 (T1 1,80, T2 – -1,64, T3-0,015) – não, f. 159 (T1 – 0,57, T2 – -1,87, T3 – 1,45) – com, f. 158 (T1 – 1,15, T2 – -0,84, T3 – -0,23) – um, f. 152 (T1 – -0,12, T2 – -0,44, T3 – 0,61) – os, f. 135 (T1 – -0,91, T2 – -0,27, T3- 1,21) – para, f. 115 (T1 – -0,01, T2 – -1,55, T3 – 1,69) – as, f. 95 (T1 – -0,97, T2 – 0,244, T3- 070) – no; f. 84 (2*84 = 168) (T1- -1,07, T2 – 1,65, -0,72) como, por, f. 74 (T1 – 0,95, T2 -1,65, T3 – 0,84) – na, f. 67 (T1 – 1,00, T2 0,47, T3 – -1,51) – mais, f. 41, (T1 -1,65, T2 – 1,83, T3 – -0,35)- das, f. 39 (2*39=78) (T1 – 0,89, T2 -1,76, T3 – 1,02) – sobre, eu, f. 32 (2*32=64) (T1 -0,11, T2 – -0,21, T3- 0,33) – nos, pelo, f. 31 (T1 -1,62, T2 – -1,45, T3 – -0,03) – vida, f. 28 (T1 – -0,85, T2 – 1,63, T3 – -0,91) – só, f. 26 (4*26) (T1 – -1,06, T2 – -0,51, T3 – 1,61) – sempre, meu, sem, aos, f. 21 (5*21=105) (T1 – 0,82, T2 – -0,99, T3 – 0,25) – homem, nunca, seu, entre, lá, f. 20 (20*2) (T1 – 0,87, T2 – 0,24, T3 – -0,60) – este, outros, f. 16 (7*16) (T1 – 1,62, T2 -0,18, T3 -1,42) – homens, num, mesmo, nas, pessoas, essa, todo, f. 13 (5*13 = 65) (T1 – 0,08, T2 – -0,56, T3 – 0,52) – palavra, mulher, depois, sei, jornal, f. 7 (32*7) (T1 -1,12, T2 – 0,55, T3 -1,71) – pena, televisão, falta, (f. 7 **tempo nas palavras** – exclusivo Alçada), (f. 7 **cambaleio das palavras** exclusivo Júlio), dele, simples, no entanto, *neste* (Alçada e Inês), pois, *Deus* (Alçada e Júlio), enquanto, *possível* (Alçada e Inês), escrever, grandes, *estes* (Alçada e Inês), alguns, *então* (Alçada e Inês), alguma, (f. 7 – **Martin Amis** – exclusivo Inês), lhes, ideia, (f. 7 **Portugal** – exclusivo Inês), (f. 7 **Cuba** – exclusivo Inês), (f. 7 – **Crônica Feminina** – exclusivo Inês), *mesma* (Inês e Júlio), bom, algum, *fez* (Inês e Júlio), *aqui* (Alçada e Inês), *viver* (Alçada e Inês), *marido* (Inês e Júlio), f. 2 (542*2) (T1 – -1,18, T2 – -0,64, T3 – 1,87).

Após o levantamento do vocabulário centrado comum aos três textos, foram arrolados com precisão os vocábulos comuns aos três grupos de *crônicas*:

- *a, e, não, com, um, os, para, as, no, como, por, na, mais, das, sobre, eu, nos, pelo, vida, só, sempre, meu, sem, aos, homem, nunca, seu, entre, lá, este, outros, homens, num, mesmo, nas, pessoas, essa, todo, palavra, mulber, depois, sei, jornal, pena, televisão, falta, dele, simples, no entanto, pois, enquanto, escrever, grandes, alguns, alguma, lbes, ideia, bom, algum.*

O *eu*, o sujeito enunciador do discurso, é uma palavra comum aos três grupos de *crônicas*. A presença do *eu* surge ainda no pronome possessivo da 1.^a pessoa (*meu*), na forma verbal da 1.^a pessoa do singular (*sei*). Aparece, igualmente, a 3.^a pessoa do singular nos pronomes possessivos e pessoais (*seu, dele, lbes*).

Os lexemas (*palavra, escrever, ideia*) apontam para temáticas ligadas à escrita, aos meios de comunicação (*jornal, televisão*), à humanidade (*homem, homens, pessoas, mulber*), à vida em geral (*vida*), etc.

Podemos alistar os vocábulos comuns a dois autores. Os vocábulos comuns aos textos de Alçada e Inês são: *neste, possível, estes, então, aqui, viver*. Os textos de Alçada e Júlio têm, apenas, um vocábulo comum: *Deus*. Os textos de Inês e Júlio apresentam três vocábulos comuns: *mesma, fez, marido*.

No vocabulário banalizado, encontrou-se algum vocabulário exclusivo a cada um dos escritores. A expressão (*Tempo nas palavras*) é particular ao vocabulário de Alçada, pois refere-se ao título das suas *crônicas*. As expressões e vocábulos (*Martin Amis, Portugal, Cuba, Crónica Feminina*) são específicos do vocabulário utilizado por Inês. A expressão *Crónica Feminina* reporta-se ao título das *crônicas* da escritora. A expressão (*Cambaleio das palavras*) remete, de igual modo, para o título das *crônicas* de Júlio Machado Vaz, por isso só poderia ser vocabulário particular a este escritor. Repare-se que os títulos das *crônicas* dos três escritores apresentam vocábulos diretamente relacionados com a temática da escrita, da linguagem – (*palavras* – Alçada e Júlio) e (*Crónica* em Inês).

3.2. Leitura vertical das crônicas

Numa segunda etapa, passou-se a uma leitura vertical de cada variável. Foi feito o repertório do vocabulário preferencial e diferencial de cada autor. O vocabulário preferencial apresenta valores “significativos positivos”, por seu turno, o diferencial apresenta valores “significativos negativos”, como foi referido atrás (CAMLONG, 1991: 122-123).

Numa terceira fase, levantou-se o vocabulário comum aos três textos, em seguida, de dois em dois e, por último, o vocabulário exclusivo, particular de cada escritor, ou seja, o que só aparece no discurso desse escritor.

Pela impossibilidade de alistar todos os tipos de vocabulário que foram repertoriados, apresenta-se somente o vocabulário particular, exclusivo de cada variável, com um mero exemplo da metodologia aplicada à análise lexicométrica das *crônicas*.

3.2.1. Vocabulário exclusivo

O vocabulário exclusivo ou particular é o vocabulário que aparece tão-somente num texto e, por isso mesmo, é peculiar à escrita, ao discurso do autor que o emprega.

A título exemplificativo, referirei apenas as palavras e expressões próprias do discurso de Júlio Machado Vaz, nas suas sete *crônicas*. Na frequência 3, arrolei 7 formas: *de imediato, Benfica, ternura, permita, jogo, conhecer; Pianista*. Na frequência 2, recenseei as seguintes palavras e expressões particulares ao discurso do mesmo autor:

- *devem, interpretação, destinado, dormir, genial, defesa, passos, serenidade, irra, decretar; Guilherme, guloso, invocam, gozo, drogas, anos, direitas, dois filhos, aquela, sorriu, avesso, Tim Page, ajuda, atentos, estante, lei, secretária, quarenta anos, sorte, música, abortadeiras, sem dramas, aborto, lapso, lapsos, António, salientar; variações, outras, atendendo, sobreviveu, ouvido, mosca, aparentemente, bolos de bacalbau, voou, seis meses, Boavista, ao tempo, vivo, tiro de partida, sol, telejornais, vagar, tomar, olhos fechados, vento, mau, anfitriões, croquetes, Açores, angústia, abarrotar; elevador, rumo, apertado, contudo, entusiasmo, fizera, fôlego, recordo, conceder; cumprimento, rápido, receio, cliente, crime, cumprir; prazer, diferenças, de 55, hóspede, caçula, juntos, claro, carro, passei, PS, Chopin, nove anos, preocupações, primeiro, participantes.*

Na frequência 1, nos hápax, recenseei 1366 vocábulos e expressões particulares, específicas do discurso de Júlio Machado Vaz. Pela extensão textual não poderei dar aqui o levantamento desse vocabulário.

3.3. Vocabulário preferencial

Procederei de uma forma semelhante à anteriormente exposta para exemplificar o levantamento do vocabulário preferencial. Relativamente a este tipo de vocabulário também só podemos apresentar o vocabulário de um autor, para não alongar a análise, por isso escolhemos o vocabulário privilegiado nas *crônicas* de Alçada Baptista.

3.3.1. Vocabulário preferencial das crônicas de Alçada Baptista

O vocabulário preferencial ou privilegiado no discurso de Alçada Baptista é constituído pelas seguintes palavras e expressões:

- f. 153 (2,5) uma, f. 47 (2,4) era, f. 44 (4,1) ele, f. 42 (2,02) porque, f. 37 (3,01) tem, sua, f. 35 (2,6) muito, ser, f. 33 (2,1) foi, f. 30 (2,2) quando, até, f. 27 (2,3) tudo, tinha, pela, mundo, quem, f. 24 (5,8) há me, está, f. 23(2,9) onde, f. 22 (2,5) história, minha, ela, lhe, f. 19 (4,1) ter, tempo, nossa, nem, esta, agora, assim, f. 18 (2,4) todos, numa, tão, violência, f. 15(4,9) nosso, esse, f. 14 (2,6) nós, pode, seus, estava, isto, outro, f. 12 (5,3) vai, talvez, menos, fazer, relação, grande, isso, mãe, bem, alma, f. 11 (3,6) dizer, seria, nome, parece, melhor, mim, sido, dinheiro, qualquer, amigos, ver, f. 9 (3,0) toda, temos, será, fosse, mal parte, pensar, desde, casa, diz, suas, quase, liberdade, logo, aquilo, às, verdade.
- F. 6 (2,7) – fim, você, tinham, depois de, posso, começou, países, dela, pelos, nossos, **relógio** (exclusivo – Alçada), maneira, capaz, seja, **coisas** (Alçada e Júlio), poderia, consciência, dúvidas, **creio** (Alçada e Júlio), sociedade, **guerra** (exclusivo – Alçada), qual, confiança, **guerras** (exclusivo – Alçada), gente, estavam, **avô** (exclusivo – Alçada), **acho** (exclusivo – Alçada), vivemos, carta, aquele, amigo, teria,

- f. 5 (2,5) – tempos, **Jorge Amado** (exclusivo – Alçada), vi, própria, deles, coisa, primeira, **camimbo** (Alçada e Júlio), **paz** (exclusivo – Alçada), **Nelson Rodrigues** (exclusivo – Alçada), delas, **apesar de** (Alçada e Júlio), público, nazismo, universo, sair, ficou, uns, chamar, **sentimentos** (Alçada e Júlio), (**estamos** – exclusivo Alçada), **reconhecer** (exclusivo – Alçada), capacidade, conta, senhor, questão, quanto, evidente, idade, caso, segundo, lembro-me, infelizmente, **longe** (Alçada e Júlio), exactamente, olhos, olhar, outra, boa, **aspone** (exclusivo – Alçada), todas.

Os vocábulos que ocorrem a negro são particulares e, simultaneamente, preferenciais, por conseguinte os vocábulos acima levantados foram escolhidos, seleccionados por Alçada Baptista na escrita das suas oito *crônicas*.

4. Breve comparação das opções discursivas dos três escritores

As preferências estilísticas de Inês Pedrosa e de Júlio Machado Vaz estão muito próximas. Os dois escritores tomam opções discursivas muito semelhantes, sobretudo, ao nível das escolhas vocabulares e temáticas.

O estilo de Júlio Machado Vaz sobressai no conjunto dos três discursos pelo uso de recursos estilísticos mais variados e com mais literariedade do que os usados nos textos de Inês Pedrosa e também no de Alçada Baptista. Enquanto nos textos de Inês Pedrosa e Alçada ocorre, preferencialmente, adjectivação simples, nos de Machado Vaz assoma a dupla adjectivação, a anteposição da adjectivação e o uso de adjectivação sugestiva, para além do emprego de vários figuras de estilo (animismo, personificação, imagem, metáfora).

De uma forma geral, os temas tratados nas *crônicas* dos três escritores são comuns, por exemplo, os domínios dos sentimentos, do vocabulário amoroso, da transcendência, mas também divergentes. Alçada prefere os temas da *transcendência*, da *guerra* e do *vocabulário religioso*, Machado Vaz os do *futebol*, o do *aborto*, o dos assuntos *automobilísticos*, o dos assuntos *médicos*, Inês Pedrosa o do *jornalismo*, o dos *meios de comunicação social*, o da *educação* e da *crítica político-social*.

Os espaços privilegiados num e noutro autor são diferentes. Machado Vaz refere com frequência o espaço geográfico, social e cultural do *Porto*, Alçada Baptista prefere os espaços geográficos, sociais e culturais do *Brasil* e de *Lisboa*, Inês Pedrosa escolhe, sobretudo, o espaço urbano generalizado (*cidade inteira*) e a praia (*praia da Altura, praias*).

Depois da breve análise vocabular realizada, podemos responder à questão levantada na introdução.

Na verdade, a escrita jornalística destes três autores anuncia a escrita literária de cada um deles. Por que motivo? Porque, apesar dos três autores nas *crônicas* escreverem artigos jornalísticos, em termos linguísticos, textos “com características de *texto utilitário* (“não-literário”) (VILELA 1999, 487), pelas temáticas tratadas (o *amor*; a *transcendência*, a *guerra*, os *problemas sociais*, o *feminismo*, etc.), e, sobretudo, pelo vocabulário selectado na escrita de cada um e comum à escrita dos três: a adjectivação, a diversidade verbal, a abundância adverbial, a presença marcante da 1.^a pessoa, etc., os seus estilos são marcadamente literários.

O *eu*, o sujeito enunciador do discurso, aparece frequentemente nos três grupos de *crônicas*. A 3.^a pessoa ocorre, igualmente, nos pronomes possessivos, pessoais e em inúmeras formas verbais. O *eu* e o *ele* são parâmetros discursivos ancorados directamente na situação de enunciação. Nos discursos analisados, encontramos também os elementos enunciativos indicadores do espaço (*aqui*) e do tempo (*agora*).

As temáticas mais tratadas nos discursos dos três escritores são a *escrita*, os *sentimentos*, a *vida*, a *transcendência*, a *crítica político-social*, a *comunicação social*, etc.

Esperamos que esta mera ilustração de uma análise lexicométrica mais desenvolvida, por nós executada noutra âmbito, possa deixar concluir que os textos jornalísticos aqui analisados são lexical e discursivamente anunciadores de escritas literárias surpreendentes.

BIBLIOGRAFIA

- ARGAMON, Shlomo; KOPPEL, Moshe; FINE, Jonathan e SHIMONI, Anat Rachel (2003), “Gender, Genre, and Writing Style in Formal Written Texts”, *in Text*, 23 (3), August 2003.
- ARGAMON, Shlomo; KOPPEL, Moshe e SHIMONI, Anat Rachel (2002), “Automatically categorizing written texts by author gender”, *in Literary and Linguistic Computing*, 17 (4), p. 401-412.
- BUSSE, Winfried e VILELA, Mário (1986), *Gramática de Valências: apresentação e esboço de aplicação à língua portuguesa*, Coimbra, Livraria Almedina.
- CAMLONG, André (1986), *Le vocabulaire du Sonnet Portugais*, Paris, Fundação Calouste Gulbenkian, Centro Cultural Português.
- _____ (1991), *Stablex Pratique: indexation des textes, traitement statistique des lexiques, extraction des séquences, création des dictionnaires*, Toulouse, Teknea.
- CARVALHO, Dulce; MARQUES, Maria Emília Ricardo e SILVA, Maria Fátima (orgs.) MARRAFA, Palmira, MOTA; Maria Antónia (1999), “Discurso: Práticas lexicométricas”, *in Linguística Computacional: Investigação Fundamental e Aplicações*, Lisboa, Edições Colibri / Associação Portuguesa de linguística, pp. 255-262.
- COUTINHO, Maria Antónia (2001), “Perspectivas Linguísticas sobre Noção de Estilo”, *in Actas do Encontro Comemorativo dos 25 anos do Centro de Linguística da Universidade do Porto*, vol. 2, pp. 41-54.
- VILELA, Mário (1999), *Gramática da Língua Portuguesa: gramática da palavra, gramática da frase, gramática do texto/discurso*, 2.^a edição, Coimbra, Livraria Almedina.

