

A PROPÓSITO DE UM «QUADRO PARA A ALFABETAÇÃO»

por JORGE PEIXOTO

A necessidade de distribuir as fichas pelos vários catálogos de uma biblioteca é trabalho moroso, fatigante, pois torna-se necessário agrupá-las de acordo com os seus *princípios, orientadores*, como a *ordem alfabética*, no caso dos ficheiros dispostos alfabeticamente; a *ordem sistemática*, seguida nos vários catálogos classificados, etc.

O crescimento é tal e as dificuldades são tantas que, amanhã, só com a mecanização nas bibliotecas é que o problema poderá ter uma mais cabal solução.

A própria Biblioteca do Congresso está a envidar esforços no sentido de se passar à mecanização dos seus ficheiros preservando, porém, grande parte dos princípios do seu manual *Filing Rules for the dictionary catalogs of the Library of Congress*. Washington, 1956 ⁽¹⁾, para assim não se causarem profundas e revolucionárias alterações que poderão levar à anarquia e ao caos.

No entanto, como vão resolver os estabelecimentos sem a *dimensão* da Biblioteca do Congresso, as dificuldades criadas pelo grande quantidade de fichas?

Enfim cada um vai adoptando as soluções que lhe são possíveis, *dimensionadas* sempre pela falta de disponibilidades financeiras e de pessoal habilitado.

Ora enquanto a mecanização não surge ainda para este aspecto, os estabelecimentos têm de fazer frente à dificuldade empregando os meios ao seu alcance, aliás bem modestos e quase sempre empíricos. Mas a realidade é essa pobreza e não o sonho desmedido, irrealizável para já...

Para lá da previsão do aumento dos ficheiros, cujas taxas de crescimento se podem avaliar em números proporcionais, o mais importante é facilitar-se a distribuição das fichas para serem em seguida intercaladas.

Nas bibliotecas de larga produção de fichas, um dos problemas mais graves é o da intercalação das mesmas nos respectivos ficheiros. Assim, uma biblioteca que atire com 1000 ou 1500

⁽¹⁾ Ver sobre o assunto: NUGENT, William R.— *The mechanization of the filing rules for the dictionary catalogs of the Library of Congress*. «Library Resources and Technical Services», Richmond, 11 (2) Spring 1967, p. 145-166.

espécies catalogadas por mês para as suas estantes, vê-se com um problema gravíssimo de intercalação, pois são talvez umas 10 000 fichas a introduzir nos vários ficheiros — nos de autores, de matérias, catálogo oficial, etc.

Bem sabemos que os espíritos mais atrevidos — e por vezes estes espíritos correspondem apenas ao dos que pretendem *épater le bourgeois*... — preconizam logo processos de mecanização ou até de automatização — e confundem também estes dois conceitos, pois tais espíritos abusam da ignorância do público a quem se dirigem, gozando de total impunidade... — para solucionar a questão. Infelizmente, a realidade nos nossos estabelecimentos — e em todos os estabelecimentos, sem excepção! — só permite por enquanto a intercalação manual.

Ora, já que os meios materiais são insuficientes, temos de procurar processos que facilitem tal acção manual. A *dimensão* e as disponibilidades financeiras não permitem ir-se mais além...

*
* * *

Um desses processos consiste na utilização de um quadro para alfabetação que apresenta a seguinte distribuição e adoptado por sugestão de Victor Penna (1):

X	Y	Z	U	Q	W	K
J	F	T	L	H	O	Dúvidas
I	G	A	B	C	D	Erros
V	M	S	R	P	N	E

As dimensões totais são as seguintes: 40×126 cm., e as de cada casa para as letras são 10×18 cm., o que permite perfeitamente a colocação das fichas bibliográficas internacionais, 7,5×12,5 (2).

É evidente que este quadro tem de estar relacionado com o estudo estatístico da frequência de letras na língua portuguesa. Na verdade, só com estudos aprofundados de *entrópia* da língua é que poderemos determinar com algum rigor científico o que desejamos alcançar.

Infelizmente os estudos na nossa língua não abundam.

(1) *Cadernos* editou este Quadro para alfabetação que vende ao preço de 20\$00.

(2) Estas exigências surgem sempre que há que fazer uma distribuição alfabética das letras e de acordo com a sua taxa de frequência de utilização. São exemplos claros aqueles que se empregam no teclado das máquinas de escrever ou nas caixas de tipo das oficinas tipográficas.

No caso das máquinas de escrever que dispõem do chamado *teclado português* temos a seguinte ordenação, conhecida por *Cesar*:

H	C	E	S	A	R	O	P	Z
Q	T	D	I	N	U	L	M	X
Y	Ç	J	B	F	V	G	K	W

Sabemos de um jovem que, num trabalho apresentado ao Instituto Tecnológico de Aeronáutica de São Paulo (Brasil), se ocupou, em 1965, do assunto, havendo dado alguns elementos que vamos, de seguida, citar (1). Atente-se, porém, que estes grupos estatísticos são obtidos em textos de política, de jornais, de sonetos e de romances.

Assim temos que as letras mais frequentes são por ordem decrescente as seguintes:

A	T	F
E	M	Q
O	U	Ç
S	C	Z
R	L	J
I	P	X
N	G	
D	B	

Passando depois aos bigramas mais correntes na nossa língua, conforme aquele autor apurou, temos:

DE	NT	NA
RA	DA	EM
ES	TA	QU
OS	TE	TO
AS	OR	IN
DO	AN	RI
RE	OO	IA
AR	SE	CA
EN	AO	AD
ER	MA	UE

Estabelecendo um quadro comparativo da frequência das letras para as línguas principais da Europa ocidental, obteve os resultados indicados na página à frente.

*
* * *

Ora a utilização do quadro de alfabetação tem-se feito por nós a propósito de dois catálogos, o de autores e o alfabético de matérias para facilitar a distribuição das fichas a intercalar nestes catálogos.

(1) NAWA, Óscar Akio — *Determinação da entropia na língua portuguesa*. São Paulo, Instituto Tecnológico de Aeronáutica, 1965. Trabalho policopiado.

(*) Por bigrama entende-se aqui toda a combinação de duas letras. Não foram considerados neste estudo os bigramas formados por duas palavras consecutivas (por exemplo: o grupo *md* na expressão — *um dia*).

FREQUÊNCIA %

Letra	Português	Italiano	Francês	Espanhol	Inglês	Alemão
A	14,68	11,24	8,15	12,53	8,15	6,15
B	1,18	0,72	0,88	1,42	1,44	2,57
C	3,34	4,30	3,06	4,68	2,76	2,84
D	5,03	3,96	4,12	5,86	3,79	5,41
E	12,37	11,20	17,56	13,68	13,10	16,69
F	0,49	0,82	0,96	0,69	2,92	2,04
G	1,20	1,71	1,05	1,00	1,99	3,65
H	0,97	0,89	0,72	0,70	5,26	4,06
I	6,55	11,70	7,56	6,25	6,35	7,81
J	0,37	—	0,60	0,44	0,13	0,19
L	2,92	6,61	5,73	4,97	3,39	2,82
M	4,49	2,61	2,99	3,15	2,54	3,01
N	5,38	7,23	7,32	6,71	7,10	9,90
O	10,99	9,42	5,29	8,68	8,00	2,29
P	2,48	2,83	2,98	2,50	1,98	0,94
Q	0,91	0,38	1,36	0,87	0,12	0,06
R	7,08	6,45	6,29	6,87	6,83	6,54
S	7,72	5,19	8,01	7,98	6,10	6,76
T	4,53	7,00	7,35	4,63	10,47	6,74
U	3,97	2,96	5,99	3,93	2,46	3,70
V	1,75	1,67	1,56	0,89	0,92	1,07
X	0,25	—	0,35	0,22	0,17	0,02
Z	0,39	0,99	0,07	0,52	0,08	1,00
Ç	0,66	—	—	—	—	—
K	—	—	0,04	0,004	0,42	1,88
W	—	—	0,02	0,02	1,54	1,40
Y	—	—	0,12	0,89	1,98	0,03
Totais	100,20	99,88	100,13	100,08	99,99	99,94

Como resultado inicial podemos dizer que o quadro serve perfeitamente para a primeira distribuição de acordo com a primeira letra, quer seja num quer seja noutra daqueles catálogos. Facilita de maneira extraordinária a arrumação das fichas, pois as letras iniciais de maior frequência nos dois catálogos encontram-se dentro de um movimento circular feito pelos braços, sem exigência de grandes esforços físicos, como se pretende igualmente diante de uma caixa de compor de tipografia ou de uma máquina de escrever.

O processo, porém, já se torna difícil quando se trata de fazer a distribuição pela segunda letra, ou mais exactamente pelo respectivo bigrama. Então o *Quadro para a alfabetação* já não dá tão bons resultados, pois a segunda letra predominante em português é uma vogal. E aquele quadro foi concebido tendo apenas em atenção a primeira letra.

Esta dificuldade é ainda maior quando as fichas são pertencentes ao catálogo alfabético de matérias cujas epígrafes utilizam os termos mais correntes.

Portanto, podemos dizer que a experiência da utilização do presente quadro para a alfabetação é positiva relativamente à distribuição das fichas pela primeira letra, quer no caso das de autores (1), quer nas de matérias.

Já se torna, porém, insuficiente no dos bigramas, pelo que será de estudar a criação de um segundo quadro para estes.

A estabelecer-se este *segundo quadro para a alfabetação dos bigramas* ter-se-ia de atender à frequência com que eles aparecem na onomástica portuguesa e nos outros tipos de palavra.

Mas caberá aqui ainda uma pergunta da maior importância: mas não terá para nós ainda mais interesse obter-se a triagem a partir do *trigrama*, já que o bigrama na língua tem uma vogal regra geral como a segunda letra do conjunto? Realmente a terceira letra é que passa a ser a decisiva, a que vai dar, para lá dos *cinco lugares* das vogais, a posição mais avançada no quadro da alfabetação.

Como se vê, este problema da alfabetação entronca directamente com duas disciplinas às quais nós pedimos imediato apoio: à filologia aplicada e à automatização, que programando a questão que temos entre mãos, nos irão permitir, em futuro mais ou menos breve, resolver a dificuldade, que assume proporções volumosas, na casa das centenas de milhar e até dos milhões.

No momento de concluirmos esta nota, vem ainda a propósito salientar o facto, por todos reconhecido, da alta importância e da delicadeza da operação que consiste em intercalar as fichas nos respectivos ficheiros. Se há operações aí que se podem fazer mecânicamente, a maior parte delas é do maior interesse para o estabelecimento e sobretudo para a catalogação.

(1) Para o caso da onomástica da língua francesa temos as seguintes percentagens:

A — 3,5%; B — 12,2%; C — 9%; D — 8,7%; F — 4%; G — 7,5%; H — 3,2%; L — 9,8%; M — 9%; P — 7%; R — 5,5%; S — 5%; T — 3,5%; V-W — 4,1%, veja-se: DUMOULIN — *La division des fichiers alphabétiques*, in «Bulletin O et M», Paris, n.º 3, Janvier-Mars 1962, p. 42-43; nota analítica in Bulletin des Bibliothèques de France, Paris, 7(5) Mai 1962, p. 287.

Só então é que se verificam erros, disparidades de critérios de catalogação, etc. Portanto, esta operação que é fastidiosa, requer uma qualificação do pessoal a que nem sempre damos grande importância, determinando até que ela seja feita por pessoal menos habilitado.

Se pusermos duas equipas a intercalar fichas logo obteremos um teste da maior importância e decisivo: a equipa que assinala os erros e chama as atenções para as disparidades de critérios é constituída pelos melhores elementos. A outra equipa, a que se limita a intercalar sem atender nas dificuldades, é realmente má — e não interessa aos serviços de catalogação que carecem, por isso, de pessoal capaz.

A concluir, queremos ainda falar num outro problema que trataremos em breve: o do tempo gasto nas intercalações de fichas nos vários catálogos. Ficará porém, para outra oportunidade.