Tamanho de palavras e facilidade de leitura para uma melhor compreensão mútua entre aqueles que trabalham em uma organização

SYLLA M. CHAVES *

1. Introdução; 2. Observações qualitativas; 3. Observações quantitativas; 4. Conclusões finais e sugestões.

1. Introdução

1.1 *Comunicação e administração*

O descaso com que nossos administradores ainda consideram os problemas de comunicação força-nos a iniciar nosso trabalho como advogados do óbvio: a comunicação é um instrumento essencial do administrador moderno. Sem ela, todo seu trabalho está fadado ao insucesso.

Não basta ter funcionários bem selecionados e bem treinados. É indispensável que eles estejam informados sobre o que devem fazer; que seus manuais de serviço, determinações ou o que quer que seja atravessem todas as barreiras encontráveis no processo de comunicação e se transformem em conhecimento. Não basta dar as ordens adequadas no que diz respeito à substância. Elas também precisam ser adequadas quanto à

* Professor da EBAP.

forma, isto é, preencher certos requisitos para serem compreendidas, lembradas e, até certo ponto, aceitas e, então, transformadas em comportamento.

O processo decisório requer dados corretos em que se possa basear, seja para avaliação de desempenho, seja para planejamento sobre o que deve e pode ser feito. Sem relatórios que reflitam a realidade, mas que, no entanto, não afoguem o administrador em um mar de informações inúteis, ele não terá como decidir.

A coordenação de esforços em órgãos ou entidades congêneres ou complementares exige intercâmbio de informações. Como conseguir o rápido fluxo das informações indispensáveis sem roubar o tempo necessário para a execução das tarefas que determinaram a criação desses órgãos ou entidades, e sem causar um desgaste emocional desnecessário?

Se está demonstrado que a boa informação não figura dentre os principais fatores de satisfação do empregado, está também demonstrado que os mal-entendidos e a desinformação figuram entre as principais causas de insatisfação. E se considerarmos que, em última análise, as organizações existem para benefício da comunidade em que estão situadas, e que a maior parte do tempo útil dos habitantes da comunidade é passado dentro dessas mesmas organizações, é claro que transformar-se em um ambiente agradável passa a ser uma finalidade essencial de cada organização, finalidade essa que depende de boas comunicações — no seu aspecto descendente, dando a cada um a sensação de contribuir para o todo, e no seu aspecto ascendente, proporcionando a cada um a oportunidade de ser ouvido por aqueles que o dirigem.

Além disso, se considerarmos cada organização como parte de um supersistema, a consecução de seus objetivos dependerá da qualidade das influências que receber e exercer, através de suas comunicações periféricas, que também precisam ser analisadas e aperfeiçoadas.

E, se a comunicação escrita é essencial para a administração moderna (obrigada pela distância física e facilitada pela alfabetização geral), temos de reconhecer que contribuir para que seus funcionários leiam mais facilmente é hoje uma tarefa essencial do administrador.

1.2 *Como medir a facilidade de leitura?*

Em todo ato comunicativo, quer verbal quer não verbal, a responsabilidade pela incompreensão final pode ser atribuída tanto ao comunicador quanto ao destinatário, e o sucesso pode ser obtido por meio de correções efetuadas em qualquer dos extremos. No caso da comunicação escrita, pode-se tornar a descodificação mais fácil aprimorando-se a capacidade do leitor (com cursos de enriquecimento do vocabulário e aperfeiçoamento dos métodos de leitura), ou facilitando-se o seu trabalho (ensinando-se os redatores a utilizarem subtítulos e ilustrações, bem como um estilo e vocabulário ao nível do público a ser atingido).

Em sua atuação no ponto de partida da comunicação, o administrador precisa de uma unidade de medida da facilidade de um texto. É a carac-

terística que os norte-americanos chamam de *readability* e que estamos traduzindo como *facilidade de leitura*. Como medi-la?

A primeira forma de medição, a mais óbvia, consiste em apresentar o texto a uma amostra dos prováveis destinatários, submetendo-os, a seguir, a um teste de compreensão. Além das dificuldades relativas à avaliação da compreensão (diferenciação entre o essencial e o acessório, entre substância e forma, dosagem do *quantum* indispensável em cada categoria, em relação ao tipo de comunicação), esta solução apresenta ainda dois outros problemas. Primeiro, é demasiado onerosa, necessitando elaboração de teste e utilização de amostra representativa para cada comunicação a ser expedida. Segundo, apenas constata os pontos de incompreensão, não indicando a maneira de corrigi-los e fazendo com que cada correção posterior exija nova amostragem e aplicação de novo teste.

A segunda forma de medição tem a primeira como ponto de apoio, mas estabelece um critério geral e unidades de medida a serem utilizadas no próprio texto. Entre essas unidades podem figurar, por exemplo, a freqüência de inversões, de conectivos de subordinação, de vocábulos raros, de termos técnicos e de verbos em voz passiva, o número médio de palavras por período e o número médio de sílabas por palavra.

De todos os possíveis índices, são estes dois últimos os mais utilizados, principalmente em sua forma combinada, conhecida como *fórmula de facilidade de leitura de Rudolf Flesch* (1948).[1] Essa fórmula é a seguinte:

$$\text{FACILIDADE DE LEITURA} = 206{,}84 - 0{,}85\ W - 1{,}02\ S$$

Pouco importa, neste trabalho, saber como Flesch chegou a cada um de seus coeficientes. *W* (*words*), em sua fórmula, corresponde ao *número total de sílabas em uma amostra de 100 palavras; S* (*sentences*) corresponde ao *número médio de palavras por período*. Baseando-se em notas obtidas por estudantes em testes de compreensão de leitura, ele interpreta os seus escores da seguinte forma:[2]

Escore do texto	Grau de dificuldade	Valor médio		Público em potencial	(% dos E. U.)
		S	*W*		
0 a 30	muito difícil	29	192	universitário	(4,5 %)
30 a 50	difícil	25	167	colegial	(24 %)
50 a 60	um pouco difícil	21	155	ginasial	(40 %)
60 a 70	normal	17	147	7.º nível	(75 %)
70 a 80	um pouco fácil	14	139	6.º nível	(80 %)
80 a 90	fácil	11	131	5.º nível	(86 %)
90 a 100	muito fácil	8	123	4.º nível	(90 %)

[1] Apud Miller, George A. *Language and communication*. New York, McGraw-Hill, 1951, p. 131-9.

[2] Flesch, Rudolf. *The art of readable writing*. New York, Harper, 1949. p. 149.

Em curso para aperfeiçoamento de assistentes sociais, que ministrei na PUC-GB, tive oportunidade de verificar, com minhas alunas, a adequação do índice S. Utilizando algumas dezenas de textos (livros, jornais e revistas), previamente classificados por nós como difíceis, médios ou fáceis, calculamos o número de palavras por período e verificamos que ele correspondia perfeitamente à nossa classificação. Entre as médias 8 e 40 os textos apresentavam-se em ordem crescente de dificuldade.

O segundo elemento da fórmula de Flesch, entretanto, não nos permitiu nenhuma conclusão definitiva. No curto tempo de que dispúnhamos, só nos foi possível constatar que as palavras fáceis em português tinham, em geral, mais sílabas que as indicadas pelos autores de língua inglesa, mas a interpretação desse fato teve de ser adiada para outra oportunidade.

A correlação entre palavras curtas e palavras fáceis basea-se na *lei do menor esforço* ou *lei da abreviação*, de Gerge K. Zipf. Depois de analisar o tamanho das palavras em diversos idiomas (alemão, chinês, latim e inglês), valendo-se para isso de quadros estatísticos apresentados por diversos outros pesquisadores (todos indicando freqüência com que a palavra ocorre e número de sílabas), Zipf concluiu que "o tamanho de uma palavra apresenta relação inversa com sua freqüência". [3] Verificando que a maioria das palavras comumente usadas é formada por monossílabos, a maioria dos autores norte-americanos passou a advogar a conveniência de utilizar palavras monossilábicas. Essa é a linha de pensamento que resultou na fórmula de Flesch. Estudaremos sua adequação ao nosso idioma à luz de dois tipos de considerações – qualitativas, baseadas em nossas observações em momentos diversos, e quantitativas, baseadas em pesquisa que realizamos na Universidade de Stanford.

2. Observações qualitativas

2.1 *Vocábulos longos e freqüentes*

Em minha primeira tradução sobre este assunto, verifiquei haver alguma diferença entre o inglês e o português. Tratava-se da tradução destes versos de Arthur Kudner, dirigidos a seu filho:

> Não tenhas medo das palavras grandes,
> pois se referem a pequenas coisas.
> Para o que é grande os nomes são pequenos:
> Assim a vida e a morte, a paz e a guerra,
> a noite, dia, a fé, o amor e o lar.
> Aprende a usar, com grandeza, as palavras pequenas.

[3] Zipf, George K. *The psichobiology of language*, reed. Cambridge, Mass., MIT, 1968. p. 38.

Verás como é difícil fazê-lo,
mas conseguirás dizer o que queres dizer.
Entretanto, quando não souberes o que queres dizer,
usa palavras grandes,
que geralmente servem para enganar os pequenos. [4]

Tive de alterar a lista do autor, que era a seguinte:

All big things have little names
such as life and death, peace and war
or dawn, day, night, hope, love, home. [5]

A solução oferecida por um dicionário bilingüe era *aurora* ou *amanhecer* para traduzir *dawn*, e *esperança* para traduzir *hope*. Para não transformar a argumentação do autor em absurdo, fui forçado a abandonar a palavra *dawn* e substituir *esperança* por *fé* (em inglês, *faith*). Era claro que em nosso idioma há coisas grandes com nomes igualmente grandes.

Advogando o uso de monossílabos, Miller dá como exemplo quatro palavras que "merecem toda nossa atenção" [6] — *thing*, *man*, *time* e *day* — todas elas monossilábicas em inglês. Acontece, porém, que, em português, elas são dissilábicas (coisa, homem, tempo e dia). Temos aqui um indício de que talvez nossa língua, fugindo à regra de Miller, prefira os dissílabos aos monossílabos.

2.2. *Vocábulos curtos e difíceis*

Os dados estatísticos comprovam que os vocábulos mais freqüentes são monossílabos, mas daí não decorre uma regra geral de que os monossílabos são sempre mais fáceis e mais convenientes, como querem Miller e Flesch. Se considerarmos, por exemplo, os vocábulos *Arzt* (pronunciar *artzt*, médico, em alemão) e *chtchi* (transliteração do russo, espécie de sopa), verificaremos que eles são extremamente difíceis para a maioria dos estrangeiros. Nossos patrícios, com sua aversão natural pelos grupos consonantais, seriam levados pela lei do menor esforço, a transformá-los (como fizeram com *sport*) em trissílabos: *ártiste* e *chitichi*.

Da mesma forma como há monossílabos difíceis de pronunciar, há também monossílados que são mais difíceis de entender do que seus sinônimos mais longos. Não nos deteremos em analisar as razões por que um vocábulo curto é desprezado pelo falar quotidiano. Apenas verificaremos que se usa mais freqüentemente

[4] Apud Redfield, Charles E. *Comunicações administrativas*. Trad. Sylla Chaves. Rio de Janeiro, Fundação Getulio Vargas, 1967. p. 32.

[5] Apud Redfield, C. *Communication in management*. 2 ed. Chicago, University of Chicago, 1963. p. 19.

[6] Miller, G. op. cit. p. 94.

cachorro	do	que	cão,
passarinho	do	que	ave,
navio	do	que	nau,
piedade	do	que	dó,
embocadura	do	que	foz,
saudável	do	que	são,
pelado	do	que	nu,
sozinho	do	que	só,
desprezível	do	que	vil,
e descansa	do	que	jaz.

Muitos outros exemplos de vocábulos curtos e difíceis encontramos nos adjetivos eruditos, de origem latina: tredo, prisco, ledo, torvo, etc.

2.3 *Ritmos de cada língua*

A observação da poesia euro-americana mostra uma grande incidência de ritmos trissilábicos (anfíbracos, dáctilos e anapestos) nos idiomas latinos e eslavos, e uma quase exclusividade dos dissilábicos (iambos e troqueus) no idioma inglês. A linguagem coloquial também parece seguir esses ritmos com grande freqüência.

A predominância das oxítonas, em francês, combina-se bem com o ritmo anapéstico (breve-breve-longa), abundante em seus alexandrinos:

> Je suis belle, ô mortels! comme un rêve de pierre [7]
> Autrefois, j'ai connu Ferdousi dans Mysore [8]

Idiomas eslavos e latinos meridionais apresentam ainda maior abundância de ritmos trissilábicos, principalmente anfíbracos (breve-longa-breve). Esse parece ser, também, o ritmo natural do esperanto, língua uniformemente paroxítona.

> Soiúz neruchímyi respúblik svobódnyh [9]
> Al vi, ho potenca senkorpa mistero [10]
>
> Um velho timbira, coberto de glória,
> guardou a memória
> do moço guerreiro, do velho tupi [11]

[7] Baudelaire, Charles. *La beauté.*
[8] Hugo, Victor. *Autrefois j'ai connu...*
[9] Início do hino nacional soviético.
[10] Zamenhof, L. L. *Prece sob o verde estandarte.*
[11] Gonçalves Dias, Antônio. *Y-Juca-Pyrama.*

Amélia não tinha a menor vaidade, [12]
Amélia é que era mulher de verdade.

Julgo não ser necessário comprovar a abundância de anfíbracos e anapestos em nossos autores, nem sua natural correlação com nosso falar quotidiano. Sua monotonia é substituída, em inglès, pela de certas seqüências de iambos (breve-longa) e troqueus (longa-breve), que eles utilizam mais cadenciadamente do que nós:

Once upon a midnight dreary,
while I pondered weak and weary [13]
If you can keep your head when all about you
are losing theirs and blaming it on you [14]

É óbvia a correlação entre os ritmos mais usados e as dimensões e acentuação tônica das palavras disponíveis em cada língua. Alguns algarismos são fornecidos pelo poeta esperantista Kalocsay, em sua comparação dos idiomas francès, italiano, esperanto, russo, polonês, alemão e inglês. [15]

As diferenças constatadas levam-nos a crer que os idiomas anglo-germânicos, em sua preferência por monossílabos, talvez ocupem uma situação peculiar entre os idiomas europeus.

2.4 *Direção das alterações introduzidas*

Segundo Zipf, a lei do menor esforço leva os povos a alterarem sua língua no sentido da redução do tamanho dos vocábulos de uso freqüente. Exemplifica com o ocorrido em inglès com os vocábulos automóvel (usa-se, preferentemente, *car*) e gasolina (abreviada, por eles, para *gas*). O primeiro exemplo é, certamente, generalizável (usando-se, em vários idiomas auto, carro, coche, etc.), mas não o segundo.

E cabe, agora, perguntar se a lei do menor esforço não atua diferentemente do indicado por Zipf — no sentido de uma dimensão ótima (que pode não ser o monossílabo), variável conforme o idioma. Já vimos, anteriormente, que *sport*, em nosso idioma, torna-se esporte. E podemos notar, também, que a familiaridade — que sempre reduz os nomes em inglês — em nosso idioma pode aumentá-los. Enquanto *Joseph*, em inglês, torna-se *Joe*, o nosso José torna-se Zezinho.

É, pois, bem provável, que as observações de Zipf e a fórmula de Flesch não tenham validade além dos idiomas anglo-germânicos. Obser-

[12] Alves, Ataulfo e Lago, Mario. *Saudades da Amélia.*
[13] Poe, Edgar Allan. *The raven.*
[14] Kipling, Rudyard. *If.*
[15] Kalocsay, K. et alii. *Parnasa gvidlibro.* reed. Varsóvia. Pola Esperanto-Asocio, 1968. p. 17.

vações quantitativas irão nortear-nos melhor, na terceira parte deste trabalho.

3. Observações quantitativas

3.1 *Metodologia*

A exigüidade do tempo disponível para esta pesquisa limitou nossa mensuração a apenas dois idiomas, o inglês e o português. Algumas de nossas observações qualitativas permitem supor que o português, no que tange ao tamanho das palavras, representa mais tipicamente os idiomas europeus do que o inglês. Essa confirmação, entretanto, exigirá uma pesquisa mais ampla.

Para esta comparação utilizamos duas revistas equivalentes, uma do Rio de Janeiro [16] e outra de Nova Iorque. [17] Ambas as edições foram vendidas antecipadamente, mais ou menos na mesma ocasião — meados de novembro. A impressão, o estilo, as ilustrações, o conteúdo e, provavalmente, também os públicos, eram bastante semelhantes.

Dos 29 artigos de *Manchete*, em amostra randômica sistemática, selecionamos os de números 1, 4, 7, 10, 13, 16, 19, 22, 25 e 28. Dos 12 artigos de *Look* não houve escolha, pois dois deles eram demasiado curtos para contagem. Os 10 restantes foram utilizados. Pelos títulos abaixo podemos verificar a grande variedade dos assuntos tratados em ambos os idiomas.

Em *Manchete*:

1. O baile da morte
2. Os maiores detetives do mundo
3. Brito × Yustrich
4. Suzy é uma jóia
5. O esqui tropical
6. Orson Welles escreve
7. Henrique Pongetti
8. Gilberto, o homem e o Senado
9. Aqui nasceu a República
10. Escolas em Volta Redonda

Em *Look*:

1. Do we need a new Pledge of Allegiance?
2. Pilgrims' Odyssey
3. Hope for America
4. Fired at 49
5. More electric power

[16] *Manchete*, 21-11-1970.
[17] *Look*, 1.º-12-1970.

6. Face-lifts for men
7. The good guys wear war paint
8. Sex and racism in the NFL
9. Letters to the editor
10. Looking at books

Já que a regra do *lead* é utilizada igualmente em ambas as revistas, tendendo a aproximar os resultados, e não a afastá-los, decidimos computar as 100 primeiras palavras [18] encontradas em cada artigo. Obtivemos, assim, $10 \times 100 = 1.000$ palavras por idioma (incluindo nesse cômputo as repetições), distribuídas nas seguintes categorias:

	Português	Inglês
a) substantivos próprios	70	33
b) substantivos comuns	250	221
c) adjetivos	105	112
d) advérbios	39	52
e) verbos	121	192
f) outros	415	390
Total	1.000	1.000

Na categoria *f* (outros) agrupamos artigos, numerais, pronomes, adjetivos possessivos, preposições e conjunções, isto é, aquelas palavras que, segundo Miller "determinam a forma de nossas afirmações enquanto que os substantivos, adjetivos, advérbios e verbos contribuem para o conteúdo". [19] Tivemos vontade de ser mais rigorosos do que Miller, incluindo verbos de ligação, como *ser,* e certos advérbios, como *não,* na categoria *f.* Finalmente, optamos por uma solução intermediária entre a exatidão e a simplicidade, introduzindo na classificação de Miller apenas as modificações indicadas acima — inclusão de numerais e adjetivos possessivos na categoria *f.* Nenhuma interjeição foi encontrada. Os poucos numerais grafados como algarismos (como *1948*) não foram computados, pois o número de sílabas (relativo à percepção auditiva) dá, nesse caso, uma falsa idéia da facilidade ou dificuldade de leitura (percepção visual). Também abandonamos, pelos mesmos motivos, abreviações como *Jr.* e *Dr.* Ambos esses casos tiveram pouca relevância em nossa contagem final. Os mesmos critérios foram sempre utilizados para as duas línguas consideradas.

Computamos, para cada uma das 2 mil palavras selecionadas, tanto o número de letras quanto o número de sílabas. A divisão silábica inglesa

[18] Em inglês, *word tokens* (inclui repetições das mesmas palavras), e não *word types* (palavras diferentes).

[19] Miller, G. op. cit. p. 94.

foi feita de acordo com o dicionário Webster. Critério semelhante de prioridade da pronúncia sobre a grafia foi também adotado em nosso idioma, para homogenização. Por exemplo, a palavra IN-CÊN-DIO tem 3 (e não 4) sílabas, da mesma forma como *PRAYER* tem uma (e não duas).

3.2 *Resultados globais*

Nossos resultados globais, na forma de número médio de sílabas por palavra e de letras por palavra, em relação a cada idioma e cada categoria gramatical, foram os seguintes:

	Português		Inglês	
	síl/pal	let/pal	síl pal	let/pal
a) subst. pr.	2,6	6,2	2,1	6,5
b) subst. com.	2,9	6,8	1,5	6,8
c) adjetivos	3,0	7,7	2,1	6,4
d) advérbios	2,3	5,8	1,8	5,0
e) verbos	2,7	6,3	1,5	5,0
f) outros	1,2	2,5	1,1	2,8
Média Geral	2,1	4,9	1,5	4,7

A primeira conclusão que podemos tirar do quadro acima é que o idioma inglês é mais *compacto* que o português na pronúncia, mas não na escrita. O mesmo número de palavras que em inglês precisa de 100 letras, em português precisa de 104, uma diferença quase imperceptível. Entretanto, o mesmo número de palavras que em inglês se pronunciam em 100 sílabas exigem 140 em nosso idioma. Em linguagem estatística, a diferença em letras entre os dois idiomas — nas amostras utilizadas — pode ser atribuída ao simples acaso, enquanto a diferença em sílabas (analisada à luz de um teste T entre as duas médias) tem menos de uma chance em mil de que tal aconteça.

A extrema improbabilidade do acaso produzir tão grande diferença no número médio de sílabas por palavra pode também ser observada na constância com que essa média se mantém nos dez textos computados em cada idioma:

Artigos	n.º 1	n.º 2	n.º 3	n.º 4	n.º 5	n.º 6	n.º 7	n.º 8	n.º 9	n.º 10	Média
Em português	2,1	2,2	2,1	2,2	2,3	2,0	2,1	2,3	2,1	2,1	2,1
Em inglês	1,6	1,6	1,5	1,3	1,6	1,5	1,6	1,4	1,8	1,5	1,5

Em virtude desta primeira conclusão, *decidimos restringir nossas comparações, a partir deste ponto, às medidas feitas com a unidade sílabas por palavra*. Esta decisão baseou-se, ainda, nas seguintes razões adicionais:

3.2.1 Zipf, Miller e Flesch referem-se especificamente a sílabas quando consideram o tamanho das palavras, e é exatamente a teoria desses autores que está em pauta. Para compararmos seus dados com os nossos, precisamos de uma unidade comum de medida.

3.2.2 Como o português é um idioma de fonética mais simples do que a inglesa, em virtude de uma constante evolução ortográfica, em que estamos mais adiantados que nossos vizinhos do norte, uma comparação baseada em letras e não em sílabas estaria medindo características provisórias e deixando de lado as mais duradouras. Considerando-se as atuais oscilações ortográficas da língua inglesa entre *night* e *nite, through* e *thru, -borrough* e *-boro, -ous* e *-os*, etc., o número de letras passa a não ter nenhuma importância para a freqüência da palavra (que é uma só em cada par). E podemos antever o dia em que uma reforma ortográfica anglo-americana reduzirá a atual média de 3,1 letras por sílaba do seu idioma para algo próximo da nossa, de 2,3 letras por sílaba, bastando para isso suprimir as inúmeras letras mudas, como nos exemplos citados.

Nossos dados podem, também, ser apresentados globalmente sob a forma de porcentagem de palavras que tem cada dimensão:

Sílabas	Português	Inglês
1	39%	66%
2	27%	21%
3	21%	8%
4	10%	4%
5 ou mais	3%	1%

Deste novo quadro podemos tirar nossa segunda conclusão: nossos resultados enquadram-se na observação geral de Zipf de que, em qualquer idioma, os vocábulos de maior freqüência são os monossílabos.

3.2.3 Nossa terceira conclusão, entretanto, apresenta uma ressalva quanto a esta última: a quantidade de monossílabos em inglês é bem maior do que em português, o que torna duvidosa a afirmação de que a grande freqüência dos monossílabos de uma língua se deve a uma lei universal do menor esforço atuando sempre no sentido do encurtamento dos vocábulos.

3.3 *Palavras essenciais e palavras secundárias*

Para melhor compreender a grande diferença entre a freqüência de monossílabos em inglês (66%) e em português (39%), precisamos examinar

o que ocorre em cada categoria gramatical. No primeiro quadro em que apresentamos nossos dados globais, podemos observar, em todas as categorias, exceto a *f*, grandes diferenças entre os dois idiomas no número de sílabas por palavra. Fundindo essas categorias em uma só, as diferenças aparecem com nitidez ainda maior. E ficamos, finalmente, com a divisão básica de Miller, ligeiramente alterada por nós: 1) palavras essenciais na transmissão do conteúdo (*content words*), i. e., categorias *a, b, c, d* e *e;* e 2) palavras secundárias (*non-content words*), i. e., categoria *f*.

Subdivididas assim e de acordo com o número de sílabas, as 1.000 palavras inglesas e as 1.000 palavras portuguesas por nós examinadas apresentam a seguinte distribuição (em que ess = essenciais, sec = secundárias e TOT = total):

Sílabas	Palavras em					
	português			inglês		
	ess	sec	TOT	ess	sec	TOT
1	44	341	385	290	364	654
2	291	67	268	188	22	210
3	208	6	214	82	2	84
4	98	1	99	42	2	44
5 ou mais	34	0	34	8	0	8
Total	585	415	(1.000)	610	390	(1.000)

Do quadro acima já tínhamos examinado os totais, apresentados, sob forma percentual, no quadro anterior. Vemos, agora, que esses totais se decompõem em dois grupos completamente diversos. É essa observação já nos permite localizar o acerto e o engano das generalizações de Zipf e da fórmula de Flesch.

É perfeitamente aplicável a lei de Zipf às palavras secundárias. Sua correlação com qualquer assunto é causa de sua freqüência, e esta, a de suas pequenas dimensões. É, pois, natural que palavras como *o, a, e, de* e *que* figurem entre as menores e mais freqüentes de qualquer idioma. Não vale a pena ocupar o pensamento com sinonímia nem gastar as cordas vocálicas para acrescentar tão poucos *bits* à nossa informação. É compreensível que os *outrossim* e *todavia* só apareçam como sinais de erudição.

E também podemos aceitar, *em relação às palavras secundárias,* a regularidade de sua diminuição com o passar do tempo e o aumento do uso. O falar do povo, nestes casos, recusa ou reduz as palavras grandes. É perfeitamente plausível, por exemplo, que nosso *aquele* se reduza, um dia, a *quele,* ou mesmo *cle,* da mesma forma como *vossa mercê* se transformou em *você* e, pouco a pouco, vai-se reduzindo a apenas *cê.*

Em nosso último quadro podemos observar que mais de 80% das palavras secundárias utilizadas em qualquer dos dois idiomas são constituídas por monossílabos (82% em português e 93% em inglês). É, pois, aceitável a inferência de que as palavras secundárias mais longas, em sua maioria, ainda não foram adotadas pela linguagem coloquial e, por conseguinte, servem para tornar um texto mais difícil.

O erro de Zipf e Flesch está em imaginar que as palavras essenciais comportam-se de modo idêntico em qualquer idioma. Se as palavras mais freqüentes em qualquer idioma são monossílabos, isso se deve ao fato de serem elas sempre palavras secundárias, e não a uma tendência ao monossílabo, atribuível às palavras essenciais. Desde os primeiros balbucios, as crianças são encaminhadas aos dissílabos repetitivos — mamãe, papai, neném, vovô, auau, etc. — e não aos monossílabos. A predominância de palavras essenciais monossilábicas é uma peculiaridade do idioma inglês, que não deve ser generalizada.

Não comparando os idiomas separadamente em relação a palavras essenciais e secundárias, Zipf não pôde ver o que está bem claro nos histogramas da página seguinte: que há grande variação de idioma para idioma no tamanho das palavras essenciais.

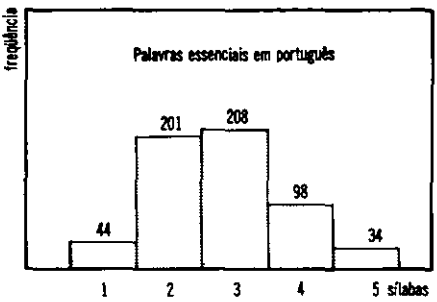
4. Conclusões finais e sugestões

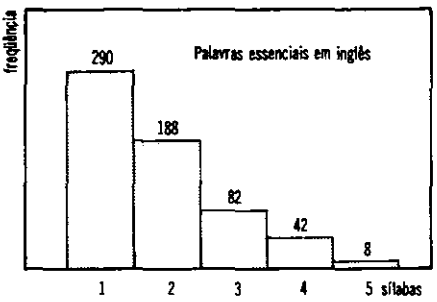
Embora neste trabalho só tenhamos podido examinar o tamanho das palavras em um tipo especial de linguagem, a jornalística, as diferenças quantitativas aí encontradas foram tão grande e tão coerentes com nossas observações qualitativas, que consideramos já poder apresentar algumas conclusões finais e sugestões sobre o assunto, pelo menos relativamente ao nosso idioma e até que estudo mais completo possa ser realizado:

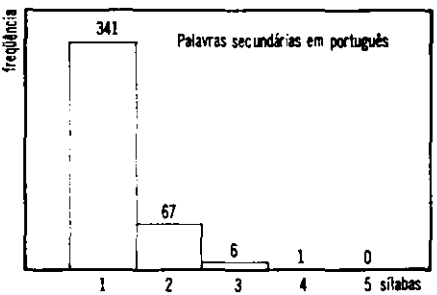
4.1 A fórmula de Flesch não se adapta às características do nosso idioma e, se for utilizada para nortear a escolha de palavras a serem usadas em nossas comunicações, poderá resultar na maior dificuldade do texto assim elaborado.

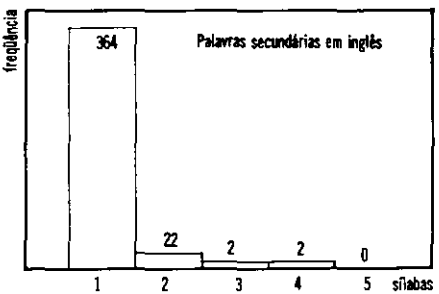
4.2 A lei de Zipf — de redução do tamanho das palavras à proporção que aumenta sua freqüência — aplica-se perfeitamente às palavras secundárias, mas não às essenciais, em que o fenômeno é mais complexo.

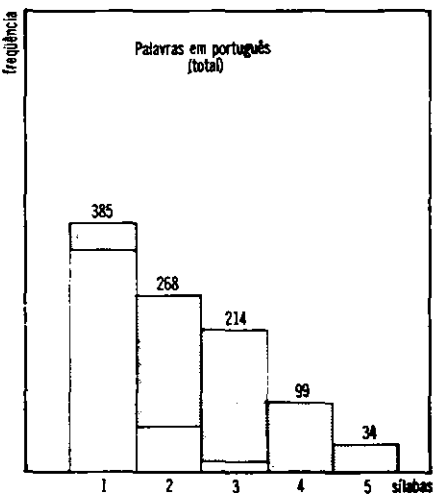
4.3 As palavras essenciais mais curtas são mais fáceis que o seu sinônimo mais longo, em inglês, mas não em português. Essa observação geral talvez possa ser explicada com a história britânica medieval, em que os nobres normandos contribuíram para o vocabulário erudito do idioma em formação, enquanto os seus súditos anglo-saxões contribuíram para o vocabulário popular. O termo de origem latina, geralmente mais longo, permaneceu, assim, menos usado do que o seu sinônimo mais curto, de procedência germânica.

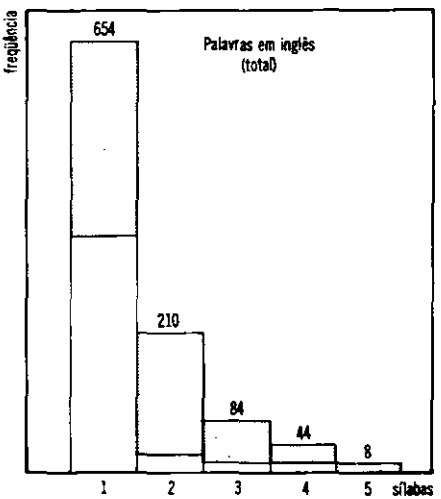

freqüência
Palavras essenciais em português
44
201
208
98
34
1
2
3
4
5 sílabas


freqüência
Palavras essenciais em inglês
290
188
82
42
8
1
2
3
4
5 sílabas


freqüência
Palavras secundárias em português
341
67
6
1
0
1
2
3
4
5 sílabas


freqüência
Palavras secundárias em inglês
364
22
2
2
0
1
2
3
4
5
sílabas


freqüência
Palavras em português
(total)
385
268
214
99
34
1
2
3
4
5
sílabas


freqüência
Palavras em inglês
(total)
654
210
84
44
8
1
2
3
4
5
sílabas

4.4 A lei do menor esforço foi mal interpretada por Zipf, em suas generalizações partidas do estudo da língua inglesa. Buscando, simultaneamente, a brevidade, a clareza e a redução dos sons difíceis, o esforço resultante parece transformar nossas palavras essenciais mais freqüentemente em dissílabos e trissílabos do que em monossílabos.

4.5 Um índice de facilidade de leitura aplicável ao nosso idioma não pode fugir ao cálculo da freqüência individual de cada palavra, nem à elaboração de listas de vocábulos aceitáveis para cada nível e em cada região ou grupo.

4.6 Provisoriamente, índices que considerem o tamanho do período médio, e a freqüência de conectivos, inversões, voz passiva e outras formas mais complexos do pensamento já estarão oferecendo uma medida razoável do grau de dificuldade de cada texto.

4.7 E, como última sugestão, os administradores deveriam envidar maiores esforços na avaliação do grau em que a pletora de papéis que circulam entre seus subordinados é lida e compreendida, e na determinação das razões para que a leitura e a compreensão não sejam maiores. Como em outros casos, perceber-se-á que o dinheiro gasto na identificação e correção dos erros existentes nas comunicações administrativas é recebido com juros, sob a forma de maior eficiência e satisfação.

Summary

The author presents qualitative and quantitative arguments against the universality of the Zipf law and the readability formula of Rudolf Flesch. His quantitative comparison between English and Portuguese shows a similarity in the dimensions of non-content words, but an enormous difference in relation to content words. While English content word-tokens are modally monosyllabic, Portuguese content word-tokens are modally trisyllabic. The Author shows a parallel between this fact and the predominant rhythms in each language (amphibrachic and anapaestic in Portuguese, iambic and trochaic in English) and shows that the easiest Portuguese words are usually not monosyllabic. Both at the beginning and at the end of his paper, the author suggests that the administrator should sponsor readability studies in order to contribute to easier reading and better understanding of the enormous flow of written communications which exist throughout the organization under his control.

Conheça as obras de Administração editadas pela Fundação Getulio Vargas. Peça nosso catálogo. O endereço é: F.G.V., Serviço de Publicações, Praia de Botafogo, 188, Caixa Postal 21.120, ZC-05, Rio de Janeiro, GB, Brasil

AVALIAÇÃO DE PROJETOS DE DESENVOLVIMENTO

Samuel P. Hayes Jr.

A Conferência de Genebra, realizada em 1954, deu origem a este manual, dirigido aos responsáveis pela elaboração de projetos de desenvolvimento.

Seu autor, Samuel P. Hayes Jr., é presidente da Associação de Política Externa desde 1962, tendo sido ainda diretor de programas de reabilitação econômica e social e de desenvolvimento na África do Norte, Europa e Sudeste Asiático.

Neste livro são demonstradas algumas técnicas de mensuração das ciências sociais que permitem avaliar as condições iniciais do projeto, medir a extensão do sucesso de tentativas para produzir mudança social, determinar o resultado global dos esquemas de desenvolvimento social e identificar os fatores importantes para o sucesso dos programas. Uma vez que este desenvolvimento exige certas modificações de ordem psicológica, educacional, tecnológica, sociológica, econômica e política, grande parte do livro dedica-se aos métodos de coleta e análise de dados nessas várias disciplinas.

Eis o plano do livro: 1) descrição do projeto de desenvolvimento e especificação de seus objetivos; 2) escolha dos dados a serem usados para indicar os resultados do projeto; 3) coleta de dados: antes, durante e depois; 4) análise e interpretação dos resultados.

Em apêndice, alguns conceitos estatísticos, procedimentos elementares e referências bibliográficas.