

**A propósito de se rever um
algoritmo de transcrição para o
português brasileiro:
considerações sobre o
*Vocabulário Ortográfico da
Língua Portuguesa (1998)*¹**

Jedicaia Sabará Silva
Maria Carlota Rosa
Fernando Gil V. Resende Jr.
Filipe Leandro Barbosa
Universidade Federal do Rio de Janeiro

Abstract

This paper focuses on the second edition of the *Vocabulário Ortográfico da Língua Portuguesa*, the reference book for Brazilian Portuguese orthography. The new edition, published in 1998, has absorbed a great number of foreign words with their original spelling. There follows from that novelty a confusing guide to Portuguese orthography.

1. INTRODUÇÃO

O presente trabalho retoma o transcritor grafema-fone do sistema TTS (*text-to-speech*), proposto em Barbosa *et al.* (2003). O sistema TTS pode ser entendido como um programa de computador que recebe, como entrada, um texto eletrônico e faz sua leitura, dando, como saída, a voz sintetizada. Um transcritor, nos trabalhos de síntese de voz, é um algoritmo de conversão de grafemas em fones. O transcritor foi a primeira fase de um projeto que procura desenvolver um sistema TTS para textos em português. O trabalho publicado em 2003 foi elaborado em conformidade com o *Formulário Ortográfico* de 12 de dezembro de 1943 e com as modificações introduzidas pela *Lei 5765*, de 18 de dezembro de 1971. Apresenta-se, aqui, o resultado da análise das grafias aceitas na última edição do *Vocabulário Ortográfico da Língua Portuguesa* (doravante, *VOLP*), publicada pela Academia Brasileira de Letras (ABL) em 1998, para posteriores adequações do transcritor. O *VOLP* é a obra que, em última análise, decide o que pode ser considerado uma grafia correta no português do Brasil.

Quando se pensa na ortografia do português brasileiro, vêm à mente os princípios gerais estabelecidos no *Formulário* e as ligeiras modificações introduzidas em 1971, que dizem respeito à exclusão do trema em hiatos átonos (como *saüdade*) e dos acentos diferenciais (como em *côr*; que se distinguia de *cor* pelo acento). A edição do *VOLP* de 1998 aceita como grafias legítimas do português, porém, uma grande quantidade de inovações ortográficas, ao acrescer aos cerca de 340.000 vocábulos da edição anterior, de 1981, mais 6.242 novas formas, em especial, termos estrangeiros oriundos

do hebraico. Com a adoção de grafias estrangeiras, o *VOLP* trouxe consigo uma enxurrada de problemas, ao admitir seqüências estranhas aos princípios apresentados pelo *Formulário Ortográfico*. A ABL, como nota Luiz Antônio Marcuschi (*c.p.*), passou a pautar-se, nessa edição, por não propor uma escrita que fizesse a adaptação fonética de termos estrangeiros a uma forma do português considerada padrão, como no passado. Ao contrário disso, a nova proposta pressupõe a incorporação pura e simples de uma forma gráfica que, sendo estrangeira, tem outras correlações grafema-fone.

O presente trabalho resulta da leitura de todo o *VOLP* e do confronto das grafias encontradas com os princípios que as regem – ou que deveriam regê-las. As secções que se seguem compararão as normas elencadas pelo *Formulário* com as práticas observadas na edição de 1998.

2. OS 12 PRINCÍPIOS E OS PROBLEMAS ARROLADOS PELO FORMULÁRIO

Os 12 princípios que guiaram o *Formulário* dizem respeito à fixação da grafia correta, especialmente para palavras que não tinham uso corrente em Portugal, como brasileirismos, variantes brasileiras e termos estrangeiros. A fixação procurava banir a duplicidade gráfica, “dando-se a cada vocábulo uma única forma, salvo se nele há pronúncia legitimada pelo uso ou pela etimologia”² (item 6º), e a duplicidade prosódica, ao regulamentar a pronúncia culta, pelo registro do timbre, aberto ou fechado, das vogais médias, bem como da sílaba tônica em casos cuja grafia pudesse gerar dúvidas. Seguem-se aos princípios 17 pequenos capítulos, em numeração romana, que tratam dos aspectos considerados pela ABL como aqueles que poderiam causar dúvidas, a saber:

- | | |
|----------------------|--|
| I. Alfabeto | X. Parônimos e vocábulos de grafia dupla |
| II. K, W, Y | XI. Nomes próprios |
| III. H | XII. Acentuação gráfica |
| IV. Consoantes mudas | XIII. Apóstrofo |
| V. SC | XIV. Hífen |
| VI. Letras dobradas | XV. Divisão silábica |
| VII. Vogais nasais | XVI. Emprego das iniciais maiúsculas |
| VIII. Ditongos | XVII. Sinais de pontuação |
| IX. Hiatos | |

3. A EDIÇÃO DO VOLP DE 1998

O que se nota na edição de 1998 é a multiplicação de variantes gráficas, ao admitir-se, como parte da ortografia portuguesa, um sem-número de grafias estrangeiras. As letras especiais, <k, w, y>, ganham presença num grande número de palavras, que vão admitir uma forma variante, da qual elas não fazem parte, e, por vezes, ainda uma terceira variação, em que a grafia estrangeira é nativizada pela simples introdução de algum acento gráfico.

Em decorrência também da nova atitude da ABL, um grafema estranho ao português passa a ser permitido, <ñ> para [ɲ], como em <velardeñita>. Surgem, ainda, grupos consonantais imprevistos, como <jr>, possível em <utitavajrássana>; <cx>, como em <vricxássana>; <ck>, como em <duckeodendrácea>; <zl>, como em <beazliwoodita>; <cg>, como em <gincgoáceo>, <gincgo>; ou <cqu>, como em <acqueta>.

Surgem travamentos gráficos de sílabas como <cf>, em <ecfma>. A representação da assimilação da nasalidade antes de consoante, que prevê <m> antes de <p, b>, mas não antes das outras consoantes, não é sempre seguida: <uranpicloro>, <gentnerita>.

Algumas letras recebem leituras imprevisíveis: <g> antes de <i> pode ser oclusiva, [g], não fricativa [ʒ], em razão de se apresentar como variante de <gu>: <attapulgita>/<attapulguita>; <ragita>/<raguita>; <ramelsbergita>/<rammelsberguita>. Por seu turno, <gu> pode apresentar a variante <gw>: <gwala> ou <guala>;

<gwamba> ou <guamba>; <gwen> ou <güeno>. Também a possibilidade de variação indica que <j> pode ser lido [i], como em <sjoegruvita>/ <sioegruvita>; <thjorsauíta>/ <tiorsauíta>. E surgem outras antigas impossibilidades, como <ss> fora do ambiente intervocálico, como em <ericssonita>.

O emprego de acentos pode permitir leituras imprevistas pelo *Formulário Ortográfico*. A edição de 1998 admite o acento grave para a indicação de abertura de vogal média, como em <ampère> ou <ampère-giro>, adicionando à grafia portuguesa um produto do sistema proposto para a grafia do francês. O acento agudo em <ái> indica a sílaba tônica, mas o conjunto pode representar uma vogal tônica média fechada, [e], como para as formas <daváinea>, <daváineo>. Entretanto, <ai> pode aparecer sem o acento e também deverá ser produzido como [e] em <davainítico>, <davainito>. Para que não haja dúvidas quanto à pronúncia, todas essas formas apresentam-se seguidas de (vê). Em meio às novas possibilidades gráficas, a acentuação ganha o importante papel de marcar a grafia como brasileira, a despeito de poder vir assinalada na quarta última sílaba (<trépolia>), na sexta (<álcaligranada>), na sétima (<álcaliastingsita>) ou mesmo na nona última sílaba (como em <álcaliferroastingsita>). No entanto, palavras que teriam acento se levada em conta a grafia original, podem perder o acento na obra, como o grego <hybris>.

4. AS NOVAS GRAFIAS

Nas seções que se seguem, levantam-se os aspectos em que não há completa concordância entre o *Formulário* e as grafias admitidas como corretas no português brasileiro na nova edição do *VOLP*.

4.1. K, W, Y

Segundo o *Formulário*, as três letras <k, w, y> só podem ser usadas em casos especiais (I, 2). O <k> deve ser substituído por <qu> antes de <e> e de <i>, e por <c> antes de qualquer outra letra (II,3). O *VOLP* de 1998 admite, porém, formas como aquelas a seguir:

- *gymkhana;*
- *petscheskita;*
- *corkita;*
- *akaganeíta;*
- *akatoveíta;*
- *akdalaíta;*
- *akermanita;*
- *aksaíta;*
- *aktashita;*
- *charnockito;*
- *kigélia;*
- *okenita;*
- *okenite;*
- *pakistani;*
- *quilstkes;*
- *rickardita;*
- *rickia;*
- *ricksha;*
- *rijkeboerita;*
- *rucklidgeíta;*
- *saka;*
- *sakharovaíta;*
- *saktismo;*
- *valkéria;*
- *zetzzerita;*
- *zykaíta;*
- *zhemchuzhnikovita.*

Quanto ao <w>, deve ser substituído por <u> ou <v>, salvo os casos de derivados vernáculos de nomes próprios estrangeiros (II,6). São comuns na nova edição, no entanto, exemplos como aqueles a seguir, cuja pronúncia, em geral, leva a dúvidas, uma vez que nem sempre vem indicada e que nem sempre há um dúplex para ela:

- *dawsônia* a par com *dausônia*;
- *lawsônia* a par com *lausônia*;
- *dewara*;
- *dewattado*;
- *dewindtita*;
- *lawrêncio* a par com *laurêncio*;
- *owyheeíta*;
- *rewdjanskita* (paralelas a *revdanskita*, *refdankista* [sic]?);
- *sewamono*;
- *paulównia* a par com *paulównia*;
- *rauwólffia* a par com *rauwólffia*.

O mesmo tipo de prescrição é previsto para <y>: deve ser substituído por <i>, excetuados os derivados de nomes próprios estrangeiros, como <byroniano>, por exemplo. E novamente são muitos os casos em que essa norma não é seguida:

- *aidylyita*;
- *coysevoxiano*;
- *eylettersita*;
- *eytlandita*;
- *fynchenita*;
- *inyoíta*;
- *lyddita*;
- *lyddítico*;
- *lyellita*;
- *lyndochita*;
- *lyonita*;
- *vôyria*;
- *tyndallização*;
- *syssfingídeo*;
- *syssfinge*;
- *peyritschia*;
- *cymraeg*;
- *cyathus*;
- *seymôuria*.

4.2. <H> e indicação de hiatos

Afora os dígrafos <ch, lh, nh>, o <h> só deveria ser empregado no início de palavras e no final de algumas interjeições, como <ah!>, <puh!>. A nova versão do *VOLP* admite, porém:

- (a) <h> final indicando a tônica, como em <*sawah*>;
- (b) <h> como travamento de sílaba não-final, como em <*gabnita*>, <*gabnítico*>, <*gabnoespinélio*>, <*lebnerita*>, <*sablinita*>, <*ubliguita*>;
- (c) depois de <c>, para a representação da oclusiva velar surda [k], como em <*plúchea*>, <*púchara*>, <*púcharo*>, <*tichita*>, variantes, respectivamente, de <*plúquea*>, <*púcara*>, <*púcaro*> e <*tiquita*>;
- (d) depois de <p>, para representação da fricativa labiodental surda [f], como em <*phot*>, <*josephinita*>;
- (e) depois de <t>, sem representar som algum, como em <*iolanthita*>, <*benthâmia*>, <*ebertbemia*>, <*ebertbela*>;
- (f) depois de <s>, para a representação de [ʃ], como em <*shabat*>, <*shamisen*>, <*sbeik*>, <*shunta*>, <*kosher*>, <*elshóltzia*>.

Admite ainda:

- (g) que <h> apareça em outros encontros que não são previstos pela ortografia do português, como <bh>, <gh>, <sch>, <kh>, <rh>, <zh>: <*abhenry*>, <*leightonita*>, <*legborne*>, <*tauschéria*>, <*gymbhana*>, <*mukbinita*>, <*perbamita*>, <*sazbinita*>, <*zbemchuzbnikovita*>.

Com as novas grafias incluídas na obra, nem sempre é claro se <h> deverá simplesmente não ser pronunciado, ou se deverá receber alguma aspiração, como nos exemplos a seguir: <*beazlewoodita*>, <*hawleyta*>, <*hardvíqu*>, <*ballstático*>, <*báckia*>.

O *VOLP* de 1998 admite, também, <h> medial para a indicação de hiatos, átonos ou não: <*bebaísta*>, <*behierita*>, <*beboíta*>.

<lehiita>, <robaíta>, <robâmia>, <utabita>, <sabamalita>, <sabel>. Volta, ainda, a admitir o trema para a indicação de alguns casos de hiatos átonos, banidos pela Lei 5.765: <baüüinita>, <baüüítico>, <baüüyna>, <baüüynita>.

4.3. Letras dobradas

Embora o *Formulário* previsse apenas <rr>, <ss>, <cc> e <çç>, sendo que essas últimas estruturas devem ser pronunciadas como [ks], a nova edição do *VOLP* inclui a possibilidade de várias consoantes dobradas:

- *attapulgita*;
- *allemantita*;
- *allenita*;
- *allegânico*;
- *beggiátoa*;
- *bazzita*;
- *billardiérea*;
- *billietita*;
- *callaghanita*;
- *tlapallita*;
- *canizzarita*;
- *bonnetia*;
- *botallackita*;
- *brammallita*;
- *brannockita*;
- *buttnéria*;
- *cannellone*;
- *toddália*;
- *uddewalita*;
- *ecciese*;
- *rauwólfia*;
- *rafflésia*;
- *eggonita*.

5. CONSIDERAÇÕES FINAIS

Embora o presente trabalho resulte de pesquisa para elaboração de um *software*, as questões aqui levantadas têm conseqüências para aqueles que trabalham com ensino de português ou mesmo com lexicografia. Ao tornar legítima uma enorme quantidade de grafias estrangeiras, quando muito nativizadas pela colocação de alguma acentuação gráfica, o *VOLP* faz com que o *Formulário Ortográfico* deixe de ser o guia para grafar palavras em português. Diferentemente da posição assumida no passado, a nova versão tornou-se um híbrido, já que a escrita do português do Brasil se ajusta ora à pronúncia dos empréstimos, ora à grafia na língua de origem, permitindo, assim, diferentes soluções gráficas para uma mesma palavra.

As conseqüências começam a aparecer em obras lexicográficas. Santos (2001: 24-28) propõe em seu *Dicionário essencial da língua portuguesa* uma “adequação fonográfica” para palavras estrangeiras, que lhe permite introduzir, por exemplo, <sofituer>, <peiperviú>, <badiborde> e <bluxe> para, respectivamente, <software>, <pay per view>, <body-board> e <blush>. Permite-lhe mesmo chegar a sugerir que a antiga revista do Modernismo, a *Klaxon*, fosse transformada em <Clécson>, sequer respeitando a entrada do termo pelo francês.

A multiplicidade gráfica presente no *VOLP* remete o leitor à lembrança de épocas da história da Língua Portuguesa anteriores à padronização das primeiras décadas do século XX no Brasil (FRANÇA, 2003), e até mesmo a fases bem mais antigas. Diferentemente, no entanto, dos ortógrafos dos séculos XVII e XVIII, as novas inclusões não se voltam a uma “ortografia sónica” (GONÇALVES, 1992: 12). As novas duplicações introduzidas, por exemplo, devem ser lidas como letras simples. As possibilidades de pronúncia, oriundas das relações estabelecidas no interior de sistemas gráficos estrangeiros e variados, são geradoras de dúvidas “e empedimento aos que não sabem essas línguas donde ellas vierão” (OLIVEIRA, 1536: 29).

Tampouco se pode invocar, como fazia Gândavo (1574: fol.C^v), por exemplo, a manutenção de uma grafia que fugisse ao esperado

devido “ao uso desta nossa linguagem”. O uso pode justificar, por exemplo, <pizza> em lugar da forma <píteça>, sugerida por Sacconi (1998: 528) em seu *Minidicionário*, mas apresentada como cabeça de verbete na versão ampliada da obra, o *Dicionário essencial*. Mas não se pode invocar o uso para justificar a inclusão, como grafia própria do português, de *gymkbana* ou *szmikita*.

Por outro lado, não há qualquer tentativa de se propor uma representação ortográfica portuguesa de sons raros, caso dos cliques, por exemplo. Assim, como incluir num dicionário a referência a línguas africanas da família *coisan* (ou seria <khoisan>?) como *ba/ /om*, *//au-//en*, em que as barras oblíquas representam o clique lateral alveolar? Por sua vez, em *xbosa*, denominação de outra língua africana, a seqüência <xh> representa uma ejetiva velar. Ao se reproduzir essa grafia, proposta para representar o nome da língua em inglês, adota-se uma seqüência até então inexistente em português. Dificilmente não se levará o leitor a relacionar <xh>, <x> e [ʃ], como em <xavante>, uma vez que o <h>, afora os dígrafos, não deveria ser pronunciado.

As questões aqui levantadas ficaram ao largo das preocupações do *VOLP* na edição de 1998. Acreditamos que a ABL deveria rever a obra, não só pelas muitas falhas na execução gráfica, mas pela desorientação sobre o que deveria reger o sistema ortográfico brasileiro.

NOTAS

¹ Versão preliminar apresentada na Jornada de Iniciação Científica/2003, na Faculdade de Letras da UFRJ. Nossos agradecimentos a Luiz Antônio Marcuschi, pelos comentários à primeira versão deste artigo.

² Atualizamos a grafia do texto de 1943.

REFERÊNCIAS BIBLIOGRÁFICAS

ACADEMIA BRASILEIRA DE LETRAS. *Vocabulário ortográfico da língua portuguesa*. 2. ed. Rio de Janeiro: Bloch Editores, 1998. 795p.

BARBOSA, Felipe *et al.* Grapheme-phone transcription algorithm for a Brazilian Portuguese TTS. In: J.M. MAMEDE *et alii.* *Computational Processing of the Portuguese Language, Lecture Notes in Computer Science*, Proceedings of 6th. Workshop on Computational Processing of the Portuguese Language - Written and Spoken. PROPOR 2003. Heidelberg: Springer, 2003. p. 23-30.

FRANÇA, Angela. *Para uma história da resolução de problemas: da “arte de dizer” na fala carioca às descrições da variante oral do português brasileiro (1937-1960)*. 261p. Tese Tese (Doutorado em Linguística). USP, São Paulo, 2003.

GÂNDAVO, Pero Magalhães de. *Regras que ensinam a maneira de escrever e Orthographia da lingua Portuguesa, com hum Dialogo qu a diante se segue em defensam da mesma lingua*. 1574. Lisboa: Biblioteca Nacional, 1981. Edição facsimilada.

GONÇALVES, Maria Filomena. *Madureira Feijó, ortografista do século XVIII: para uma história da ortografia portuguesa*. Lisboa: Ministério da Educação, 1992. 142p.

OLIVEIRA, Fernão. *Gramática da linguagem portuguesa (1536)*. Ed. crítica, semidiplomática e anastática por Amadeu Torres e Carlos Assunção. Pref. de J. V. de Pina Martins. Lisboa: MM, 2000. 339p.

SACCONI, Luiz Antônio. *Minidicionário Sacconi da língua portuguesa*. São Paulo: Atual, 1998. 686p.

SACCONI, Luiz Antônio. *Dicionário essencial da língua portuguesa*. São Paulo: Saraiva, 2002. 992p.

SANTOS, Volnyr (Coord.). *Dicionário essencial da língua portuguesa (DELP)*. Porto Alegre: Rígel, 2001. 1072p.