

O PROGRAMA DE CONVERSÃO AUTOMÁTICA GRAFEMA-FONEMA NHENHÉM 1.0 E SEU USO COMO FERRAMENTA PARA AUXILIAR O ENSINO DE LÍNGUA MATERNA

Vera VASILÉVSKI¹

RESUMO

Este artigo apresenta os princípios de funcionamento do programa computacional Nhenhém© v.1.0 – que descodifica o sistema verbal escrito do português do Brasil em símbolos fonológicos – e discute sua utilidade na alfabetização. O Nhenhém reflete o alto nível de previsibilidade dos valores que os grafemas do sistema alfabético do português do Brasil têm, uma vez que esse sistema é solidamente estruturado, em grande medida. Norteou a construção do programa a idéia de que essa transparência deve ser levada em conta na escola, a fim de basear o aprendizado em fatos do sistema alfabético do português, que o aluno pode descobrir e testar no computador ou acompanhar os testes de seu professor. Assim o aluno pode tirar conclusões próprias e conscientizar-se de que o sistema verbal é composto por duas modalidades distintas e interligadas: a codificação (escrita) e a descodificação (leitura). O programa considera os princípios do sistema alfabético do português do Brasil conforme descritos por Scliar-Cabral (2003) e a partir disso adota outras definições e redefinições complementares. A conversão é feita nas fontes do Alfabeto Fonético Internacional (IPA). O aplicativo reproduz a parte do sistema verbal escrito oficial que é previsível por regras de descodificação fonológicas com aproximadamente 98% de acerto e o sistema verbal escrito todo, com acerto mínimo de 95%. Ao fornecer a prosódia das palavras da língua portuguesa, o acerto do programa supera 99%.

PALAVRAS-CHAVE: Fonologia; Sistema alfabético do português do Brasil; lingüística computacional; alfabetização.

Introdução²

Apresenta-se um programa eletrônico que descodifica o sistema verbal escrito oficial do português do Brasil em símbolos fonológicos, de acordo com o que é previsível na língua portuguesa e discute-se sua utilidade como ferramenta para auxiliar o ensino-

¹ UFSC, Centro de Comunicação e Expressão, Pós-graduação em Lingüística
Rua José Real Prado, 670, 83501-470, Almirante Tamandaré, Paraná, Brasil (sereiad@hotmail.com)

² Este texto baseia-se em Vasilévski (2008a).

aprendizagem da língua materna. Trata-se do sistema eletrônico de conversão grafema-fonema Nhenhém (2008), em sua versão 1.0.

A construção do Nhenhém envolveu, principalmente, lingüística computacional (linguagem de programação de computador), fonética, fonologia e gramática. A fonologia, dadas suas delimitações, é promissora para ser investigada por meio de um programa eletrônico. Foi decisivo para idealizar o programa tomar conhecimento da clareza do sistema ortográfico do português, tendo em vista seus princípios, que revelam intuições fonológicas. O nível de transparência de um sistema ortográfico mensura-se pela previsibilidade dos valores de seus grafemas, e, nesse sentido, o sistema alfabético do português do Brasil tem grande transparência.

A partir disso, norteou a construção do programa a idéia de que, se o nível de previsibilidade do sistema alfabético do português do Brasil é elevado, então, esse sistema é lógico, ou seja, fundamenta-se em regras sólidas, em grande medida, de forma que é possível reproduzir o sistema ortográfico do português do Brasil em ambiente eletrônico e obter, na tradução automática grafema-fonema, acerto proporcional a tal nível de previsibilidade.

O nome dado ao programa, “nhenhém”, vem do tupi e significa o prosseguir, o repetir-se interminavelmente de um ato – como expressar-se –, um movimento – como o dos lábios –, um som – como a voz –, portanto, “falar, falar”. É abreviatura de “nhenhenhém”, que também significa falatório interminável (AURÉLIO, 2004).

O aplicativo está disponibilizado para pesquisadores e professores que trabalham com ensino de língua materna, que podem usá-lo em suas aulas, em diversas situações. Estima-se que o Nhenhém contribua para que o aluno formule conclusões próprias acerca do funcionamento do português do Brasil e conscientize-se de que o sistema verbal é

composto por duas modalidades distintas e interligadas: a codificação (escrita) e a descodificação (leitura).

Metodologia

O Nhenhém trabalha com a ortografia oficial vigente em 2008 e está em fase de adaptação para trabalhar com as regras do novo acordo ortográfico, que entrará em vigor no próximo ano.

Basicamente, o código fonte do sistema de conversão grafema-fonema Nhenhém 1.0 contém a maioria das regras de descodificação da língua portuguesa, como são apresentadas por Scliar-Cabral (2003, p.81-109). Não foram utilizadas algumas regras referentes às vogais, as regras correspondentes aos ditongos e aos tritongos, bem como não foi usado o arquifonema |R| na tradução de palavras em que “r” sucede “l”, “n” e “s” e em início de palavra. Já que, nessas circunstâncias, ocorre somente /r/ forte (CÂMARA JR., 1986, CAGLIARI, 2002), a letra “r” foi traduzida como fonema.

Algumas adaptações e redefinições teóricas tiveram de ser feitas, pois se trata da conversão de um ambiente (racional, subjetivo) para outro (lógico, objetivo). A partir disso, há alterações nas regras, na forma como são apresentadas, nos símbolos usados, por motivos coerentes e diversos, os quais são detalhados neste artigo.

A descodificação é muito mais simples do que a codificação (SCLIAR-CABRAL, 2003), por isso, parte-se do princípio de que o usuário tem certeza da grafia de o que informa ao programa. O sistema garante a conversão de o que é previsível na língua, para o que há regras. Isso significa que parte da língua não pode ser convertida propriamente.

Dada a grande transparência do sistema alfabético do português do Brasil, para a leitura, os maiores problemas são os valores abertos ou fechados dos grafemas “e” e “o” e os valores de “x” (SCLIAR-CABRAL, 2003). O desenvolvimento do sistema reafirmou isso.

Conforme mostra a literatura da área, há divergência na pronúncia, na separação silábica, na classificação do número de vogais e consoantes da língua portuguesa falada no Brasil. Se é assim na teoria e na prática da teoria, em um sistema eletrônico, essas disparidades ficam mais evidentes.

Padronizações

Enquanto um sistema desenvolve-se, ele exige padronizações e mostra certas incoerências, que são, de fato, incompatibilidades entre sistemas de naturezas distintas ou, até, no mesmo sistema. Como o aplicativo trabalha com lógica, em alguns casos de previsibilidade parcial, houve como criar uma regra lógica que fosse uma sub-regra de outra, assim, uma parte ficou previsível e, então, é convertida pelo sistema.

A maioria das regras de descodificação independentes do contexto foi inserida no sistema, com poucas exceções, em que se considerava o nível da sílaba, que implicavam erro em outros vocábulos, que admitiam outra definição válida e mais harmonizável com outras regras. A partir disso, não se usa o arquifonema |W|, em seu lugar, dependendo do caso, está a vogal /u/, ou /o/, ou a semivogal /w/.

Considera-se hiato a maioria dos chamados ditongos crescentes, à exceção de alguns que sucedem consoantes velares. Da mesma forma, tritongos são somente os inseparáveis, o que os limita aos que sucedem consoante velar, os outros casos configuram

ditongos. Assim, o princípio adotado é o de que o ditongo, bem como o tritongo, é inseparável. Se houver separação, trata-se de hiato, portanto.

Em alguns casos, o sistema preserva uma possibilidade de pronúncia, mesmo que não seja generalizada. Algumas variantes sociolingüísticas não foram consideradas, como a transposição /w/ → /r/: “barde”, por não serem oficiais.

O enfoque do aplicativo é no fonema e nos fenômenos que o rodeiam, e não na sílaba. O fonema revelou-se mais fácil para trabalhar, pois a organização silábica do português extrapola os moldes propostos, e é mais simples e rápido para o processamento eletrônico comparar os caracteres anteriores e posteriores a outro. Desse modo, majoritariamente, o programa é regido por fonemas e, raramente, morfemas, mas não pela sílaba, apesar de conceitos sobre a sílaba ser levados em conta no desenvolvimento das regras.

Por ser basicamente fonológico, o programa tenta abarcar uma pronúncia geral, válida para todos os falantes do português brasileiro. Todavia, algumas decisões em prol de certas variedades podem ser privilegiadas, dado o comportamento da linguagem de programação e o funcionamento das regras em conjunto no sistema, para evitar conflitos entre elas. Ainda há o fator humano, nesse caso, a influência das concepções lingüísticas do desenvolvedor, seu idioleto.

O sistema Nhenhém parte do princípio de que o usuário está certo da grafia do texto que digita como entrada. Se o usuário erra, o sistema descodifica a palavra da mesma forma, com o erro que ela contém. Fornece, portanto, a transcrição de como ela seria se aquela escrita fosse correta.

Outras padronizações, bem como detalhamentos de alguns desses pontos, estão discriminadas no manual de instruções do programa (ver VASILÉVSKI, 2008b).

Tonicidade

Não foi complexo inserir a prosódia brasileira em ambiente eletrônico, porque ela leva em conta a terminação da palavra, na grande maioria dos casos. Com a adaptação da teoria lingüística para a teoria de computação e redefinição de alguns pontos divergentes da teoria lingüística, o Nhenhém 1.0 ficou preparado para lidar com a acentuação de qualquer palavra da língua portuguesa, com raríssimas exceções, estima-se, pois em todos os testes feitos não ocorreu erro de prosódia.

Uma das definições necessárias, nesse aspecto, foi com relação ao local onde apareceria a marca fonológica de tonicidade / / . Se, como ensina a teoria, o pico da sílaba é a vogal, então, quando a sílaba for tônica, a vogal dela pode receber o acento, em vez da primeira letra dessa sílaba. Essa decisão tornou o sistema apto a procurar a vogal tônica a partir da terminação do vocábulo.

O que é acentuado graficamente não precisa de regra no programa, pois o usuário insere o acento, à medida que digita ou cola o texto a ser traduzido. O que não é acentuado graficamente deve ser acentuado pelo Nhenhém. Assim, o sistema verifica as palavras que não são acentuadas graficamente e a elas aplica as regras de acentuação não marcada da língua portuguesa, antes de fazer a tradução fonológica, porque a acentuação é regida pela escrita. Por exemplo, a palavra “amam”, fonologicamente, termina em ditongo, mas não na escrita. Então, ela recebe acento na penúltima vogal, que está na penúltima sílaba, posto que termina em “am”, e não em “ão”. Desse modo, tem-se em “mamão” e “mamam” duas palavras inconfundíveis, exclusivamente por causa da prosódia: ĩ ĩ e ĩ ĩ, assim como “venceram” e “vencerão” ĩ ĩ e ĩ ĩ.

Em conformidade com isso, aplicam-se regras de acentuação a palavras oxítonas e paroxítonas cuja tonicidade não é marcada graficamente. Nesse caso, a terminação da palavra define o acento. Então, recebem acento na última vogal as palavras terminadas por “r”, “l”, “x”, “z”, “n”, “um”, “im”, “om” e “ã”, bem como os vocábulos terminados por “i” e “u”, por ditongo nasal e oral e por tritongo.

Recebem acento na penúltima vogal os vocábulos terminados por “a”, “e”, “o”, “am”, “em”, “ens”. O acréscimo da letra “s” não invalida as regras. As vogais “i” e “u” em posição de semivogal não podem receber acento.

Os clíticos desviam-se das regras do português, pois, se há vogais neles, eles deveriam ter tonicidade, sobretudo se há mais de uma vogal. A solução foi pô-los numa biblioteca, uma lista de palavras que são exceções da língua (“para”, “porque”). O sistema marca o acento de palavras com mais de uma vogal, então, apenas os clíticos com mais de uma vogal estão na biblioteca. Dessa lista também fazem parte palavras com pronúncias irregulares, como “exu” – que deveria ser escrita “echu”, para adaptar-se aos padrões da língua portuguesa – e vocábulos com o radical “troux”.

O hífen é excluído do texto na conversão, assim, palavras unidas por hífen aparecem separadas, com duas prosódias, se for o caso.

Letras e Fonemas

Sistematizar as consoantes é mais fácil do que sistematizar as vogais, apesar da polivalência da letra “x”, que é o maior entrave nesse sentido. Em casos não previsíveis desse grafema, foi usado o símbolo do IPA / /, para indicar que a pronúncia do grafema “x” não pode ser automática, de forma que o usuário deve editar o texto para acertá-la. Não

somente as consoantes recebem leituras diferentes (a exemplo dos valores de “x” e de “s”), isso ocorre também com vogais.

Algumas definições específicas referentes ao tratamento das vogais foram necessárias. Seja nasal ou nasalizada a vogal aparece com o mesmo diacrítico, tendo em vista que o til é o único usado com essas finalidades nas transcrições em português.

A nasalidade fonológica é sempre marcada, e a nasalidade que se dá por assimilação do traço de $\tilde{m}\tilde{n}$ ou $\tilde{n}\tilde{m}$ da sílaba seguinte somente é marcada quando ocorre com a vogal \tilde{a} tônica em paroxítonas, como em “gama” → $\tilde{g}\tilde{a}\tilde{m}\tilde{a}$, quando é inconfundível. A nasalização das outras vogais fica discreta, porque se confunde com a preparação vocal para emitir a consoante nasal seguinte. A única circunstância em que a nasalização por assimilação é marcada em todas as vogais é a mais notória, ou seja, quando antecedem / /, como em “sinhá” → $\tilde{s}\tilde{i}\tilde{n}\tilde{h}\tilde{a}$ (BECHARA, 1973).

Como não é possível pronunciar certas combinações de consoantes, insere-se uma vogal entre elas, geralmente \tilde{i} . Essa epêntese ocorre também na pronúncia de palavras estrangeiras que terminam por consoante que geralmente recebem vogal em português, como “chipe” (*chip*), “nocaute” (*knock out*), quando então se torna silábica. No interior da palavra essa vogal auxiliar é similar à vogal \tilde{i} .

Esse fenômeno epentético torna-se peculiar se a letra “s” está próxima, quando a descodificação de “s” configura uma das maiores dificuldades para o programa. Se “s” entre vogais tem som de \tilde{z} em português, quando “s” vem depois de “b” (“abster”, “observar”), “d” (“adstringente”), devia ser pronunciada como /s/. Como essa combinação é atípica em português, nesses casos, a tendência é ocorrer a epêntese de uma vogal reduzida após a consoante \tilde{b} , como $\tilde{b}\tilde{i}$. Desse modo, o “s” ficaria entre duas vogais, portanto, teria som de \tilde{z} , [], mas não ocorre isso, ou seja, há casos em que esse “s” é surdo, como em “subseção”, \tilde{s} . Portanto, não há

lógica fonêmica aparente nessas ocorrências, o que incita desenvolver estudos relacionados a esse tema.

Os arquifonemas usados no programa são |S| e |R|. O primeiro corresponde às realizações de: “s” e “z” em final de palavra; “s” em final de sílaba que antecede consoante surda ou sonora e copia dela esse traço; “x” após “e” e antes de consoante surda; e “x” no prefixo “ex”. O arquifonema |R| corresponde às realizações de “r” em final de sílaba.

Letras que não têm valor fonológico, como a letra “h” em início e em final de vocábulo, não aparecem na transcrição.

O aplicativo interpreta que há em português, em encontros vocálicos, ditongos decrescentes e os ditongos crescentes “ua”, “uo”, “üe” e “üi”, em casos especiais. Assim, os ditongos do português são todos os decrescentes oficiais e os crescentes “üe” e “üi”, antecidos por “g”; “ua” e “uo”, antecidos por “q”; e “ua” antecido por “g”. Esse último é assim considerado unicamente pela pronúncia inseparável desse par, nesse tipo de sílaba, o que não ocorre com “guo”, que se separa nas conjugações de verbos terminados em “guar”: “apaziguo” → $\text{ĩ} \quad \text{ĩ}$.

Apesar de os ditongos, na escrita, serem reservados às vogais, fonologicamente, há fenômenos que igualam consoantes a vogais e geram ditongos decrescentes por neutralização de “l” em favor de /w/. É o caso de “al” (“caldo” → $\text{ĩ} \quad \text{ĩ}$), que se iguala a “au” (“aura” → $\text{ĩ} \quad \text{ĩ}$); “el” (“tropol” → $\text{ĩ} \quad \text{ĩ}$), que se iguala a “éu” (“troféu” → $\text{ĩ} \quad \text{ĩ}$); “il” (“brasil” → $\text{ĩ} \quad \text{ĩ}$) e “iu” (“corrigiu” → $\text{ĩ} \quad \text{ĩ}$); “ol” (“cerol” → $\text{ĩ} \quad \text{ĩO}$), que em fim de vocábulo não tem oposição com vogal, por não haver “óu” em português, mas no interior de vocábulo iguala-se a “ou” (“couraça” → $\text{ĩ} \quad \text{ĩ}$, “emoldurar” → $\text{ĩ} \quad \text{ĩ}$); e “ul” (“resultado” → $\text{ĩ} \quad \text{ĩ}$), que se assemelha ao hiato final de vocábulos

terminados em “uo” (“recuo” → ĩ ĩ , “supérfluo” → ĩ ĩ), mas é inseparável (“azul”, “sul” → ĩ ĩ , ĩ ĩ).

Então, para o programa, palavras como “ vaidade” e “saudade” contêm ditongo, e não hiato, uma vez que se pronunciam as duas vogais num único esforço expiratório, a exemplo de “eunuco”. O ditongo, frisa-se, é o encontro inseparável de duas vogais ou de uma vogal seguida de consoante que se vocaliza na pronúncia (“l”, “m” e “n”), portanto, de vogal com semivogal, ou, mais raramente, semivogal com vogal, na descodificação fonológica.

A parte da língua portuguesa do Brasil que o Nhenhém menos cobre é a harmonia vocálica e o timbre aberto não marcado das vogais ĩ ĩ e ĩ ĩ . Nesse sentido, o programa considera apenas palavras terminadas em “exo”, “osa”, “osos”, “el” e “ol”, “el” tônico no interior de vocábulo e as que recebem acento gráfico.

O Nhenhém não interfere na pronúncia de “coalho” como ĩ ĩ e “alardear” como ĩ ĩ , porque isso tornaria errada a tradução de “boa”, “reagir” e “tear”, por exemplo. Então, para o programa as traduções dessas palavras são: ĩ ĩ e ĩ ĩ . Testes preliminares revelaram que há como inserir regras para a pronúncia de “e” e “o” nesses casos, mas são regras elaboradas, cujo desenvolvimento demanda pesquisa e tempo.

Algumas Limitações

Além das vogais abertas não acentuadas graficamente, há outras limitações do Nhenhém em sua primeira versão. A pronúncia das formas verbais terminadas em “ar” e “er”, como “falar”, “comer”, dispensa o “r” final na pronúncia de diversas variedades

sociolinguísticas. Da mesma forma, o “s” final das formas verbais desaparece na fala, como em “vamos” → []. Assim, “vamos ver” pronuncia-se []. O sistema não distingue formas verbais, e sempre traduz esses “r” e “s” como pronunciados. Isso não configura erro, mas sim uma possibilidade de pronúncia que pode ser considerada padrão, aliás, usada em conversas formais e na televisão. Cabe ressaltar que se trata dos arquifonemas |S| e |R|, cujo uso cobre fenômenos fonológicos peculiares do português. O usuário pode adaptar a pronúncia na edição do texto convertido, se for o caso.

Então, em algumas situações, o programa faz descodificação em uma pronúncia que nem sempre é a mais comum, mas é possível e também é usada, de forma que um usuário que a usa não é tomado por estrangeiro. Isso ocorre sobretudo com alguns encontros vocálicos.

Os morfemas “ob” e “sub” seguidos por “s” e “trans” seguido por vogal são imprevisíveis quanto ao valor fonológico de “s”, portanto, geram erro nos casos em que esse “s” deve ser descodificado com /z/. Por conta disso, a descodificação do radical “trans”, como em “transação”, fica prejudicada.

O sistema tem uma lista de sinais de pontuação e símbolos que devem ser ignorados. Os sinais de pontuação mais comuns em textos estão inseridos nessa biblioteca, porém, os símbolos são inúmeros, de forma que alguns símbolos raros podem aparecer no texto e gerar um caractere estranho na tradução.

Também há regras que o sistema não contempla, por dependerem de metalinguagem ou do contexto morfossintático e semântico, ou seja, casos imprevisíveis por regras lógicas, como os timbres abertos de /e/ e /o/ em certos verbos rizotônicos, em certos tempos verbais. Essas situações, portanto, geram necessidade de o usuário editar o texto convertido. No sistema, também pode haver regras que provocam erros advindos de conflitos com outras regras e regras que geram alguns desvios, mas ainda não apareceram

nos testes. O uso informará quais são, para que sejam resolvidos, afinal, esta é a versão 1.0, e a partir de agora está disponibilizada ao público interessado.

Alguns Percalços do Percurso

Naturalmente, a própria construção do sistema, sobretudo em seus entraves e percalços, é reveladora de comportamentos da língua. Por converter e analisar textos corridos e longos, em seu percurso, o programa, cujo funcionamento é resultado da interação linguagem de programação-configurações do computador do usuário-língua portuguesa do Brasil, apresentou comportamentos peculiares. Eis alguns, dentre tantos.

O fonema correspondente a “r” em início de palavra, “rr” e “r” após “l”, “n” e “s” é representado por / ɾ / no Nhenhém, porque houve conflito com o arquifonema |R|, uma vez que ambos usam o mesmo caractere, porém, o fonema tem tamanho menor. A solução adotada baseia-se no IPA, em que o símbolo / ɾ / indica que o fonema que o contém é surdo, e isso confere com a característica do fonema em que esse símbolo é usado no programa. Ainda, a presença de um símbolo similar ao fonema original faz o sistema ser mais intuitivo ao usuário. Por conta disso, não se utilizou / r / ou / ɾ /, como o fazem alguns estudiosos da área.

Indubitavelmente, a maior dificuldade foi sistematizar as vogais, sobretudo quando há encontro vocálico. O /l/ decodificado como /w/ conflitava com ditongos terminados em /w/, como em “lousa” e “bolsa”. Nesses casos, as transcrições são: ɨ ɨ e ɨ ɨ. Pela lógica, o valor da letra “s” depois do ditongo deveria ser o mesmo, no entanto, não é.

Outro caso similar se relaciona com a semivogal /w/, seguida pelas consoantes /r/ brando e forte, como em “aura” e “melro”. A transcrição dessas duas palavras é $\text{ɔ̃} \quad \text{r}$ e $\text{ɔ̃} \quad \text{r}$. Uma vez que um fonema sofre interferência do fonema anterior, novamente, como inserir no sistema que, apesar de se tratar, nos dois casos, da semivogal /w/, a letra posterior a ela, embora se escreva da mesma forma, configura dois fonemas distintos? Casos como esses devem ser discutidos com o aluno.

Quanto à prosódia, configurou o maior impasse inserir a tonicidade de hiatos “i” que antecedem “nh”. Esse é o único caso em que o hiato tônico /i/ não recebe acento gráfico, então, a lógica é lê-lo como semivogal. Vale comparar três casos: “polaina”, “tubaína”, “ladainha” → $\text{ɔ̃} \quad \text{ɪ}, \text{ɪ} \quad \text{ɔ̃}, \text{ɪ} \quad \text{ɔ̃}$.

Funcionamento

Um sistema de processamento da língua funciona com entradas e saídas, e a linguagem de programação utilizada para o processamento computacional é responsável pela interface requerida. A entrada é a modalidade escrita padrão da língua portuguesa e a saída é sua representação fonêmica, feita nas fontes do alfabeto fonético internacional (IPA, 2007). Assim, o sistema Nhenhém funciona da seguinte forma, como qualquer sistema eletrônico:

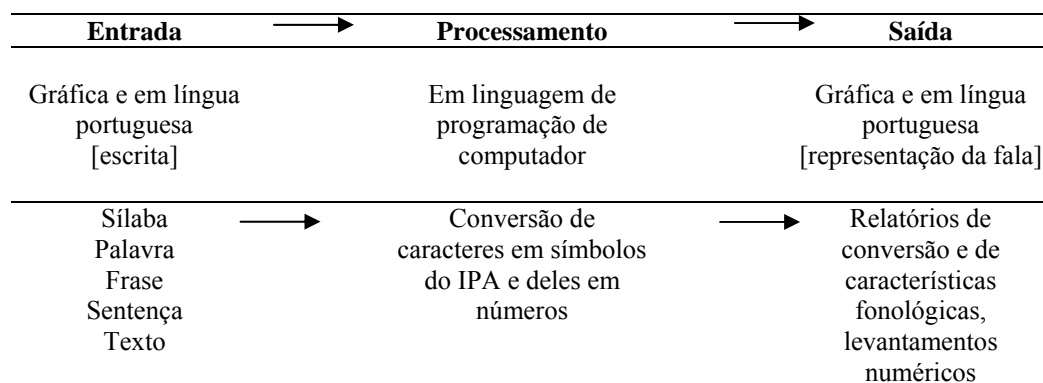


Figura 1: Esquema de processamento computacional

O aplicativo reproduz a parte do sistema verbal escrito oficial que é previsível por regras de descodificação fonológicas com aproximadamente 98% de acerto e o sistema verbal escrito todo, com acerto mínimo de 95%. Ao fornecer a prosódia das palavras da língua portuguesa, o acerto do programa beira 100%.

O funcionamento completo e detalhado do Nhenhém está descrito no manual de instruções que acompanha o programa. O Nhenhém está disponível para ser baixado gratuitamente em: <<http://br.geocities.com/sisnhenhem>>.

A seguir expõem-se os símbolos que o Nhenhém utiliza na conversão.

Quadro 1: Letras e fonemas correspondentes no Nhenhém 1.0

Graf.	Fon.	Exemplo
á	/ /	água
à	ĩ dĩ	àquela
â	ĩ oĩ	lâmpada
ã	ĩ oĩ	maçã
é	ĩ ĩ	pé
é	ĩ oĩ	também, contém
ê	ĩ ĩ	lêvedo
ê	ĩ oĩ	têmpora, ênfase
e	ĩ ĩ	era
e	ĩ ĩ	cante
í	ĩ ĩ	lívida
í	ĩ oĩ	límpido, índio
i	/y/	peito
	/ /	ad(i)vento
ó	ĩ ĩ	pó
õ	ĩõ/	anões
ô	ĩ ĩ	pôs
ô	ĩ oĩ	cômputo, cômscio
o	ĩ ĩ	somente
o	ĩ ĩ	comente
o	ĩ ĩ	mão
o	ĩ ĩ	pato
u	ĩ ĩ	pau, taquara
ú	ĩ ĩ	útil
ú	ĩ oĩ	cúmplice, anúncio
ü	ĩ ĩ	cinquenta
c	ĩ ĩ	acerola
c	ĩ ĩ	acudir
ch	ĩ ĩ	achar
g	ĩ ĩ	gente, girar
gu	ĩ ĩ	guerra, guitarra
h		hoje, ah
j	ĩ ĩ	janela
l	ĩ ĩ	anzol
l	ĩ ĩ	lenço
lh	ĩ ĩ	malha
lh	ĩ ĩ	filhinho
m	ĩ ĩ	mia
n	ĩ ĩ	ano
nh	ĩ ĩ	ninho
qu	ĩ ĩ	quente, quinta
q	ĩ ĩ	aquático
r	ĩ ĩ	cera, prata
r	R	amor
r	ĩ ĩ	bilro, enredo*

r	ĩ ĩ	rosto
rr	ĩ ĩ	amarrar
s	ĩ ĩ	sapo
s	S	mosca, lesma
ss	ĩ ĩ	assar
sc	ĩ ĩ	fascinante
sç	ĩ ĩ	nasça
s	ĩ ĩ	asa
x	ĩ ĩ ou ĩ ĩ	táxi
x	S	expor
x	ĩ ĩ	exato
xc	ĩ ĩ	exceção
z	ĩ ĩ	azedo
z	S	luz

* Por conta de incompatibilidade entre a fonte do IPA utilizada e as configurações internas do computador, o símbolo correspondente ao fonema forte / / adotado nessa versão do programa é ĩ ĩ.

Fonologia, alfabetização e o Nhenhém

O alto nível de transparência do sistema ortográfico do português do Brasil deve ser usado como ferramenta auxiliar na alfabetização. Por exemplo, o aluno aprende que se escreve “m” antes de “p” e “b”, como se isso fosse uma regra para complicar o português escrito. Ensinar assim demonstra despreparo do professor em relação a seu objeto de ensino. Nesse caso, o aluno deve saber que o motivo para haver essa “regra” vem do respeito à fala. Afinal, se /p/ e /b/ são bilabiais, inevitavelmente, ao preparar a boca para articular esses fonemas, a melhor letra para codificar a nasalização é “m”, porque, como “p” e “b”, “m” é bilabial. Na conversão, as letras “m” e “n” desaparecem quando terminam sílaba, pois, se nasalizam a vogal que as antecede, no lugar delas fica a vogal nasalizada, como em “compra” → ĩ ĩ e “manto” → ĩ ĩ, pois, pelo fenômeno da co-articulação, a vogal nasalizada, ao fechar-se, já prepara os articuladores, de acordo com a consoante seguinte.

Se o aluno perceber essa lógica, certamente procurará outras na língua, tentará perceber razões, questionará. Assim, desperta-se a consciência lingüística, que é primordial para superar dificuldades ortográficas e até mesmo o analfabetismo funcional.

É preciso fazer exercícios com os alunos em sala, em que pronunciem as palavras e sintam como os fonemas se convertem na fala, para que tomem consciência de que muitas regras ortográficas obedecem à fala. Então, elas foram feitas para facilitar a leitura e a escrita, e não para dificultá-las, como muitos professores e, conseqüentemente, alunos podem pensar, quando aprendem algo que não lhes é explicado de forma racional.

Para avaliar a competência lingüística do aluno, o professor pode comparar um texto padrão com o mesmo texto, mas escrito pelo aluno. Os relatórios de conversão elucidarão as diferenças entre fonemas e a estatística revelará em traços e números os desvios e acertos que tais textos apresentam. Pode ficar mais fácil lidar com alguns problemas de alfabetização. Por exemplo, ao pedir, oralmente, para um usuário digitar a palavra “ócio” ele digitou “ósseo” no programa. Isso confirmou involuntariamente a coincidência de sons entre vogais agrupadas, que interfere na concepção do usuário sobre o que é dito, pois a transcrição de ambas as palavras é ĩ ĩ.

É difícil explicar que uma letra pode assumir mais de um valor, dependendo das letras que a cercam, mais difícil ainda é explicar que uma letra, circundada pelas mesmas letras, pode ter mais de um valor. Isso reafirma a ineficácia de alfabetizar pelo nome das letras, e não pelo valor que elas assumem nos contextos em que figuram (SCLIAR-CABRAL, 2003). Talvez por isso uma pessoa semi-alfabetizada disse à desenvolvedora que conhece todas as letras, mas não sabe uni-las. Ora, para uni-las é preciso estar atento às transformações de valores que sofrem e provocam ao juntar-se a outras. Brincando com o Nhenhém, aluno e professor podem ver algumas dessas transformações e constatar como são lógicas e necessárias ou não, ou seja, poderão formar juízo. Poderão conferir as pronúncias, discordar delas. Para isso também é necessário reflexão.

Despertar a consciência do aluno não significa que ele encontrará respostas lógicas para tudo, mas que procurará respostas, entenderá que há motivos para as regras ser como são. Um passo a dar nesse sentido é conscientizar o aluno de Letras, de Pedagogia, os futuros professores, de que ao aluno deve ser dada a opção de conhecer e manipular eficientemente a língua padrão, sem que isso gere discriminação dos muitos falares brasileiros. Essa é a língua que ele usará para escrever de forma a ser entendido

por todos no país. Essa é a língua que ele usará para entender o que está escrito em jornais, revistas, livros. Essa é a língua que ele usará para entender os programas televisivos, os programas sobre agricultura, sobre o campo. Essa é a língua que ele usará para escrever a esses programas, pedindo ajuda e informações sobre situações de sua área.

A função da escola é habilitá-lo para isso, sem agredir variações sociolingüísticas, sem agredir a língua padrão. A escola deve tentar garantir-lhe inserção social, preparação para lidar com situações mais amplas da vida, e não mantê-lo encarcerado em seu mundo, sem iniciativa para melhorar, encarar uma sociedade diferente da sua, a qual lhe pode trazer benefícios e prejudicá-lo.

Cabe resgatar uma questão do ensino que a criação do Nhenhém evidencia. Na escola o aluno aprende o sistema de acentuação como regras para complicar, mas o professor deve lidar com isso mostrando ao aluno que esse sistema facilita a leitura e é lógico e econômico – tanto que foi possível inseri-lo completamente no Nhenhém. O aluno precisa entender a parte do sistema de acentuação que se refere ao que não é acentuado graficamente, pois assim se escreve a maioria das palavras em português. Ensinar somente as regras dos acentos gráficos é negligenciar a maior parte da língua.

Por conta de falhas como essa, ao escrever, o aluno desconsidera o trema, por exemplo, nos vocábulos corriqueiros, porque não lhe foi explicado que o trema é um ótimo tira-dúvidas da língua, quando esse aluno mesmo lê um texto. Situação análoga ocorre com acentos gráficos na codificação, e também inversa, pois não é raro ver em textos acentos gráficos e outros diacríticos em usos não oficiais. O trema deixará de fazer parte da língua sem ter sido compreendido por muitos de seus usuários.

A língua tem de ser ensinada na escola, e, como anota o lingüista francês Tonnelat (*apud* CÂMARA JR., 1986), o ensino escolar tem de assentar necessariamente numa regulamentação imperativa. Não adianta ignorar isso.

Aliás, a fuga disso tem provocado lacunas danosas ao ensino de língua materna e facilitado o estabelecimento do analfabetismo funcional, sobretudo porque a escola, na falta de o que ensinar e como ensinar em lugar dos padrões, perde-se no conteúdo, e o aluno não aprende nem padrões. Falta bom senso e preocupação com desenvolver melhores formas de ensinar. Falta uma proposta para ocupar o lugar do famigerado e mal falado padrão. Falta reconhecer que não há como evitar as regras, e mesmo a lingüística contemporânea mais radical desdobra-se em busca de padrões para suas novas regras.

Conhecer bem a língua materna é a melhor forma de lidar com estrangeirismos, ou seja, quebras de padrão. O cidadão poderá fazer reconhecimento da palavra, notar incompatibilidades entre ela e o sistema de sua língua materna, enfim, poderá refletir, formar opinião, posicionar-se com segurança.

Espera-se que o Nhenhém contribua nesse sentido, como ferramenta para ajudar a entender a língua escrita, bem como para submetê-la a testes e brincadeiras. Que tal fazer reconhecimento das diferenças entre os alfabetos ortográfico e fonológico do português? Que tal achar rimas testando a terminação das palavras? Que tal testar as rimas criadas pelos poetas brasileiros, portugueses e de outros países em que se fala português? Que tal analisar siglas e ver quais formam palavras que se desviam do sistema fonológico? Que tal achar a forma ortográfica correta para elas? Que tal soltar a criatividade?

Referências bibliográficas

AURÉLIO. **Novo Dicionário Eletrônico Aurélio**. Versão 5.0. 3.ed. 1. imp. Editora Positivo, 2004.

BECHARA, E. **Moderna gramática portuguesa**. 19.ed. São Paulo: Cia. Editora Nacional, 1973.

CAGLIARI, L. C. **Análise fonológica**: introdução à teoria e à prática. Campinas: Mercado das Letras, 2002.

CÂMARA JR. J. M. **Estrutura da língua portuguesa**. 16.ed. Petrópolis: Vozes, 1986.

INTERNATIONAL PHONETIC ALPHABET – IPA. Disponível em: <<http://www.sil.org/sil/>>. Acesso em: out. 2007.

VASILÉVSKI, V. **Construção de um sistema computacional para suporte à pesquisa em fonologia do português do Brasil**. 2008. 166f. Tese (Doutorado em Lingüística aplicada: fonologia e lingüística computacional) – Programa de Pós-graduação em Lingüística, Universidade Federal de Santa Catarina, Florianópolis.

VASILÉVSKI, V. Manual de instruções. In: **Nhenhém** – programa eletrônico de conversão grafema-fonema. Versão 1.0, 1a. atualiz. Curitiba-Florianópolis, Brasil, agosto, 2008b. Disponível em: <<http://br.geocities.com/sisnhenhem>>.

NHENHÉM: Sistema de Conversão Grafema-Fonema (Curitiba/PR). Vera Vasilévski. **Programa eletrônico para pesquisa em lingüística**. Versão 1.0. Em registro no INPI, Florianópolis-Curitiba, Brasil, março de 2008.

SCLIAR-CABRAL, L. **Princípios do sistema alfabético do português do Brasil**. São Paulo: Contexto, 2003.