SETFON: O PROBLEMA DA ANÁLISE DE DADOS PROSÓDICOS, TEXTUAIS E ACÚSTICOS

Alexsandro Meireles[*]
Ana Cristina Fricke Matte[**]
Rubens Takiguti Ribeiro[***]

RESUMO

SetFon é um sistema de informação *Web* livre para coleta de dados em pesquisas sobre a fala. O sistema aborda o problema central, por meio de componentes, e nasceu da necessidade de se aumentar, significativamente, a quantidade de dados acústico-fonológicos para atender a demandas de estudos estatísticos de expressão da emoção e de estilo. Além disso, apresentamos, neste artigo, o programa e seus componentes, dos pontos de vista da coleta de dados em Fonética Acústica, da Fonologia, da Semiótica, da Tecnologia da Informação e da Computação.

PALAVRAS-CHAVE: Tecnologia. Fonética Acústica. Fonologia. Fonoestilística.

ABSTRACT

SetFon is an open source web information system for data collecting in the field of speech sciences. Its architecture is component based and it emerged from the necessity of raising the acoustic-phonological data quantity up in order to solve the problem of statistic significance in expression of emotion in speech studies. Moreover, we present in this paper the software and its components, in different approaches of the data collecting: Acoustical Phonetics, Phonology, Semiotics, Information Technology and Computation.

KEYWORDS:Technology. Acoustical Phonetics. Phonology. Phonostylistics.

O QUE É O SETFON?

As pesquisas sobre fala, geralmente, dividem-se em abordagens acústicas, fonético-acústicas, fonológicas e expressivas (emoção, atitude, etc.). Essas últimas, especialmente, dão atenção especial ao conteúdo do que foi falado, buscando, de um jeito ou outro, relacionar uma "intenção" comunicacional com uma determinada expressão sonora. Portanto,

[*]Universidade Federal do Espírito Santo, UFES, Faculdade de Letras, PPGEL, Vitória, ES, Brasil, meirelesalex@gmail.com.

[**]Universidade Federal de Minas Gerais, UFMG, Faculdade de Letras, POSLIN, Belo Horizonte, MG, Brasil, anacrisfm@ufmg.br.

[***] Universidade Federal de Lavras, UFLA, Faculdade de Ciência da Computação, TecnoLivre, Lavras, MG, Brasil, rubens@tecnolivre.com.br.

informações físicas, linguísticas e semióticas são importantes para determinar a existência dessa relação e, se houver, grau e tipo de relação. A coleta de dados, portanto, não pode focalizar apenas um dos aspectos da produção de fala, o que aumenta, significativamente, o número de parâmetros a serem colhidos e analisados.

SetFon é um sistema de informação *Web* livre para coleta de dados em pesquisas sobre a fala. O sistema aborda o problema da coleta e gerenciamento de dados, por meio de componentes de *software*, e nasceu da necessidade de aumentar, significativamente, a quantidade de dados acústico-fonológicos, a fim de atender a demandas de estudos estatísticos de expressão da emoção e de estilo. Além disso, por ser on-line, o SetFon permite a criação de um banco de dados nacional sobre o tema, compartilhável entre pesquisadores das ciências da fala.

Este trabalho não teria sido possível sem uma equipe interdisciplinar, com preocupações bem diversas, mas objetivos comuns. Este artigo pretende explorar as diferentes facetas do trabalho. Para isto, vamos apresentar o programa, seu contexto e a tecnologia da sua produção.

DADOS EM FONÉTICA ACÚSTICA

Trabalhar com fonética sempre foi, segundo muitos estudiosos, permanecer no limiar entre o que seria e o que não seria um estudo linguístico. A situação complica-se ainda mais quando se trata de Fonética Acústica e Articulatória, dada a quantidade de conhecimentos de Física e de Anatomia Humana envolvidas nesses trabalhos. No entanto, o trabalho do foneticista linguista não ocorre desvinculado da Fonologia. Nesse caso, não se justificaria a não aceitação da fonética como estudo linguístico por parte dos fonólogos.

A nosso ver, a visão de Fonética como não integrante da ciência linguística é, em boa parte, resultado de uma inversão de prioridades, em termos de tempo de dedicação do foneticista, durante a pesquisa linguística, que pode ser dividida em levantamento de hipóteses e preparo do experimento, coleta de dados e análise. A elaboração do problema e a análise dos dados são epistemologicamente e linguisticamente embasadas, mas ocupam um terço ou menos do tempo de pesquisa do foneticista, que passa a maior parte do tempo recolhendo dados acústicos ou articulatórios dos sons de fala. O trabalho de segmentação e etiquetação das amostras de som é responsável pela maior parte dessa má interpretação do trabalho do foneticista linguista.

Infelizmente, não é pelo interesse linguístico que muitos pesquisadores, por exemplo, os engenheiros eletricistas, elaboram segmentadores automáticos de fala. Os segmentadores automáticos que vêm sendo desenvolvidos pela Engenharia, geralmente, visam à síntese ou ao reconhecimento de fala e, por isso, restringem a transcrição fonética como único elemento vinculado ao segmento fônico. Por outro lado, programas como o Praat – *software* livre analisador de fala voltado à comunidade científica – permitem o vínculo de diversos níveis de informação a segmentos de um arquivo de som, mas todo o processo de etiquetação é, necessariamente, manual (concebendo-se a etiquetação como o processo de vincular informações – como, por exemplo, transcrição fonética, taxa de elocução solicitada ao locutor, emoção relatada pelo locutor – a uma unidade sonora pré-determinada). Esses segmentadores automáticos, sem dúvida, representam grandes avanços para o linguista, mas são insuficientes para a inversão das prioridades no cronograma da pesquisa em fonética. Inversão esta imprescindível para que a pesquisa fonética no campo das tecnologias de fala alcance competitividade no cenário internacional.

SETFON: PROPOSTA DE OTIMIZAÇÃO

Para resolver esse problema, elaborou-se uma proposta de etiquetação parcialmente automática que não só atende às necessidades do foneticista linguista como também prevê a interação futura com pesquisas de reconhecimento e de síntese de fala. A proposta é o SetFon, um algoritmo para produção e organização de semioetiquetas fonológicas.

As semioetiquetas fonológicas são produto de uma ferramenta para anotação e organização de dados resultantes de análises textuais, sintáticas e semânticas, informação sobre a gravação e quaisquer outras informações relevantes para análise de dados de Fonética Acústica, como a própria etiqueta fonológica ou fonética, sempre necessária nas pesquisas da área. Essa ferramenta, o SetFon, visa agilizar o processo de preparação dos dados para a análise fonética propriamente dita, o que representa um ganho importante para a pesquisa em Fonética no Brasil, dada a originalidade da proposta nos meios internacionais de pesquisa.

O algoritmo do SetFon conjuga um arquivo de som (.wav ou outro formato) e um arquivo de texto (.txt), a fim de obter uma segmentação etiquetada com acesso às informações prosódicas como duração e curva entoacional, etiquetação – essa, passível de receber novas informações, conforme a necessidade do pesquisador – e, finalmente, retornar, em tabelas, as informações sobre cada segmento. Em virtude da natureza desse processo, do ponto de vista computacional, a ferramenta consiste num sistema gerenciador de diversos programas, cada

qual responsável por uma das tarefas necessárias ao trabalho de etiquetação de amostras de som e obtenção de tabelas. Alguns desses programas preexistiam, podendo-se citar, a título de exemplo, o Praat[1], o Ortofon (ALBANO; MOREIRA, 1996) e o SilWeb (MATTE, MEIRELES; FRAGUAS, 2006).

O SetFon visa à análise linguística da fala, sendo opcional o objetivo de suporte às pesquisas sobre síntese de fala propriamente dita. Portanto, não se trata da criação de um segmentador, mas da concepção de um algoritmo para um programa capaz de automatizar e gerenciar o vínculo de vários níveis de informação linguística e extralinguística a segmentos, cujo tamanho será determinado pela necessidade de cada pesquisa.

É importante ressaltar que, em virtude da versatilidade do SetFon, essa proposta de etiquetação é mais do que um simples programa: trata-se de montar um servidor de aplicações que pode ser facilmente adaptado às mais diversas finalidades, na pesquisa com a Fonética Experimental, devendo-se, também, ressaltar a facilidade de manutenção de seus componentes. Como o objetivo do projeto é a concepção desse algoritmo geral, a implementação da análise, inicialmente restrita à sentença, em termos de duração e curva entoacional, viabiliza a realização imediata de testes com pesquisas reais. A utilização de *corpora* reais e controlados permitiu obter resultados imediatamente utilizáveis em pesquisas fonoestilísticas do Português do Brasil (MENDES, 2009). Tratando-se de um sistema viável na web, suas aplicações na criação de tecnologias de fala são muito abrangentes, em termos de interação homem/máquina: telefonia, *Web*, eletrodomésticos, até produtos auxiliares no tratamento de patologias de fala.

DESENVOLVIMENTO DO *SOFTWARE*

A concepção do semioetiquetador fonológico seguiu o processo de desenvolvimento de software baseado em componentes, proposto por Brito *et al* (2005), sendo dividida nas seguintes etapas: (i) análise do domínio; (ii) modelagem dos componentes; (iii) implementação dos componentes; (iv) testes dos componentes; (v) implementação da interface *Web*; e (iv) teste de integração.

Durante o desenvolvimento de cada componente do Setfon, foram criadas interfaces para utilização individual (*scripts* em *shell*), tornando possível realizar testes independentes. A Programação Orientada a Componentes é uma abordagem técnica para solucionar

[1] Disponível em: <http://www.praat.org>.

problemas de forma computacional através de estruturas lógicas atômicas e bem definidas por meio de interfaces. Componentes encapsulam processos ditos de caixa preta, ou seja, processos que não exigem conhecimento aprofundado da estratégia de implementação, já que, em nível modular, não têm acoplamento. As semioetiquetas fonológicas formam um conjunto de dados complexo; para obtê-los, cada atributo foi tratado por um componente diferenciado.

O processo baseado em componentes tem grande proximidade com as atividades manuais e semiautomáticas realizadas, por pesquisadores, para obtenção de dados acústicos. A maioria das operações é atômica, e tem entradas e saídas bem definidas. Neste sentido, foram identificados 4 (quatro) componentes essenciais: (i) segmentador de som; (ii) fonotranscritor de texto tradicional para texto fonológico; (iii) manipulador de TextGrid1; e (iv) extrator de dados acústicos. Esses componentes são manipulados por uma camada *Web* que funciona tanto como controlador das etapas envolvidas no processo de extração quanto interface direta com o usuário pesquisador.

A principal ferramenta do SetFon é representada na Figura 1. Trata-se de uma interface *Web* que dá início ao processo a partir do envio de um arquivo de som de fala e um arquivo de texto (com valores semânticos correspondentes). Para se obter os dados acústicos,

Figura 1: Página principal do Setfon.

é preciso avaliar o arquivo de som com o seu respectivo TextGrid, preenchido com segmentos fonológicos e outros dados relevantes. Para se obter o TextGrid, por sua vez, são necessárias 3 (três) subetapas: (i) realizar uma transcrição do arquivo de texto para um texto fonológico e segmentá-lo com a estratégia VV; (ii) gerar um TextGrid apenas com os dados obtidos do arquivo de som, mas ainda sem os segmentos fonológicos; e (iii) inserir os segmentos fonológicos nos respectivos espaços reservados no TextGrid.

A principal estratégia para abordar a solução desse problema foi definir as entradas e saídas de cada componente como arquivos de diferentes tipos. Cada componente, portanto, recebe um ou mais arquivos de entrada e produz um arquivo resultante. A camada *Web* apresenta os arquivos (na região central da Figura 1) e as possíveis operações sobre esses arquivos (na parte inferior da Figura 1). Para realizar uma operação, é necessário selecionar os arquivos de entrada (clicando-se sobre eles) e, depois, acionar a operação desejada. Cada componente desse processo utilizou as técnicas e tecnologias mais apropriadas para o propósito.

UMA LINHA DE DESMONTAGEM: ANÁLISE DE FALA.

O SetFon trabalha como uma linha de desmontagem: o produto "fala" é desmembrado no tempo e suas qualidades analisadas e dispostas separadamente, a fim de permitir visualização das partes, antes da visualização do conjunto. Dois tipos de segmentação são necessários para a obtenção de dados de Fonoestilística: uma macrossegmentação, baseada no grupo acentual, e uma microssegmentação, baseada no VV. Embora a automação do processo tenha sido determinante na escolha dos tipos de segmentos e tenha conduzido à utilização de segmentos como a frase e a sílaba, o método de inclusão de dados foi criado tendo em vista valorizar a abordagem semiótica que considera o texto como um todo.

O ponto de partida do SetFon foi o programa SilWeb, originalmente concebido para retornar, para cada palavra e para cada sílaba, sua classificação acentual. O trabalho com UML, embora não tenha sido levado a termo, possibilitou um algoritmo enxuto e compatível com outras aplicações, algumas das quais acabaram por ser incorporadas ao programa.

Iniciada no MatLab e concluída em PHP, a programação foi feita com base nos estudos fonológicos de Mattoso Câmara (1970) e as regras foram organizadas de maneira a analisar, com 99% de acerto, qualquer palavra do Português Brasileiro (doravante, PB) ou logatoma que siga as regras fonotáticas do PB. O programa aproveitou sua estratégia de análise por caracteres e vizinhos para retornar, também, as sílabas e as máscaras consoante-vogal e consoante-vogal-semivogal.

O programa foi testado em um grande *corpus*, o CETEN-Folha, durante o período em que os 3 (três) pesquisadores envolvidos trabalharam no projeto temático "Integrando Parâmetros Contínuos e Discretos em Modelos do Conhecimento Fônico e Lexical", sob a coordenação de Eleonora Albano, com sede na UNICAMP, financiado pela FAPESP até janeiro de 2005.

O comportamento de unidades maiores do que o fone, tais como a taxa de elocução e o grupo *inter-perceptual-center* – constituído por unidades do tamanho da sílaba desde a primeira unidade após a posição acentuada até a unidade em posição acentuada seguinte (MARCUS, 1981) –, está significativamente relacionado ao comportamento tensivo potencial do texto, conforme indicam os resultados de Matte (2005), em suas pesquisas sobre fala emotiva.

A aplicação de semioetiquetas fonológicas, com a implementação do SetFon, veio ao encontro da necessidade do pesquisador de testar quais variáveis independentes implicariam variação no plano da expressão, quando sob emoção, ao invés de se ater a uma única hipótese de trabalho.

Como exemplo, podemos citar a hipótese da curva tensiva de temporalidade M (MATTE, 2004a), proposta em 2001. M resulta da combinação de 3 (três) elementos diretamente resultantes de uma análise semiótica de 5 (cinco) níveis de temporalidade no conteúdo do texto (MATTE, 2004b), dois dos quais são descartados pela fórmula de M em virtude de uma obsolescência teórica. Durante essa fase da pesquisa, o componente prosódico taxa de elocução mostrou-se significativamente correlacionado à variação de M.

O SetFon agiliza o processo de obtenção e organização dos dados de modo a permitir a realização de testes com diferentes hipóteses, por exemplo um teste com cada componente temporal isolado e em diferentes combinações, incluindo-se aqueles descartados pela hipótese original da fórmula de M. Além disso, possibilita um salto em direção a uma etapa previsível e desejável da pesquisa tendo em vista a análise semiótica do léxico, por meio de teste da relação entre palavras-chave da análise semiotemporal e os resultados prosódicos. É possível prever que a análise de um possível conteúdo tensivo do léxico, vinculada a uma análise sintática, possibilite uma automatização da análise semiótica, com vistas à síntese de fala, calcada na hipótese de caricatura vocal (MATTE, 2004a). Além de possibilitar o teste de maior número e variedade de hipóteses em menor tempo, a agilidade garantida pelo SetFon permite mudanças de estratégia sempre que os resultados apontem para isso, sem acarretar atrasos significativos à pesquisa.

O projeto pode ser dividido em 3 (três) blocos: estudo fonológico, estudo computacional do gerenciador e programação das ferramentas subsidiárias, na interface entre a Computação e a Linguística.

OS COMPONENTES

O estudo linguístico, por meio dos métodos da Fonética Experimental e da Fonologia, do comportamento da taxa de elocução, das pausas silenciosas e da curva entoacional (f0) no escopo da sentença possibilita uma segmentação calcada no sentido linguístico da prosódia da frase. A segmentação da sentença obedece ao conceito de unidade VV, iniciando, portanto, no início da primeira vogal e terminando no início da última vogal da sentença, dada a maior precisão perceptiva da transição entre consoante e vogal subsequente do que entre vogal e consoante, como comprovado em trabalhos na área (BARBOSA, 1996; CUMMINS, 2002; POMPINO-MARSCHALL, 1989).

A duração da sentença assim segmentada só pode gerar informação a respeito da taxa de elocução se as unidades utilizadas forem também unidades VV, de modo que foi necessário remodelar o programa Silweb, criado, em 2004 (MATTE; MEIRELES; FRÁGUAS, 2006), para análise acentual e decomposição fônica de palavras com transcrição fonológica e atualmente concebido para separação de sílabas CV, para uso no banco de dados do projeto temático citado acima. O uso desse programa, o SilWebVV, também permite a obtenção de dados para o cálculo do z-score da sentença, uma medida relativa da duração que leva em consideração a duração intrínseca dos segmentos fônicos, o que também será feito automaticamente.

A concepção geral do programa orientada a componentes permite que se possa vincular outros programas já existentes ao processo, incrementando o resultado final. Funciona como uma grade de texto que vincula diferentes camadas de informação a cada trecho de som-sentença. Em estrito senso, é uma rede de classes informativas de diferentes naturezas, vinculadas à mídia contínua; no caso, o som, por meio de identidades digitais.

Implementado desta forma, o semioetiquetador fonológico SetFon é passível de atualizações, algumas das quais já previsíveis, dentre as quais interessam, sobremaneira, à comunidade de pesquisa em Fonética a substituição do segmentador de sentenças por um segmentador de fones, bem como a implementação de analisadores de f0 atualizados.

Os 3 (três) blocos que organizam este projeto foram desenvolvidos conforme a necessidade; muitas vezes, simultaneamente.

ETIQUETAS FONOLÓGICAS

O conceito de semioetiqueta fonológica, aqui proposto, é uma abordagem da análise de fala que trabalha com os segmentos de som de fala como objetos. Uma semioetiqueta fonológica é uma classe de objetos cujos atributos são dados intrínsecos ou adquiridos.

Os objetos são segmentos de som de fala, que podem ter tamanhos diferentes. Adotamos, aqui, o segmento VV (vogal a vogal) (MARCUS, 1981) e o grupo acentual como base para a segmentação (BARBOSA, 2006). Esses objetos são obtidos pela análise automática de trechos de fala acompanhados de uma transcrição ortográfica (BARBOSA, 1996). O grupo acentual é uma sequência de segmentos VV obtida pela análise quantitativa e qualitativa da duração dos segmentos VV, portanto, uma análise dependente já dos atributos da original. Assim, definem-se a classe Segmento VV e a classe Grupo Acentual.

Por um lado, os dados intrínsecos são variáveis independentes, essencialmente quantitativas, e que podem ser obtidos por análise acústica automática. Já os dados adquiridos são parâmetros cuja automatização ainda é uma possibilidade pouco explorada, dada sua dependência de uma análise qualitativa. Atualmente, é possível dispor de *parsers* sintáticos e semânticos para auxiliar o processo, mas a análise semiótica é totalmente manual.

Tanto o Grupo Acentual quanto o segmento VV podem receber atributos intrínsecos e adquiridos, embora de naturezas diferentes. O primeiro atributo adquirido do segmento VV é uma etiqueta fonológica, obtida pela transcrição fonológica e segmentação do texto correspondente a um som de fala. Somente com a obtenção desta etiqueta fonológica é possível calcular seus atributos intrínsecos (duração, intensidade, frequência, configuração formântica) e criar os objetos da classe Grupo Acentual. Já os objetos desta última possuem atributos intrínsecos da ordem do prosódico, tais como taxa de elocução, curva melódica variação de intensidade, variação de duração, posição do acento e número de segmentos VV. Os atributos adquiridos são totalmente dependentes do tipo de resultado almejado, podendo advir de *parsers* sintáticos, *parsers* semânticos e/ou de análises de conteúdo específicas como análise semiótica tensiva ou narrativa, só para citar alguns exemplos.

ORTOSIL

Conforme descrito em Matte, Meireles e Fraguas (2006), elaboramos um analisador fonológico silábico-acentual para aplicações linguísticas: o SilWeb. Resumidamente, esse programa retorna, a partir de uma entrada fonológica, as seguintes informações lexicais: 1) classe acentual da palavra; 2) número e tipo de sílaba (tônica, pré-tônica e pós-tônica); e 3)

máscaras silábicas (com ou sem a presença de semivogais). Um exemplo dessa análise é apresentado na Figura 2, abaixo.

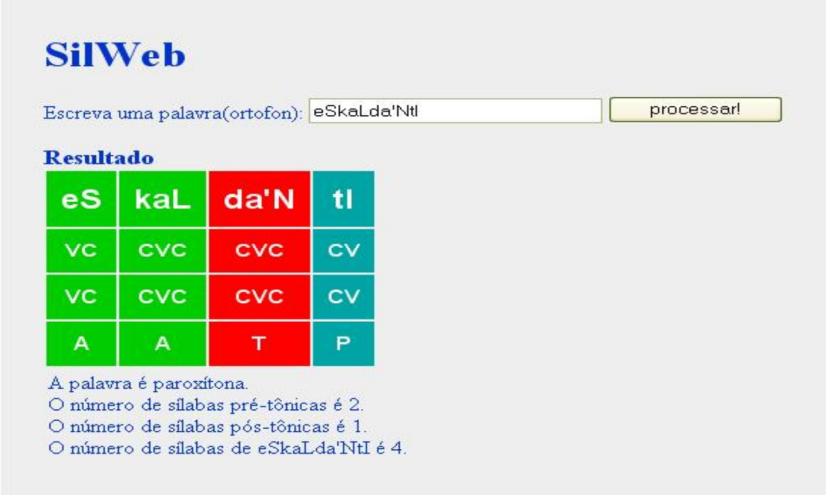

Figura 2: Análise fonológico silábico-acentual de uma palavra escrita em transcrição "ɘOrtofon" (MATTES; MEIRELES; FRAGUAS, 2006, p. 47).

Na Figura 2, podemos notar que o pesquisador digita a palavra transcrita em "Ortofon" (eSkaLda'NtI) e o programa retorna as informações linguísticas pré-programadas no código-fonte. Esse tipo de transcrição (conversão letra-fone) foi proposto por Albano & Moreira (1996), para fins de síntese de fala, e era efetuada pelo programa Ortofon, de uso restrito.

Sendo assim, apesar de o SilWeb gerar informações linguísticas, para que linguistas e demais interessados em linguística de *corpora* possam usufruir dos benefícios do programa, os mesmo deveriam saber transcrever em "ortofon", o que dificulta, bastante, a aplicação prática do programa. Assim, com o intuito de facilitar a utilização do programa para a comunidade científica, elaboramos um programa próprio para conversão de dados ortográficos em fonológicos: Ortosil. Essa ferramenta computacional segue uma linha teórica similar, mas independente, dos princípios do Ortofon.

O Ortosil surgiu a partir de nossa experiência com o Ortofon (ALBANO; MOREIRA,*ibidem*); no entanto, como o nosso intuito era elaborar um programa que refletisse o conhecimento fonológico do Português, baseamos nossa transcrição, mais precisamente, na análise fonológica proposta por Mattoso Câmara[2]; porém, com algumas modificações fundamentais.

[2] Sobre o porquê da utilização da análise fonológica de Mattoso Câmara, vide Matte, Meireles e Fraguas (*ibidem*).

De acordo com Mattoso Câmara Jr. (1970), o sistema fonológico do Português é composto pelos seguintes fonemas: 19 (dezenove) consoantes, 7 (sete) vogais orais e 2 (dois) arquifonemas. A partir desse quadro fonêmico, transcreve-se qualquer palavra do PB. Apresentaremos, a seguir, uma análise comparativa de nossa análise fonológica com a de Mattoso Câmara Jr.

Análise das consoantes:Mattoso Câmara Jr. propõe 19 (dezenove) fonemas consonantais para o PB: /p b t d k g f v s z ʃ ʒ m n ɲ l ʎ ɾ R/.Todos esses fonemas ocorrem em início de sílaba e, portanto, possuem pouca variabilidade articulatória (vide entre outros, TAUROSA, 1992; KEATING et al*,* 1999). Nossa transcrição para esses fonemas é idêntica a essa análise, exceto por modificações simbólicas relacionadas a maior facilidade de implementação computacional, a saber: /p b t d k g s z sh zhm n nh l lh R/. Além disso, optamos por representar a vibrante simples [ɾ] com o mesmo símbolo do arquifonema /R/, por ser uma de suas possíveis pronúncias.

Como podemos notar, o quadro fonêmico consonantal para o contexto de início de sílaba e/ou palavra (exceto o fonema /r/), devido à questão da estabilidade articulatória, é incontroverso. No entanto, há grande variação dialetal na pronúncia de consoantes em final de sílaba e/ou de palavra, no Português Brasileiro. Para explicar essa variação, Mattoso Câmara Jr. fez uso da noção clássica de arquifonema do estruturalismo russo.

Arquifonemas consonantais: Segundo Trubetzkoy (1939), arquifonemas são símbolos que representam a perda de contraste fonêmico em determinados contextos fonéticos. Sua representação é feita em maiúscula pelo fonema não-marcado. Por exemplo, [s z] são fonemas do Português, pois encontramos pares mínimos como "saca" e "Zaca", em que a troca de um pelo outro gera mudança de significado. No entanto, esse contraste fonêmico é neutralizado em posição de final de sílaba ou palavra. Consideremos a palavra "mas". Alguns dialetos a pronunciam como [mas]; outros, como [maz], o que indica perda da distinção fonêmica, nesses contextos. Sendo assim, o conceito de arquifonema foi introduzido para representar casos como esse. Mattoso Câmara (1970, p. 42), baseado nessa argumentação, considera, pois a ocorrência dos seguintes arquifonemas no PB. /S N/. O /S/ lida com a variação entre /s z ʃ ʒ/ e o /N/ lida com a variação entre /m n ɲ/. Da mesma forma, nossa transcrição é feita utilizando exatamente esses símbolos. Ex.: "pasta" é transcrita como /pa'StA/[3]. Além dessas consoantes pós-vocálicas, há outras consoantes também instáveis que, no entanto, não se

[3] Em nossa transcrição, o acento sempre vem após a vogal, mesmo se a sílaba for fechada.

encaixam na definição de arquifonema. A esses sons dá-se o nome de arquisegmento; ou seja, um segmento não-especificado (vide ARCHANGELI, 1988).

Arquisegmentos consonantais: Arquisegmentos, conforme também apontado por Albano e Moreira (1996), são elementos subespecificados fonologicamente; ou seja, sua realização fonética varia de acordo com características dialetais, sociais ou mesmo individuais da linguagem. No entanto, diferentemente desses autores, consideramos um número muito maior desses elementos fonológicos, a saber: /P B T D K G R L/. Todos esses arquisegmentos são necessários para representar variações na fala erudita do Português. Por exemplo, a palavra "pacto" pode ser pronunciada num contínuo entre [paktʊ] e [pakɪtʊ][4]. Essa palavra, por exemplo, é transcrita como /pa'KtO/.

Análise das vogais:De acordo com a análise de Mattoso Câmara, temos: (a) 7 (sete) fonemas vocálicos orais (/a ɛ e i ɔ o u/), (b) 5 (cinco) fonemas vocálicos orais em posição pré-tônica e/ou pós-tônica medial (/a e i o u/) e (c) 3 (três) fonemas vocálicos orais em posição pós-tônica final (/a i u/). Nossa análise considera o mesmo conjunto de fonemas para as 2 (duas) primeiras posições (a, b), mas não para a terceira (c). Nesse contexto, consideramos a ocorrência de arquifonemas vocálicos.

Arquifonemas vocálicos: Da mesma forma que as consoantes, consideramos a existência de arquifonemas vocálicos, no PB, em sílaba pós-tônica final, a saber: /E O/. Nesse contexto, não há mais a distinção fonológica entre os fonemas /e i/ ou /o u/. Quanto ao /a/, no mesmo contexto, consideramos a existência de um arquisegmento variando entre [a] e [ɐ], o qual é representado por /A/. Semelhantemente, em nossa análise, as semivogais /i u/ são representadas pelos arquisegmentos /I U/.

A Tabela 1, abaixo, representa todas as possibilidades de transcrição pelo programa "Ortosil", descrito acima, com exemplos do PB, e comparando com a análise de Mattoso Câmara(1970).

[4] Consideramos, contudo, que, provavelmente, mesmo em pronúncias como [paktʊ], deve haver uma vogal [ɪ] intrusiva que, devido à sua curta duração, não é percebida pelos falantes nativos. Além disso, acreditamos que pronúncias com uma suposta oclusiva em final de sílaba não deva ocorrer em falantes iletrados ou de baixa escolaridade. Obviamente, experimentos necessitam ser feitos, a fim de que se possa confirmar ou não essas hipóteses.

TABELA 1 - COMPARAÇÃO DA ANÁLISE FONOLÓGICA DE MATTOSO CÂMARA COM A DO ORTOSIL

Mattoso	Ortosil	Exemplo	Mattoso	Ortosil	Exemplo
p	p ou P	/pa'/; /a'PtO/	ɲ	nh	/nhoh'kE/
b	b ou B	/be'/; /aBdika'R/	l	l ou L	/la'/; /ma'L/
t	t ou T	/ta'/; /eh'TnikO/	ʎ	lh	/lha'mA/
d	d ou D	/da'/; /aDmini'StRA/	r	r ou R	/ra'tO/; /po'R/
k	k ou K	/ka'/; /pa'KtO/	ɾ	R	/ka'RA/
g	g ou G	/aga'/; /aGnoh'StIkO/	N	N	/masa'N/
f	F	/feh'/	S	S	/pa'S/
v	V	/ve'/	a	a ou A	/a'/; /pa'tA/
s	S	/seh'/	ɛ	eh	/eh'/
z	Z	/ze'RO/	e	e ou E	/pe'/; /a'RtE/
ʃ	Sh	/sha'/	i	i ou I	/pi'/; /pa'I/
ʒ	Zh	/zha'/	ɔ	oh	/soh'/
m	M	/ma'/	o	o ou O	/avo'/; /a'vO/
n	N	/noh'/	u	u ou U	/tu'/; /vo'U/

ALGORITMO COMPUTACIONAL DO ORTOSIL

O algoritmo computacional do Ortosil segue os mesmos princípios teóricos do Silweb (descrito acima). Toda a análise lexical é feita, sílaba a sílaba, com base no esquema de sílaba como ataque, núcleo e coda (vide MATTE; MEIRELES; FRAGUAS, 2006, p. 42). Antes da análise, contudo, transformávamos todas as letras em minúsculas (função "tolower" da linguagem C) e, depois, fazíamos a acentuação na palavra ortográfica (função "stress", de nossa autoria).

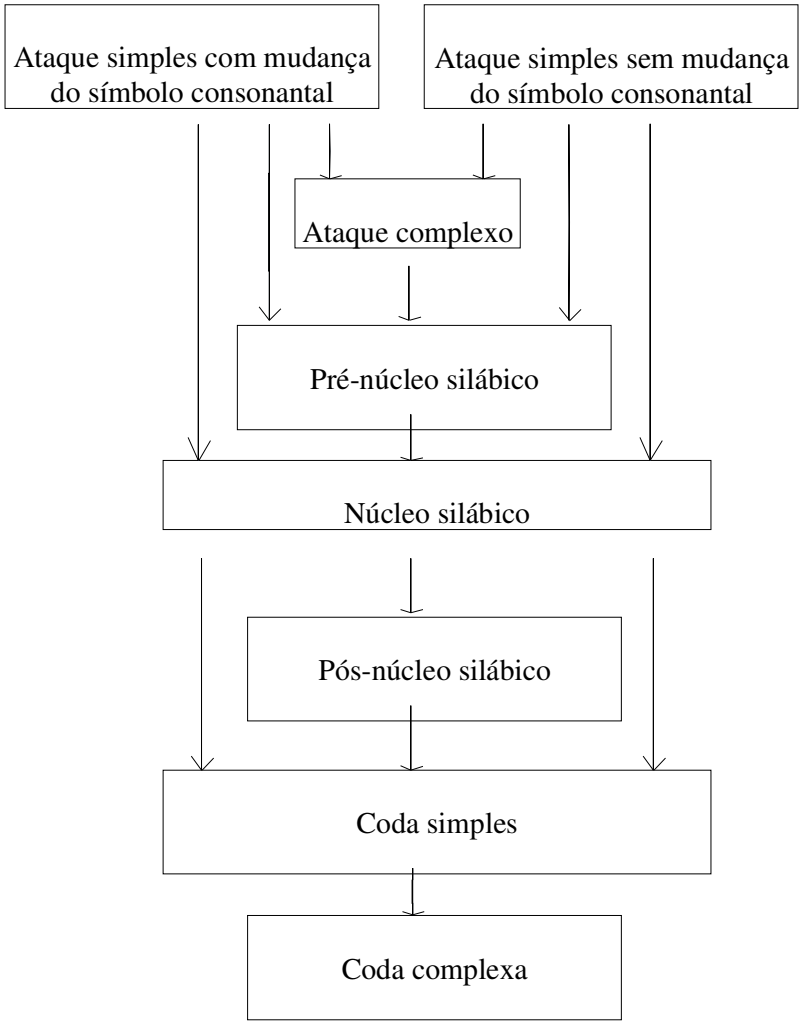

Figura 3: Algoritmo computacional do programa Ortosil.

A acentuação ortográfica das palavras tomou como base os algoritmos para detecção de classe acentual (oxítona, paroxítona, proparoxítona) das gramáticas normativas do PB. Analisemos alguns desses casos: 1) acento final: se a palavra termina em "r" e não tem acento gráfico, é oxítona. Ex.: partir, amor; 2) acento penúltimo: se a palavra termina em "o", tem mais de uma sílaba e não apresenta acento gráfico, é paroxítona. Ex.: pato, maroto; 3) acento antepenúltimo: se a palavra tem acento gráfico na antepenúltima sílaba, é proparoxítona. Ex.: parábola, abóbora.

Após essa primeira etapa de tratamento de dados, começa a análise silábica, que gerará a transcrição "ortosil". O algoritmo computacional que representa essa análise é descrito na Figura 3. Só não está descrito nessa figura o *loop* que ocorre após o núcleo silábico (ou o pós-núcleo silábico, ou a coda simples ou a coda complexa), reiniciando o tratamento silábico até o final da palavra.

ETAPAS DA LINHA DE DESMONTAGEM

O SetFon pode ser definido como uma linha de desmontagem: entra som e respectivo texto e saem informações analíticas dos componentes acústico, fonético-fonológico, prosódico e do conteúdo verbal. Suas etapas são:

- *analisador de som*: segmentação automática das sentenças. Essa segmentação é, atualmente, feita usando-se o programa *freeware* Praat, para o qual foi concebido um segmentador automático de unidades do tamanho de uma sílaba, de autoria do professor Plínio Almeida Barbosa (2004), com a colaboração de Meireles e participação de Matte na fase de *debug*. O programa, além de delimitar o tamanho dos segmentos VV e dos grupos acentuais, permite salvar amostras desses sons em arquivos independentes, criando um banco de dados passível de recuperação para outros tipos de análises. A informação sobre a duração da sentença em milissegundos é obtida nessa etapa, bem como informações sobre instantes de início e fim de sentença, para uso pelo anotador de mídia contínua. A possibilidade de audição da sentença, do meio da transição entre ela e a anterior até o meio da transição para a próxima, ainda não implementada, deve ser uma possibilidade presente em todas as visualizações das anotações.
- *etiqueta fônica*: Passar as sentenças no Fonotranscritor, analisador de texto que obtém a transcrição fonológica, máscaras Consoante-Vogal, com ou sem distinção de semi-vogal, tonicidade de palavras e de unidades do tamanho da sílaba e o número de VVs; cada uma dessas informações é cadastrada no banco de dados, podendo constituir classe ou subclasse, conforme o tipo de informação. Como o SetFon trabalha, nesse momento, com referência ao segmento do tamanho da sentença, as classes referem-se a informações com relação de um para um; ou seja, uma informação para cada sentença, enquanto as sub-classes contém mais de uma informação para cada sentença (uma identidade para cada palavra, por exemplo, vinculadas a uma única sentença).
- *z-score*: A partir das informações obtidas pelo analisador de som (camada de duração) e pelo analisador de texto (transcrição fonológica), calcular o z-score por sentença e criar a camada correspondente (BARBOSA, 2004).

- *taxa de elocução*: a partir das informações da camada de número de VVs por sentença e de duração, calcular e criar a camada de taxa de elocução (MEIRELES, 2001; MATTE, 2004a).
- *picos e médias de picos de derivadas de f0*: Essa entrada é, por enquanto, feita usando-se o programa Praat, para testar esse componente do som como fator para análise da taxa de elocução e para análise da emoção na fala. O analisador de som; no caso, o Praat, vai calcular e vincular os picos e as médias dos picos de derivadas de F0 a cada sentença. Trata-se de 3 (três) classes de informação: o número de picos, os valores dos picos e a média dos picos para cada sentença. Essa abordagem dá preferência à análise da dinâmica da curva entoacional por meio de informações da derivada no lugar da utilização de valores absolutos de F0 (BALLY; HOLM, 2002).
- *anotação semiautomática de dados não acústicos, nem fonéticos, nem fonológicos*: esta anotação é feita numa interface web a qual permite incluir informações extra, cujo tipo depende do interesse do pesquisador. Por exemplo: resultados de análises semióticas (tipo de manipulação, valor do objeto, tensividade, aspectualização, etc.), sexo do informante, contexto de obtenção de dados (instruções para o informante ou outros dados relevantes), dentre outros. O pesquisador indica a categoria de dados e insere a informação para cada grupo acentual, podendo repetir, automaticamente, a mesma informação para um conjunto de grupos acentuais ou especificar um a um, conforme a necessidade.
- *tabelas*: todos os resultados, por sua organização no banco de dados, são passíveis de recuperação em tabelas (arquivos CSV – *comma-separated values*) para análise estatística, cabendo ao pesquisador-usuário definir as informações que deseja antes mesmo do início do processo, para que tarefas desnecessárias não impliquem num indesejável aumento no custo computacional (tempo de execução). Além disso, quando o programa envia o *e-mail* de aviso de processo completado, envia, junto, uma breve análise estatística descritiva do *corpus*, obtida com funções do programa R[5].

Segmentador de som

O segmentador de som é um componente do sistema, implementado com a linguagem Praat, que recebe, como entrada, um arquivo de som, para produzir um TextGrid semipreenchido. O código foi baseado no *script* intitulado "BeatExtractor", de autoria de

[5] Disponível em: <http://cran.r-project.org/>.

Plínio A. Barbosa. Esse componente recebe uma configuração (por exemplo: sexo do falante, tipo de filtro a ser utilizado e tipo de técnica a ser utilizada) e identifica os pontos de segmentação VV, produzindo um arquivo TextGrid contendo os intervalos de tempo entre cada segmento e outros dados relevantes, conforme exemplo disponível na Figura 4. O componente tem uma pequena camada implementada com a linguagem PHP[6], que faz a comunicação entre a interface *Web* e o *script* em Praat.

Para ajustar as possíveis configurações de segmentação, a interface *Web* oferece uma ferramenta para criar e editar um arquivo INI, que armazena as possíveis configurações a serem utilizadas. Desta forma, o componente pode receber, opcionalmente, um arquivo de configurações como entrada, além do arquivo de som, que é obrigatório. Definir um arquivo de configurações é útil para reutilização por diferentes análises ou por diferentes instâncias do sistema (em servidores diferentes). A utilização de *scripts* na interface gráfica do Praat agiliza, sobremaneira, o processamento manual do sinal, mas implica realizar a operação, arquivo a arquivo, tornando o processo todo bem mais demorado. A primeira implementação do *script* de Barbosa foi via *shell script*, com significativo aumento na performance da coleta de dados, e constituiu a primeira versão do componente de obtenção de dados intrínsecos do Setfon.

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 5.9780045351473925
tiers? <exists>
size = 1
item []:
item [1]:
class = "IntervalTier"
name = "vv"
xmin = 0
xmax = 5.9780045351473925
intervals: size = 8
intervals [1]:
xmin = 0
xmax = 1.6456602147085324
            text = ""
intervals [2]:
xmin = 1.6456602147085324
xmax = 2.04477534539712
text = "O_n"
intervals [3]:
xmin = 2.04477534539712
xmax = 2.365412781849187
text = "om"
```

Figura 4: Exemplo de TextGrid feito no Praat; dois primeiros segmentos VV do texto "O nome da fruta".

[6] Disponível em: <http://www.php.net>.

Etiquetas fonológicas

O preenchimento de etiquetas fonológicas no TextGrid (produzido pelo segmentador de som) é realizado por 2 (dois) componentes: o fonotranscritor e o manipulador de TextGrid, ambos implementados, exclusivamente, com a linguagem PHP. O fonotranscritor foi totalmente baseado nas regras do Ortosil. A única diferença é que a transcrição da forma ortográfica para a forma fonológica é intermediada por uma transcrição fonética, também automática.

Primeiramente, o *fonotranscritor* é responsável por converter um texto natural em segmentos VV com escrita fonológica, passando por 3 (três) etapas: a) conversão ortográfico → fonético; b) conversão fonético → fonológico; e c) conversão frase → VV. Por exemplo, a palavra "complexo" (texto natural) é transcrita como "kõplɛˈkso" (que utiliza os símbolos fonéticos definidos pelo IPA - *International Phonetic Alphabet*); em seguida, é transcrita para "koNple'kso" (texto fonológico) e, finalmente, é dividida nos segmentos: /oNpl/ /e'ks/ /o/ (segmentos VV). O arquivo resultante desse processo é um arquivo texto, em que cada linha guarda um segmento VV.

Para realizar a transcrição inicial, foi utilizado um algoritmo, baseado em expressões regulares, que avalia uma palavra, trecho a trecho e, de acordo com uma tabela de regras gramaticais e uma lista de exceções, símbolos de algum alfabeto são convertidos em símbolos do IPA. A tabela de regras gramaticais e a lista de exceções são especificadas em um arquivo separado, já que dependem do alfabeto utilizado para representar o texto. As demais operações são relativamente simples e diretas.

Já o componente manipulador de TextGrid tem um *parser*, um manipulador e um gerador de TextGrid. Desta forma, o componente é capaz de ler o TextGrid semipreenchido gerado pelo segmentador de som e de incluir os segmentos fonológicos produzidos pelo fonotranscritor, gerando, assim, um TextGrid totalmente preenchido.

O TextGrid, originalmente, era produzido como parte da segmentação do som, sem qualquer dado qualitativo. A inserção das etiquetas fonológicas era feita, manualmente, pelo pesquisador. Inicialmente, produzimos um Praat *script* para inserção dos dados, dependente de interface gráfica e restrito a rodar um arquivo por vez. O SetFon reescreveu o código em PHP e automatizou o processo, dispensando a interface gráfica do Praat, o que também contribuiu, significativamente, para a melhoria do desempenho da coleta de dados.

Extrator de dados acústicos

O extrator de dados acústicos é um componente responsável por avaliar um arquivo de som e, com auxílio de um arquivo TextGrid completo, obter os dados acústicos, que são, então, armazenados em um arquivo CSV.

Esse componente foi escrito, essencialmente, com a linguagem Praat, e tem uma camada implementada em PHP, para comunicação com a interface *Web*. O algoritmo utilizado para extração dos dados é baseado no *script* "SGdetector", de autoria de Plínio A. Barbosa, adaptado, por Ana C. F. Matte, para a obtenção de maior número de variáveis. O *script* também foi adaptado para resultar em um comando sql, a fim de possibilitar a imediata inclusão dos resultados em um banco de dados.

```
INSERT INTO `segmentos` (`idSeg` , `arquivo` , `segmento` , `tIni` , `tFim` , `dur` , `f1media` , `f1mediana` , `f1dp` , `f2media` , `f2mediana` , `f2dp` , `f3media` , `f3mediana` , `f3dp` , `f4media` , `f4mediana` , `f4dp` , `zScore` , `zSuavisado` , `posicao` , `tamanho` ) VALUES
('1','Monicalit1_2-2','Og','72.91081323201482','73.05954890102598','0.17183238822112457','451','420','59.70','1739','1428','535.76','2674','2518',"f3#dp:2",'3819','3762','211.73','-4.212','0','-6','7'),
('2','Monicalit1_2-2','al','72.91081323201482','73.05954890102598','0.17685845326224958','451','420','59.70','1739','1428','535.76','2674','2518','299.50','3819','3762','211.73','-3.818','0','-5','7'),
```

Figura 5: Trecho inicial de resultado obtido pelo extrator de dados acústicos (SANTOS, 2008). Dados: nome do arquivo de entrada, transcrição do segmento, tempo inicial e final do segmento no arquivo, duração do segmento, dados sobre os quatro primeiro formantes, análise da duração relativa e informação sobre a posição no grupo acentual e tamanho deste último. Além disso, em outro *script*, são obtidos dados sobre intensidade, taxa de elocução e curva melódica para o grupo acentual.

Etiquetas semióticas

A inserção de etiquetas semióticas ainda é feita diretamente pelo pesquisador, por meio de uma interface web que organiza os dados como atributos da classe Grupo Acentual. A interface possibilita a replicação automática de entradas e a criação de diferentes atributos, conforme o tipo de análise. Embora tenha sido criada para a inserção de dados provenientes de análises semióticas, a possibilidade de criação de novos atributos confere à ferramenta maleabilidade suficiente para a inserção de qualquer tipo de dado qualitativo, para fins de análise estatística paramétrica ou não. Entraria nessa categoria de dado, por exemplo, orientação dada ao informante para produzir uma fala mais neutra ou mais emotiva, mais rápida ou mais lenta, etc., ao que chamamos de informações circunstanciais sobre a coleta.

Nesse sistema, podem ser acrescentadas, conforme o objetivo da análise, classificações do *corpus*, como trechos de noticiário jornalístico com tema internacional, local ou regional, etc. (MENDES, 2009).

CONCLUSÃO

O sistema, registrado no Sourceforge.net[7] como GPL v. 2, atingiu os objetivos esperados, possibilitando aumento significativo na coleta de dados para estudos fonoestilísticos. Vários processos, usualmente manuais, passaram a ser automatizados e padronizados, com a utilização do Setfon.

Embora o SetFon signifique um passo importante para o campo dos estudos da fala, existem várias melhorias que ainda podem ser exploradas, tais como: incorporar um componente de reconhecimento de fala (permitindo a geração do texto a partir do arquivo de som), oferecer um recurso que possibilite interação assíncrona (para minimizar a conectividade entre cliente/servidor, durante o processamento), criar um *framework* que ofereça recursos específicos de estudos da fala e oferecer as funcionalidades do SetFon sob uma arquitetura de *Web Service.*

[7] Disponível em: <http://www.sourceforge.net/projects/setfon>.