

# Setfon: Sistema Open Source para produção e organização de Semioetiquetas Fonológicas\*

Ana Cristina Fricke Matte<sup>1</sup>, Rubens Takiguti Ribeiro<sup>2</sup>

<sup>1</sup>Faculdade de Letras – Universidade Federal de Minas Gerais (UFMG) -  
Av. Antônio Carlos, 6627 – 31.270-970 – Belo Horizonte – MG – Brasil

<sup>2</sup>TecnoLivre – Cooperativa de Tecnologia e Soluções Livres  
Universidade Federal de Lavras (Ufla) – 37.200-000 – Lavras – MG – Brasil

*anacrifm@ufmg.br, rubens@tecnolivre.com.br*

**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 up the acoustic-phonological data quantity in order to solve the problem of statistic significance in expression of emotion in speech studies. Moreover, Setfon is on-line and, then, allows the creation of a national database about this matter, that can be easily shared between the researchers in speech sciences.*

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

## 1. INTRODUÇÃO

O presente trabalho discorre sobre o desenvolvimento de um software livre para auxiliar pesquisas relacionadas à fonoestilística e a outros estudos da fala, o Setfon. O problema central é a interdisciplinaridade: não é viável trabalhar com fonoestilística, em especial expressão da emoção na fala, sem ultrapassar os limites dos estudos linguísticos dos sons de fala, compreendidos pela fonética e pela fonologia. O processamento do Setfon, portanto, visa a elaboração de tabelas para análise estatística com a obtenção de conjuntos de dados aos quais chamamos semioetiquetas fonológicas.

Pesquisas nesta área são sustentadas por avaliações estatísticas resultantes do cruzamento de dados intrínsecos ao som (como os dados acústicos) e dados adquiridos (avaliações semióticas, normalmente relacionadas à cultura do falante ou a circunstâncias da coleta de amostras de fala). Esta vasta gama de variáveis envolvidas implica um *corpus* de magnitude muito superior à usualmente alcançada [Matte 2004].

Assim, automatizar o processo de coleta de dados é uma etapa imprescindível para via-

\* Projeto financiado pelo CNPq, processo número 473616/2007-6.

bilizar um estudo científico da emoção na fala. Estudos [Matte 2006] [Barbosa 2006] mostram que a utilização de softwares desenvolvidos por ou com pesquisadores da área, contexto comum no desenvolvimento de software livre, produzem resultados sensivelmente mais práticos para a realização de pesquisas nessa área, quando comparados com softwares proprietários. Neste artigo vamos apresentar a parte conceitual que sustenta as análises do Setfon no campo da linguística e as soluções computacionais adotadas para obtenção de dados.

## 2. SEMIOETIQUETAS 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, define-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 até o momento, 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.

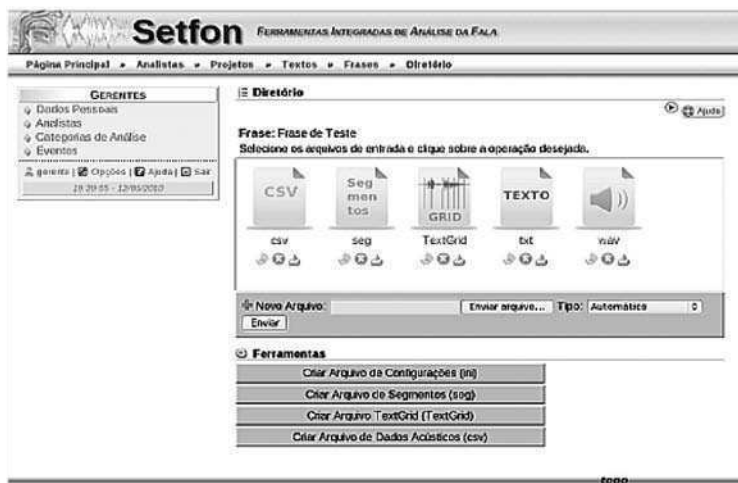
Tanto o Grupo Acentual quanto o segmento VV podem receber atributos intrínsecos e adquiridos, embora de naturezas diferentes. O primeiro atributo adquirido do segmento VV é uma etiqueta fonológica, obtida pela transcrição fonológica e segmentação do texto correspondente a um som de fala. Somente com 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.

## 3. ETIQUETADOR FONOLÓGICO

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 (vi) 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 problemas de forma computacional através de estruturas lógicas atômicas e bem definidas através de interfaces. Componentes encapsulam processos ditos de caixa preta, ou seja, processos que não

exigem conhecimento aprofundado da estratégia de implementação, já que, a nível modular, não possuem acoplamento. As semiotiquetas 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 semi-automáticas realizadas por pesquisadores para obtenção de dados acústicos. A maioria das operações são atômicas, e possuem entradas e saídas bem definidas. Neste sentido, foram identificados 4 componentes essenciais: (i) segmentador de som; (ii) fonotranscritor de texto tradicional para texto fonológico; (iii) manipulador de TextGrid<sup>1</sup>; (iv) extrator de dados acústicos. Estes 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.



**Figura 1: Interface da ferramenta principal do Setfon**

A principal ferramenta do Setfon é representada na Figura 1. Ela funciona 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, 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 três sub-etapas: (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 estratégia adotada para abordar a solução deste problema foi definir as entradas e saídas de cada componente como sendo 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 estes arquivos (na parte inferior da Figura 1). Para realizar uma operação, é necessário selecionar os arquivos de entrada (clitando sobre eles) e, depois, acionar a operação desejada. Cada componente deste processo utilizou as técnicas e tecnologias mais apropriadas ao propósito.

<sup>1</sup> Estrutura computacional responsável por representar os limites entre os segmentos de um arquivo de som, e que pode agregar anotações, como as transcrições fonológicas dos segmentos.

### 3.1. Segmentação do Som

O segmentador de som é um componente do sistema, implementado com a linguagem Praat<sup>2</sup>, que recebe como entrada um arquivo de som para produzir um TextGrid semi-preenchido. O código foi baseado no *script* intitulado “BeatExtractor”, de autoria de Plínio A. Barbosa. Este 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 na Figura 2. O componente possui uma pequena camada implementada com a linguagem PHP<sup>3</sup>, que faz a comunicação entre a interface Web com 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 em 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.

### 3.2. Etiquetas Fonológicas

O preenchimento de etiquetas fonológicas no TextGrid (produzido pelo segmentador de som) é realizado por dois componentes: o fonotranscritor e o manipulador de TextGrid, ambos implementados exclusivamente com a linguagem PHP. Primeiramente, o fonotranscritor é responsável por converter um texto natural em texto fonológico. Por exemplo, a palavra “complexo” (texto natural) é transcrita para “kõple'kso” (que utiliza os símbolos fonéticos definidos pelo IPA – *International Phonetic Alphabet*), em seguida transcrita para “koNple'kso” (texto fonológico) e, finalmente, dividida nos segmentos: /oNpl/ /e'ks/ /o/ (segmentos VV). O arquivo resultante deste processo é um arquivo texto, onde cada linha guarda um segmento.

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 para 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 possui 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 incluir os segmentos fonológicos produzidos pelo fonotranscritor, gerando, assim, um TextGrid totalmente preenchido. O TextGrid original faz parte da segmentação do som, sem qualquer dado qualitativo. A inserção das etiquetas fonológicas, que era

<sup>2</sup> Software livre destinado à manipulação e extração de dados de arquivos de som. Possui tanto uma interface de acesso por pesquisadores quanto uma linguagem de *script* que permite automatizar operações que poderiam ser realizadas pela interface tradicional. Disponível em <<http://praat.org/>>.

<sup>3</sup> Linguagem de *script* open source de uso geral e especialmente guarnecida para desenvolvimento Web, disponível em <<http://php.net/>>.

feita manualmente pelo pesquisador, passou a ser feita por um *praat script*, 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.

```
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 2: Exemplo de TextGrid.

### 3.3. Extração 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 (*comma-separated values*). Este componente foi escrito essencialmente com a linguagem Praat, e possui uma camada implementada em PHP para comunicação com a interface Web. O algoritmo utilizado para extração dos dados é baseado em um *script* intitulado "SGdetector", de autoria de Plínio A. Barbosa e adaptado por Ana C. F. Matte para obtenção de maior número de variáveis.

### 3.4. 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. Criada para a inserção de dados provenientes de análises semióticas, a interface confere à ferramenta maleabilidade suficiente para a inserção de qualquer tipo de dado qualitativo para fins de análise estatística paramétrica ou não, por exemplo, informações circunstanciais sobre a coleta e/ou classificações do *corpus* [Mendes 2009].

## 4. CONCLUSÃO<sup>4</sup>

O sistema, registrado no Sourceforge.net<sup>5</sup> como GPL v.2, atingiu os objetivos esperados, possibilitando um aumento significativo na coleta de dados para estudos fonostilísticos. Vários processos usualmente manuais passaram a ser automatizados e padronizados, com a utilização do Setfon.

Embora o sistema 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 uma 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*.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Barbosa, P. A. (1996) "At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation" In: *Cadernos de Estudos Lingüísticos*, 31, 1, Brasil.
- Barbosa, P. A. (2006) "Incurções em torno do Ritmo da Fala", Campinas: Pontes Editores, v.1, 1.
- Brito, P. H. da S.; Barbosa, M. A. M.; Guerra, P. A. de C. & Rubira, C. M. F. (2005) "Um Processo para o Desenvolvimento Baseado em Componentes com Reuso de Componentes", Brasil.
- Marcus, S. M. (1981) "Acoustic determinants of Perceptual –Center (p-center) location" In: *Perception and Psychophysics*, 30, 3.
- Matte, A. C. F. (2004) "Relating emotional content to speech rate in Brazilian Portuguese" In: SP2004 CD-ROM Proceedings , *Speech Prosody 2004*, Nara, Japão.
- Matte 2006: Matte, A. C. F. , SL: inclusão social e pesquisa de ponta na Letras da UFMG, In: *Programação do 7.o FISL*, Porto Alegre, 2006.
- Mendes, C. (2009) "A Expressão e o Conteúdo da Fala do Jornal Nacional", Programa de Pós-Graduação em Estudos Lingüísticos, Universidade Federal de Minas Gerais (UFMG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

<sup>4</sup> Agradecimentos: Adelma Araújo, Alexandro Meireles, Cecilio C. Fraguas, Conrado Mendes, Daniervelin Pereira, Leonardo Amaral, Marcos P. Rebello, Mônica Santos e Plínio A. Barbosa.

<sup>5</sup> <http://www.sourceforge.net/projects/setfon>