# PROPOSTA TEÓRICA PARA SELEÇÃO ALGORÍTMICA DE ENTONAÇÃO VOCAL BASEADA EM CORRELAÇÃO E SIMILARIDADE

Theoretical proposal for algorithm selection of vocal intonation based on correlation and similarity

**RIBEIRO, Anderson da Silva**
Faculdade Politécnica de Campinas

**MOREIRA, Juliana da Costa**
Faculdade Politécnica de Campinas

**OLIVEIRA, Vanessa Hipólito de**
Faculdade Politécnica de Campinas

**MENDELECK, André**
Faculdade de Jaguariúna

**Resumo:** Neste artigo é apresentado um modelo teórico para seleção de entonação de voz a partir de uma base de dados previamente conhecida, cujas informações são fornecidas pelo sistema de conversão texto-fala [5]. O objetivo é criar um modelo teórico funcional para selecionar um padrão de entonação interrogativo de uma frase, a ser aplicado em frases afirmativas. A metodologia utilizada baseia-se na avaliação da correlação entre a estrutura gramatical do texto de entrada e a estrutura gramatical das sentenças presentes na base de dados, aplicando-se um modelo estruturado semelhante ao implementado em compiladores.

**Palavras-chave:** entonação, compilador, padrão interrogativo, regras de ponderação.

**ABSTRACT:** This paper shows a theoretical model to select a voice intonation from one previously established database, this database contains information from the text-to-speech conversion systems. The purpose is to create the theoretical and functional model to find the intonation pattern for interrogative sentences. The applied methodology is based on the evaluation of the correlation between the grammatical structure of the input text and the

grammatical structure of sentences in the database, applying a structured model for compilers implementation

**Key-words:** intonation, compiler, question pattern, weighting rules

## 1. Introdução

Este artigo refere-se à criação de um modelo teórico funcional que tem como objetivo selecionar um padrão de entonação interrogativo de uma frase gravada na base de dados, para ser aplicado a uma frase afirmativa utilizando o produto CPqD Texto Fala, desenvolvido por pesquisadores da Fundação CPqD [1] – Centro de Pesquisa e Desenvolvimento em Telecomunicações, situado na cidade de Campinas, SP. O Texto Fala é um software especializado em converter uma frase textual em sinais de áudio. Este software utiliza uma base de dados com arquivos de áudio e os seus correspondentes elementos textuais. Os arquivos de áudio contêm frases gravadas por locutores com os perfis de entonação que se deseja aplicar. E assim, para cada perfil de entonação que se deseja aplicar à frase textual de entrada, é necessário que hajam frases gravadas na base de dados com o perfil desejado. A proposta deste trabalho é que, dada uma frase no formato textual afirmativo, o software faça a escolha do perfil de entonação interrogativo independente do cadastro ou não na base de áudio, não havendo a necessidade de realizar uma nova gravação.

Baseado nos conceitos de compiladores, propõe-se, neste artigo, um modelo de implementação, que a partir da classificação gramatical de uma frase textual em português, o software efetue a busca na base de dados com frases interrogativas e escolha um perfil interrogativo que será aplicado à frase original. O CPqD Texto Fala é utilizado por uma grande variedade de segmentos comerciais, desde pequenas a grandes empresas, tais como, em sistemas de caixas eletrônicos para ajuda a deficientes visuais, em call Center de companhias de telecomunicações, podendo ser usado para leitura de correios eletrônicos através de um aparelho de telefone. Selecionado o padrão

de entonação, devem-se efetuar estudos específicos focando a pronúncia final da fala a ser apresentada ao usuário final. Este estudo será realizado pela equipe técnica do CPqD.

## 2. Objetivo

O objetivo é propor um Modelo Teórico para Seleção Algorítmica de Entonação Vocal Baseado em Correlação e Similaridade. Utilizando como recurso o produto "CPqD Texto Fala" [5], é apresentado um modelo computacional que selecione um perfil de entonação em uma base de dados contendo frases interrogativas, e escolha a frase com maior similaridade de entonação, e então a mesma é selecionada para aplicação sobre a frase de entrada, inicialmente não cadastrada na base.

## 3. Justificativa

O produto CPqD Texto Fala contém uma base de frases gravadas por locutores com perfis de entonação afirmativos. Para se obter um novo perfil de entonação interrogativa, é necessário realizar novas gravações. Isso representa um problema, pois se para cada frase for necessário gravar amostras no padrão afirmativo, interrogativo e exclamativo isso tornaria a base muito grande e diminuiria o desempenho do sistema. Espera-se que o modelo proposto ajude a melhorar o desempenho do software, e, como conseqüência, possa servir como incentivo para mais investimentos em estudos futuros na área da comunicação.

## 4. Metodologia

O modelo que este trabalho apresenta, utiliza técnicas tradicionais aplicadas no desenvolvimento de compiladores e técnicas de correlação para a

seleção probabilística de modelos de entonação. O modelo está estruturado em cinco etapas:

1) Identificação do código fonte
2) Análise Léxica
3) Análise Sintática
4) Máquina de Validação/Ponderação Sintática - Sintática de Estado
5) Máquina de Seleção

O código fonte, ou seja, uma frase textual é submetida a um analisador léxico para validação das palavras (tokens). A sequência de tokens é submetida ao analisador sintático para validação e obtenção de uma seqüência sintática. A seqüência sintática é então ponderada com padrões sintáticos das frases/entonação previamente armazenadas no banco de dados. O resultado desta etapa é um valor de correlação entre a sintaxe de origem e as sintaxes armazenadas. A sintaxe que apresentar maior correlação é selecionada. Como resultado, a frase originalmente com entonação afirmativa é pronunciada na forma interrogativa.

## 5. Conceito de um compilador

Segundo Aho [2], um compilador é um software que tem o propósito de fazer a conversão de um programa escrito numa linguagem - a linguagem fonte – em um programa equivalente numa outra linguagem - a linguagem alvo (ver figura 1.1). Como importante parte desse processo de tradução, o compilador relata ao seu usuário a presença de erros no programa fonte.

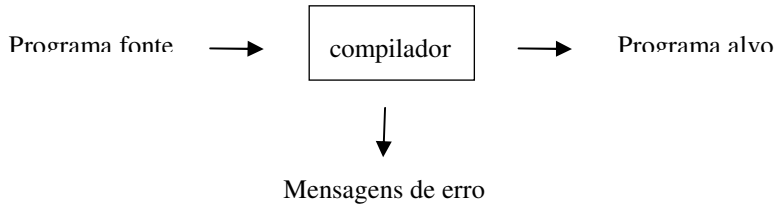


Figura 1 - Um compilador [2]

## Rotinas de Análise e fases do processo de compilação

Segundo Aho [2], a análise do processo de compilação é feita em seis etapas:

- Análise linear ou análise léxica: Lê os caracteres do programa fonte da esquerda para a direita e os agrupa em uma sequência de tokens com um significado coletivo.
- Análise hierárquica ou análise sintática: Os tokens, identificados na análise léxica, são agrupados nesta fase seguindo uma hierarquia que faz com que esse agrupamento possua um significado coletivo.
- Análise semântica: efetuar verificações a fim de assegurar a combinação significativa dos componentes do programa.
- Após as análises (léxica-sintática-semântica), é gerada uma representação intermediária do código fonte, processo ao qual dá-se o nome de “Geração de Código Intermediário”.
- Após a “Geração de Código Intermediário”, acontece o processo chamado de “Otimização de Código”. Esse processo tem a finalidade de otimizar o código intermediário de forma a torná-lo um código de máquina mais rápido em tempo de execução.
- Por fim, é gerado o código alvo. Geralmente, trata-se de um código realocável ou código de montagem. A esse processo dá-se o nome de “Geração de Código”. Para a aplicação em questão, o código gerado é uma frase interrogativa, com similaridade sintática ao código fonte.

Na figura 2 é representado um processo de compilação como descrito na proposta por Aho [2], e que norteou o modelo proposto neste trabalho.


Figura 2 - Fases de um processo de compilação [4]

## 6. CPqD Texto Fala

O CPqD Texto Fala [5] é um software capaz de transformar qualquer informação textual em fala, transformação esta feita em tempo real. É uma tecnologia que converte qualquer texto escrito em português do Brasil em sinal auditivo.

Existe uma diferença entre síntese de voz e síntese de fala. Voz pode ser qualquer som ou ruído emitido pelo aparelho fonador humano, enquanto a fala é um som emitido com significado lingüístico. Portanto, o CPqD Texto Fala sintetiza fala e não voz. A expressão inglesa correspondente a “síntese de fala” é speech synthesis.

Uma característica do CPqD Texto Fala, é a facilidade de integração com outras aplicações, além da variedade de mensagens que podem ser geradas. A seguir são apresentados alguns exemplos de aplicações:

“ATM e postos de auto-atendimento: pode ser utilizado em caixas de banco e terminais de atendimento, entre outros, facilitando a interação com usuários em geral que possuem  dificuldade de leitura" [5].

"*Auxílio a pessoas portadoras de deficiência*: Aplicações que são capazes de reproduzir sons podem utilizar o CPqD Texto Fala para interagir facilmente com pessoas portadoras de deficiência visual existindo uma comunicação mais natural e uma transferência de informação mais precisa" [5].

"*Auxílio à navegação no computador pessoal*: Pode ser usado para navegação de janelas, leituras de conteúdo de páginas da Internet, bem como para leitura de mensagens de correio eletrônico em forma de áudio" [5].

"*Leitura de informações de ajuda em softwares*: Pode ser aplicado para leitura de textos de ajuda de softwares, permitindo assim um acompanhamento das instruções passo-a-passo em forma de áudio" [5].

"*Serviços de tele-atendimento*: Pode ser usado para leitura de e-mails por telefone, conta bancária, conta de telefone, água, energia elétrica, previsão do

 ISSN 1679-8902

tempo, horóscopo, consulta de hotéis, variados serviços de call center, notícias, etc.” [5]

"*Substituição de interface visual por interface de áudio*: Útil em ambientes de monitoramento, como salas de controle, convertendo alarmes visuais de texto em fala, bem como em dispositivos inteligentes que recebam mensagens em formato textual, seja em casa, no escritório ou em um automóvel" [5].

## 7. Estrutura básica do modelo proposto

Na figura 3 é apresentado o modelo teórico proposto por este trabalho.

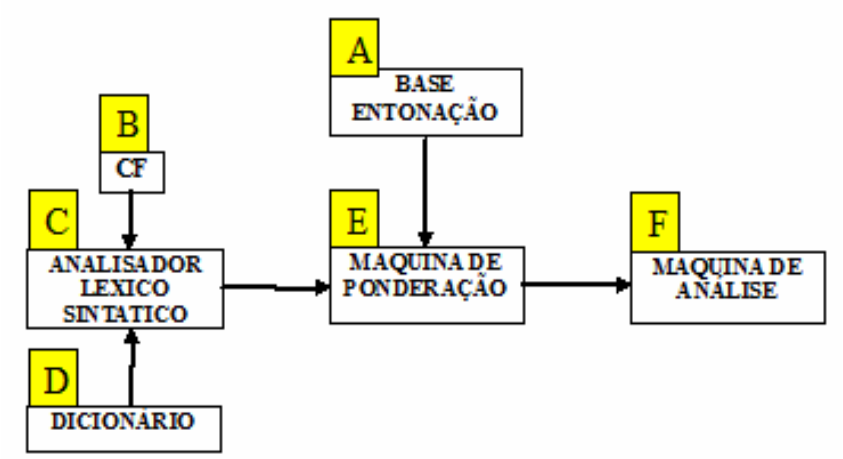


Figura 3 - Estrutura do modelo

A seguir serão explicadas cada uma das partes apresentadas na figura 3:

### A. Base Entonação - Base de Frases Gravadas com Entonação Interrogativa

Incorporado ao sistema Texto Fala [5], foi definida uma base de dados auxiliar contendo frases (amostras) gravadas com entonação interrogativa, que auxiliará no processo de seleção. Essa base será a fonte de busca para ser selecionado o padrão de entonação a ser aplicado à frase de entrada. Para

cada amostra, existem informações em modo textual e modo de áudio. O modo textual dessas amostras servirá de insumo para o processo de seleção incorporado no modelo proposto.

### B. CF – Texto de Entrada

O texto de entrada é a frase em modo textual, a qual será aplicada o processo de correlação e similaridade que visa selecionar o padrão de entonação interrogativo. Esse texto de entrada será referenciado aqui como Código Fonte (CF). Este CF é no primeiro momento só uma frase em formato textual e passará pelos analisadores léxico e sintático, tendo como base os dados de um dicionário (palavra e classe sintática dessa palavra) e um conjunto de regras sintáticas.

### C. Analisador Léxico Sintático

Para cada uma das amostras da base de entonação, será gerada uma frase sintática contendo classes gramaticais (artigo, substantivo, verbo...) para cada um dos elementos. Por exemplo: a frase “O gato pulou?” possui um artigo (o), um substantivo (gato) e um verbo (pulou). Então é montada uma frase sintática dessa frase e formam-se os tokens sintáticos “artigo substantivo verbo”. Esse processo é repetido para cada uma das amostras da base de entonação.

Da mesma forma, a saída do Analisador Léxico Sintático terá uma seqüência sintática e uma seqüência de tokens. Exemplo: a frase “A casa é azul” possui um artigo (a), um substantivo (casa), um verbo (é) e um adjetivo (azul). Esses elementos de saída serão usados como entrada para a máquina de ponderação.

O Analisador Léxico Sintático faz a separação do CF em tokens (palavras) e os classifica gramaticalmente, atribuindo para cada frase de

 ISSN 1679-8902

entrada uma frase sintática de acordo com o dicionário. Ou seja, após o CF passar pelo analisador Léxico Sintático, haverá duas saídas: 1 seqüência textual (separada em tokens) e uma seqüência Sintática (classificação gramatical de cada token). Conforme exemplo abaixo, temos três vetores (V1, V2 e V3), onde o primeiro vetor conterá o CF, no segundo vetor será gravada a seqüência sintática do CF e no terceiro vetor será gravada separadamente a seqüência sintática (tokens sintáticos) de cada token textual.

### D. Dicionário

É um arquivo contendo as palavras mais comuns da língua portuguesa e sua classificação gramatical. Será utilizado pelo analisador léxico sintático para devolver a seqüência sintática do CF. Caso a palavra não conste no arquivo, é selecionada uma classificação gramatical padrão.

### E. Máquina de Ponderação

É um conjunto de regras que será aplicado a cada uma das amostras da base usando como referência a frase inserida como CF, que no exemplo seria “A casa é azul”. Na máquina de ponderação terão que ser aplicadas regras para, por exemplo, pré-selecionar um conjunto de amostras da base que têm um perfil parecido com o perfil alvo.

#### 1) Tamanho da Amostra

A primeira regra a ser aplicada tem por objetivo selecionar amostras por tamanho (Fator de Escala), buscando frases com tamanho sintático aproximado ao do CF: Para cada amostra da base serão selecionadas aquelas que têm o tamanho próximo ao tamanho da frase alvo. A frase “A casa é azul” tem tamanho igual a 3 palavras, e nas amostras da base serão encontradas

frases com vários tamanhos, desde apenas uma palavra até dezenas de palavras.

**Seleção de frases da base pelo tamanho**

Define-se um fator de escala e aplica-se esse fator.

Na frase “A casa é azul”, o alvo tem tamanho igual a 3. Se for aplicado um fator de escala de 20% sobre 3, será obtida uma faixa de 2,4 a 3,6. Se arredondarmos os valores para os inteiros maior e menor, mais próximos, teremos 2 e 4. Então, nesse caso, serão selecionadas na base todas as amostras que têm tamanho de 2 a 4 palavras. Essa regra servirá para selecionar frases da Base de Entonação que tenham o mesmo tamanho ou tamanho similar à frase entrada (CF), usando um fator de escala de, por exemplo, 20% para mais e para menos.

**2) Regras de Ponderação por Energia**

Na próxima etapa, as frases selecionadas da base de entonação serão submetidas à máquina de ponderação.

Para cada elemento sintático das frases selecionadas, serão aplicados fatores de ponderação indicando a correlação sintática com o CF. Este fator de ponderação é chamado de **energia**, indicando o nível de proximidade sintática entre o token do CF e o token no banco de dados.

*Status*: Ao procurar na base de entonação as frases com os tamanhos desejados, usa-se um *status* para indicar se uma determinada amostra foi selecionada ou não. Ex.: Se a amostra tem o tamanho definido pelo fator de escala, atribui-se a esta um *status* = 1, e, em caso negativo, atribui-se a esta um *status* = -1.

 ISSN 1679-8902

Isso servirá para indicar a quais amostras devem ser aplicadas as regras de ponderação. Pode-se também criar um *status* = 0, que significa estar pendente a seleção daquela amostra – isso pode acontecer se houver algum critério de empate durante o cálculo de energia. Essas amostras podem ser usadas após passarem por outras regras ou análise.

*Nota:* este modelo não apresenta regras para tratar critérios de desempate. Estas regras deverão ser definidas futuramente.

### 3) Zerar todos os pontos de energia

O objetivo é encontrar a regra que tem a maior energia possível. Aquela que tiver maior energia tem a maior correlação e a maior similaridade com o alvo. Para cada token do CF, que é o alvo, verificar se existe na regra da base. Se existir, ponderar com 1. Se o token está na mesma posição da regra da base, somar 1. Equivalência funcional - todo artigo é igual (+0,1) e todo substantivo é igual (+0,5). Cálculo da distância funcional – ponderar negativamente a distância entre a posição ao token da regra da base e o token ao código fonte com -0,1. Se um token sintático não existir, ponderar fortemente com -0,8.

## Descrevendo a máquina de ponderação

Foi selecionada uma amostra qualquer da base e esta será comparada com o CF. Temos como exemplo o CF: “O gato preto pulou rapidamente”, e como amostra interrogativa, temos na base de dados a seguinte frase: “O carro saiu?”

O primeiro token da amostra é um artigo. Verifica se o primeiro token do alvo (CF) é um artigo. Se for, pondera com 1. No exemplo que este trabalho propõe, isso é verdade, então é ponderado positivamente.

O segundo token da amostra é um substantivo. Verifica se o segundo token do alvo (CF) é um substantivo. Se for pondera com 1. No exemplo que este trabalho propõe, isso é verdade, então é ponderado positivamente.

O terceiro elemento da amostra é um verbo. Verifica se o terceiro token do alvo (CF) é um verbo, se for pondera com 1. No exemplo que este trabalho propõe, isso não é verdade, então não é ponderado positivamente.

Verifica se o terceiro elemento com o quarto elemento é um verbo e se for pondera, porém negativamente. No exemplo que este trabalho propõe isso é verdade, portanto é ponderado.

| CF | | Amostra |
|---|---|---|
| **O** = artigo | ⟶ | **O** = artigo |
| **gato** = substantivo | ⟶ | **carro** = substantivo |
| **preto** = adjetivo | ⟶ | **saiu** = verbo |
| **pulou** = verbo | ⟶ | “ ” |
| **rapidamente** = advérbio | ⟶ | “ ” |

Pode-se notar que na amostra não existe adjetivo e nem advérbio, e que o verbo do CF não está na mesma posição do verbo da amostra.

Tabela ponderação

| Token Sintático (Amostra da Base de Entonação) | |
|---|---|
| 0 | Artigo |
| 1 | substantivo |
| 2 | Verbo |
| 3 | |
| . | |
| . | |
| N | |

| Token Sintático (CF alvo) | |
|---|---|
| 0 | artigo |
| 1 | substantivo |
| 2 | adjetivo |
| 3 | verbo |
| 4 | advérbio |
| . | |
| N | |

| Energia (CF alvo) | |
|---|---|
| 0 | 1 |
| 1 | 1 |
| 2 | -1 |
| 3 | -0,1 |
| 4 | -0,8 |
| . | |
| N | |

 ISSN 1679-8902

### F. Máquina de Análise

Aplicada a ponderação de energia para todas as amostras selecionadas na base de entonação, tem-se o *status* energético de cada uma das amostras. A próxima etapa consiste em selecionar a amostra que tenha a maior energia. A amostra que tiver a maior energia é a que tem a maior similaridade com a frase de entrada. Por exemplo, suponha que a base de entonação possua 100 amostras, e que foram selecionadas, por tamanho (conforme item E.1), 10 amostras e que, por exemplo, das 10 amostras selecionadas, a que tem a maior energia é a de nº 52 (que é a soma de todas as energias - positiva e negativa - atribuídas à frase). A amostra da base de entonação que possuir a maior energia será indicada como sendo a amostra que tem um possível perfil de entonação que deverá ser aplicado na frase alvo (CF). Neste exemplo, na base de entonação a amostra de número 52 foi selecionada como a que possui a maior energia, e o perfil desta amostra é aplicado ao alvo (CF).

## 8. Conclusão

Neste trabalho foi apresentada uma proposta de modelagem algorítmica para seleção de entonação de voz. Por se tratar de um modelo, neste momento, não há resultados, pois não foi realizada nenhuma implementação para testes finais. Diante de pesquisas teóricas, esta proposta de modelagem algorítmica para seleção de entonação de voz parece ser um método adequado para se obter uma pronuncia ideal, mais próxima da real (fala humana).

O modelo proposto foi submetido para a equipe responsável pelo sistema “CPqD Texto Fala”, que aprovou a sugestão. Serão realizados testes e, em um próximo trabalho, será feita implementado e testado.

## Referências

[1] FUNDAÇÃO CPqD. **CPqD.** Disponível em: http://www.cpqd.com.br. Acessado em: 27 de novembro de 2009.

[2] Compiladores: princípios, técnicas e ferramentas/ Alfred V. Aho et al.; tradução Daniel Vieira – 2. Ed.- São Paulo: Pearson Addison- Wesley, 2008.

[3] FUNDAÇÃO CPqD. **CPqD Texto Fala.** Disponível em: https://www.cpqd.com.br/textofala/telefonia/index.htm. Acessado em: 27 de novembro de 2009.

[4] Compiladores: princípios, técnicas e ferramentas/ Alfred V. Aho et al.; tradução Daniel Vieira – 2. Ed.- São Paulo: Pearson Addison- Wesley, 2008.

[5] FUNDAÇÃO CPqD. CPqD Texto Fala. Disponível em: http://www.cpqd.com.br/component/docman/doc_download/16-cpqd-texto-fala.html. Acessado em: 27 de novembro de 2009.