

PRO-GMÁTICA: uma base de dados prosódica pragmaticamente orientada em Português Europeu

Daniela Braga¹, João Paulo Teixeira², Luís Coelho³
e Diamantino Freitas⁴

¹Universidade da Coruña

²Instituto Politécnico de Bragança

³Instituto Politécnico do Porto

⁴Faculdade de Engenharia da Universidade do Porto

Resumo

Neste trabalho apresenta-se uma base de dados de fala espontânea com etiquetagem linguística, prosódica e pragmática em Português Europeu (PE). Esta informação pode ser encontrada num ficheiro de texto que é associado ao sinal de fala, sendo por isso possível analisar simultaneamente as várias dimensões em questão. O objectivo deste trabalho consiste em criar um corpus de fala etiquetado em que seja possível correlacionar a dimensão acústica e prosódica do sinal com o seu significado linguístico e pragmático, sobretudo tendo em atenção as dimensões emocionais e estilísticas do discurso.

Constituem motivações subjacentes a este trabalho a necessidade de controlar a manipular a naturalidade da fala sintética, a necessidade de implementar emoção e atitude nos actuais sistemas síntese e reconhecimento de fala e a inexistência de corpora de fala espontânea linguística e prosodicamente etiquetada e que contemple as dimensões emocional e comunicativa.

Neste artigo, faz-se uma breve introdução ao assunto em questão e apresenta-se o estado-da-arte das bases de dados em PE. Em seguida, descreve-se o processo da recolha da PRO-GMÁTICA e os critérios subjacentes à sua estruturação. Posteriormente, refere-se o tratamento e análise dos dados recolhidos, o que consiste basicamente no elenco e descrição dos níveis de etiquetagem considerados e na extracção dos parâmetros acústicos segundo o modelo Fujisaki. Finalmente, apresentam-se as conclusões e trabalhos futuros.

Desta forma, pretendemos, num futuro próximo, vir a dotar os sistemas de síntese da fala e de diálogo homem-máquina em PE de uma dimensão mais pragmática.

1. Introdução

Nos desenvolvimentos mais recentes em processamento da fala, tornaram-se frequentes as expressões “attitudinal prosody”, “emotional speech” e “affective computing”, o que vem demonstrar que, uma vez controlada a qualidade segmental dos

sistemas de síntese da fala, o passo seguinte prende-se com a manipulação do seu nível suprasegmental ou prosódico, nível este que é determinado não só por factores linguísticos (lexicais, sintácticos e semânticos), mas também pragmáticos ou comunicativos. Assim, a dimensão prosódica, responsável pela tão almejada naturalidade da fala sintética, informa também sobre o objectivo pragmático-comunicativo do locutor, sobre o grau de certeza ou incerteza que este tem em relação ao conteúdo proposicional do seu enunciado e até sobre o nível de familiaridade ou distanciamento que o locutor estabelece com o alocutário.

No entanto, não parece existir uma base de dados de fala em Português Europeu (PE) disponível para uso da comunidade científica que seja especificamente orientada para as questões que visam a interface prosódia/pragmática.

Neste trabalho, é nosso objectivo apresentar uma base de dados prosódica, elaborada segundo critérios pragmáticos, que convencionamos designar de PRO-GMÁTICA, com vista a servir de base para o estudo, descrição e futura previsão dos padrões prosódicos que estão na base dos actos de fala e de eventos discursivos relacionados com a noção de emoção e de atitude.

2. Bases de dados de fala em PE: estado da questão

Desde logo parece-nos importante sublinhar o abismo que existe entre bases de dados de centenas e milhares de horas de gravação, como as que se encontram para o Japonês (JSP/ CREST ESP Project – <http://feast.atr.jp>), por exemplo, e as bases que existem para o Português que, para além de pequenas e escassas, o que decorre naturalmente da falta de investimento económico e científico, são também, muitas vezes, de difícil acesso.

No plano nacional, gostaríamos de destacar as seguintes bases de dados de entre várias outras que poderiam ser mencionadas:

- *Corpus do Português Fundamental* devidamente transcrito e trabalhado segundo critérios de frequência (Nascimento, 1987);
- *Corpus de Referência do Português Contemporâneo* (Nascimento, 2000), desenvolvido pelo Centro de Linguística da Universidade de Lisboa, em 1997, que apesar de ser um corpus maioritariamente escrito (com uma dimensão total de 77,3 milhões de palavras), contém ainda 1 725 240 de palavras na variedade falada; abrange documentos publicados desde o século XIX até aos nossos dias e tem a particularidade de conter documentos publicados em todo o mundo de língua portuguesa;
- *Português falado, variedades geográficas e sociais* (1997), desenvolvido pelo Centro de Linguística da Universidade de Lisboa, que é composto por 4 CD-ROM com 83 gravações de português falado informal e formal, recolhido ao longo de 25 anos em Portugal, no Brasil, em todos os países africanos de língua portuguesa, em Macau e em Timor;

- FEUP-IPB (Teixeira, 2001) que contém uma etiquetagem prosódica, ao nível do acento e da frase, constituída por cerca de 100min de fala produzida por um locutor profissional a partir de textos de imprensa. A base de dados foi recolhida numa sala insonorizada e com equipamento profissional, tendo posteriormente sido etiquetada ao nível do fone, palavra, frase e com etiquetagem da frequência fundamental;
- *REDIP – Rede de Difusão Internacional do Português: rádio, televisão e imprensa* (iniciado em 2000), desenvolvido pelo Instituto de Linguística Teórica e Computacional, que é constituído por gravações em áudio e em vídeo relativas a amostras da rádio e da televisão (Ramilo, 2002);
- O corpus prosódico que está a ser construída pelo Centro de Línguas e Culturas – Grupo de Ciências da Linguagem da Universidade de Aveiro (<http://www.ii.ua.pt/cidlc/gcl/atlas/> consultado em 27-09-2005), inserida no projecto europeu *Multimedia Prosodic Atlas for the Romance Languages* (AMPER).

De um projecto conjunto entre instituições europeias, iniciado em 2000, surge a base de dados *C-ORAL-ROM Project* (<http://dit12.dit.unifi.it/coralrom/> consultado em 27-09-2005) orientada para o estudo dos níveis prosódicos e sintácticos, que possui registos comparáveis de fala espontânea em quatro línguas românicas (Italiano, Francês, Português e Espanhol). Este recurso representa um passo importante na área mas, para além de apenas estar disponível comercialmente e a um custo elevado, possui ainda alguns pontos que mereceriam alguma atenção e desenvolvimento para estar totalmente adaptado a uma utilização imediata nas aplicações de processamento de fala.

Desta forma, pelo que está documentado, nenhuma das bases de dados acima referidas parece orientar-se para a dimensão pragmática do discurso.

3. Recolha da PRO-GMÁTICA

A PRO-GMÁTICA é uma base de dados de fala espontânea organizada segundo determinados actos de fala e etiquetada de acordo com vários níveis informacionais, nomeadamente linguísticos, paralinguísticos e extralinguísticos, como será descrito no ponto seguinte.

São cada vez mais os autores que defendem a recolha de corpora de fala espontânea em detrimento de corpora de texto lido (Campbell, 2000) e, por isso, fabricados e artificiais, sempre que os objectivos subjacentes ao estudo em causa sejam analisar os eventos reais da voz humana e a sua utilização inserida numa determinada situação comunicativa. Por esta razão, e tendo em vista os objectivos futuros de investigação que já enunciámos, optámos por recolher um corpus de fala espontânea a partir de várias situações comunicativas do quotidiano. Para isso, gravámos e recortámos, nesta primeira fase, actos de fala que extraímos a partir de várias interacções verbais orais encontradas em material televisivo, nomeadamente entrevistas,

debates políticos e conversações obtidas a partir de *reality shows* e de telenovelas. Não foram, até ao momento, considerados outros formatos de programas, como o telejornal ou o documentário, por exemplo, por não se enquadrarem completamente na estrutura que seguimos de interacção verbal e porque partem da leitura de um texto em voz alta, ainda que essa leitura seja quase imperceptível.

Entendemos o conceito de interacção verbal oral tal como ele é definido por C. Kerbrat-Orecchioni (1990), ou seja, uma troca comunicativa entre pelo menos dois participantes em presença, que exercem influências recíprocas um sobre o outro e que realizam um conjunto de acções/ sinais verbais e para-verbais (prosódicos, cinésicos e proxémicos) que atestam o seu envolvimento mútuo na troca comunicativa e que asseguram a sua gestão (1990).

Os critérios que estiveram na base da estruturação da PRO-GMÁTICA assentam numa revisão da bem conhecida tipologia de actos de fala proposta por Austin (1962), aperfeiçoada por Searle (1969), posteriormente retomada e enriquecida por Kerbrat-Orecchioni (1990), que refere, embora sem descrever, os actos de fala de crítica, censura, refutação e troça, entre outros, quando descreve os taxemas verbais de marcação de relações verticais numa interacção verbal, ou seja, as relações interpessoais que se constroem em torno do eixo da dominação, do poder e da hierarquia, decorrentes do estatuto social dos participantes ou da sua habilidade discursiva (Kerbrat-Orecchioni, 1996).

No entanto, e apesar de a abordagem interaccionista ter enriquecido e alargado a tipologia inicial de actos de fala, são muito escassos os trabalhos sobre este assunto.

Neste trabalho, entendemos o conceito de acto de fala enquanto um enunciado, não necessariamente coincidente com os limites da frase, produzido com um determinado objectivo comunicativo e inserido numa determinada interacção verbal oral.

Para cada acto de fala foram consideradas dois domínios essenciais: um, segmental ou linguístico, em que se incluem as dimensões glotal, fonética e morfossintáctica; e outro, suprasegmental ou prosódico, em que se incluem as dimensões da emoção, do estilo e da atitude.

Na figura 1, em baixo, apresenta-se um esquema da tipologia de actos de fala considerados na PRO-GMÁTICA, estruturada segundo cinco objectivos comunicativos principais:

- 1) *Obter reacção do alocutário.* Trata-se de uma revisão do conceito de actos de fala directivos, segundo a tipologia de Searle. São possivelmente os primeiros actos a serem considerados quando se idealiza a arquitectura de qualquer sistema de diálogo homem-máquina.
- 2) *Cumprir funções sociais.* Trata-se de uma reformulação crítica do conceito de actos de fala expressivos proposto por Searle. Estes actos parecem-nos cumprir mais uma função social do que realmente expressiva, tal como era preconizada por Searle, na medida em que a condição de sinceridade não é necessária nem está subjacente à sua realização. Além disso, na maior parte dos casos também não exprimem verdadeiramente o estado psicológico do Locutor, uma vez que

essa dimensão pode decorrer mais do nível suprasegmental do que do nível segmental.

- 3) *Argumentar ou persuadir*. A partir do conceito de “acto de argumentar” introduzido por Ducrot e Anscombe (1997), no quadro da sua tese da argumentatividade intrínseca da língua, a que se seguiram desenvolvimentos posteriores conduzidos pela abordagem interaccionista, introduzimos aqui este objectivo comunicativo por nos parecer uma dimensão importante das interacções verbais.

Objectivo Comunicativo do Acto de Fala				
Objectivos de Searle	Comportamentos sociais	Argumentar ou persuadir	Relacionar o locutor com a verdade	Expressar compromisso
Ordem	Saudação	Crítica	Opinião	Promessa
Pedido	Pedido desculpa	Censura	Informação	Jura
Sugestão	Elogio	Refutação	Afirmacao	
Conselho	Agradecimento	Afirmacao	Negacao	
Pergunta	Despedida	Troca		
Pedido informação	Apresentação	Ironia		
Pedido confirmação	Felicitação			
	Lamento			

Figura 1. Organigrama da estruturação da PRO-GMÁTICA.

- 4) *Relacionar o locutor com o valor de verdade* da proposição expressa pelo enunciado. Neste grupo de actos de fala recuperam-se os actos de fala representativos proposta por Searle.
- 5) *Expressar compromisso*. Trata-se de uma retoma fiel dos actos de fala comissivos, segundo a tipologia de Searle.

Não foram considerados os actos de fala declarativos, nesta fase do trabalho, por nos parecerem menos aplicáveis aos sistemas de interacção homem-máquina.

Esta proposta tipológica de actos de fala não pretende ser nem definitiva nem completa, estando, por exemplo, outros actos de fala em vias de serem incluídos no nosso trabalho como o insulto, a mentira.

Foi nosso propósito alargar e reformular a tipologia tradicional de actos de fala, procurando assim vir a compreender o funcionamento desses actos de fala em todas as suas dimensões.

A PRO-GMÁTICA é, neste momento, constituída por 20 horas de material audiovisual televisivo gravado em VHS, entre os anos de 2003 e 2005, e posteriormente convertido para suporte digital. Desse material gravado foram seleccionados e recortados, com auxílio do PRAAT (<http://www.fon.hum.uva.nl/praat/>), enunciados prototípicos para cada um dos actos de fala considerados, num total de 10 frases para cada acto de fala.

As principais dificuldades que encontramos nesta fase do trabalho dizem respeito à classificação/catalogação do acto de fala, na medida em que o que se encontra, muitas vezes, é uma sobreposição de objectivos comunicativos, e logo, de actos de fala. Outro problema que surge com frequência é a necessidade de eliminar actos de fala com música, ruído de fundo ou sobreposição de vozes, por razões de preservação da qualidade acústica do exemplo, o que conduz à exclusão de bons exemplares de actos de fala.

Posteriormente, procedeu-se à etiquetagem da base de dados, como se descreverá no ponto que se segue.

4. Tratamento e análise dos dados

4.1. Etiquetagem da PRO-GMÁTICA

A registo áudio representa por si só um recurso de elevada importância mas pode ainda ser enriquecido pela adição de vários níveis de etiquetas que em sincronismo temporal com o sinal forneçam outras informações. Para a etiquetagem do sinal de fala da PRO-GMÁTICA, foram considerados os seguintes níveis informacionais:

- **Transcrição fonética estreita.** Do ponto de vista do sinal, o nível de anotação mais baixo localiza-se ao nível do fone. A tarefa de etiquetagem a este nível, que tradicionalmente é realizada à mão e por isso se torna demorada e dispendiosa, foi realizada utilizando uma ferramenta automática (Coelho, 2005) adaptada ao português que a partir do sinal de fala produz marcas temporais indicativas de início e término de cada fone e para cada intervalo compreendido entre marcas fornece uma classificação de acordo com um inventário fonético pré-estabelecido. O funcionamento desta ferramenta apresenta-se em esquema na figura 2. Os símbolos utilizados abarcam um inventário fonético constituído por 38 sons e 4 símbolos adicionais para assinalar silêncio, aspiração, oclusões e tónica. Este etiquetador, optimizado para a dimensão acústico-articulatória do PE, utiliza modelos de linguagem e regras fonológicas para a detecção e correcção de erros, o que faz aproximar os seus resultados, em condições favoráveis, dos obtidos por etiquetagem manual. O erro médio obtido é inferior a 10% na marcação de fronteiras considerando intervalos de 20ms e inferior a 5% na classificação, o que leva por vezes à necessidade de eventuais correcções manuais. Estes casos são geralmente facilmente identificados quando o registo áudio possui algum ruído ou quando a articulação não for clara.

- **Transcrição ortográfica simples**, com pontuação, apenas para permitir um melhor manuseamento do material gravado.
- **Etiquetagem morfossintáctica** ao nível da categoria gramatical da palavra (nome, verbo, adjectivo, determinante, pronome, advérbio, preposição, interjeição, conector, marcador discursivo).
- **Etiquetagem prosódica** que inclui a marcação de acento tónico (indicado no plano de etiquetagem dedicado à transcrição fonética através de um apóstrofo no início da sílaba), de fronteiras de grupo ou sintagma entoacional ou “phrasing” e com indicação de foco.
- **Emoção veiculada**, classificada entre 7 categorias (Mozziconacci, 2002): Estado Neutro, Alegria, Tristeza, Medo, Indignação, Surpresa, Raiva. São muitas as propostas de tipologias de emoções, mas decidimos concentrar-nos nas chamadas “big seven” que, por serem mais estereotipáveis, estão mais estudadas. Trata-se de uma classificação subjectiva, para a qual concorrem as informações semânticas do enunciado e a análise acústica do sinal. Futuramente, pretende-se fazer uma avaliação perceptiva com recurso a vários sujeitos para validar a classificação inicial.
- **Estilo**. Consideraram-se dois parâmetros: a proximidade/distanciamento entre locutores, e a cortesia, classificada através de uma polaridade de +/-delicado.

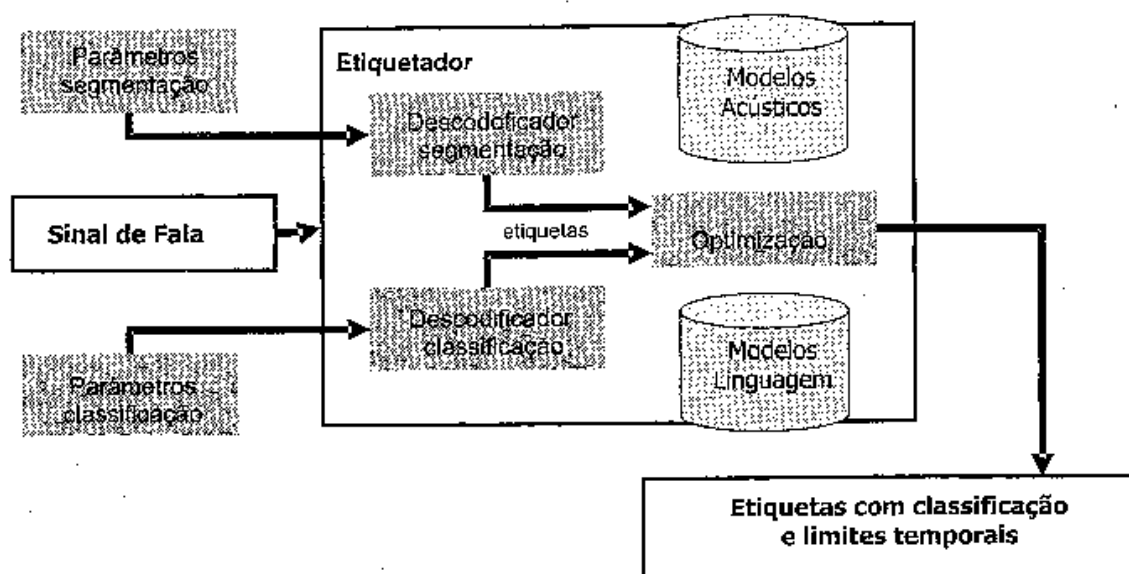


Figura 2. Funcionamento do sistema de etiquetagem automática..

- **Situação Comunicativa**, onde se dão indicações contextuais como a data, o programa de onde foi extraído, o Locutor que produz o enunciado em análise, e o tipo de interação verbal (debate, entrevista, conversação familiar, social, profissional).

Seguem-se alguns exemplos de actos de fala etiquetados de acordo com os níveis de análise descritos (figuras 3, 4 e 5):

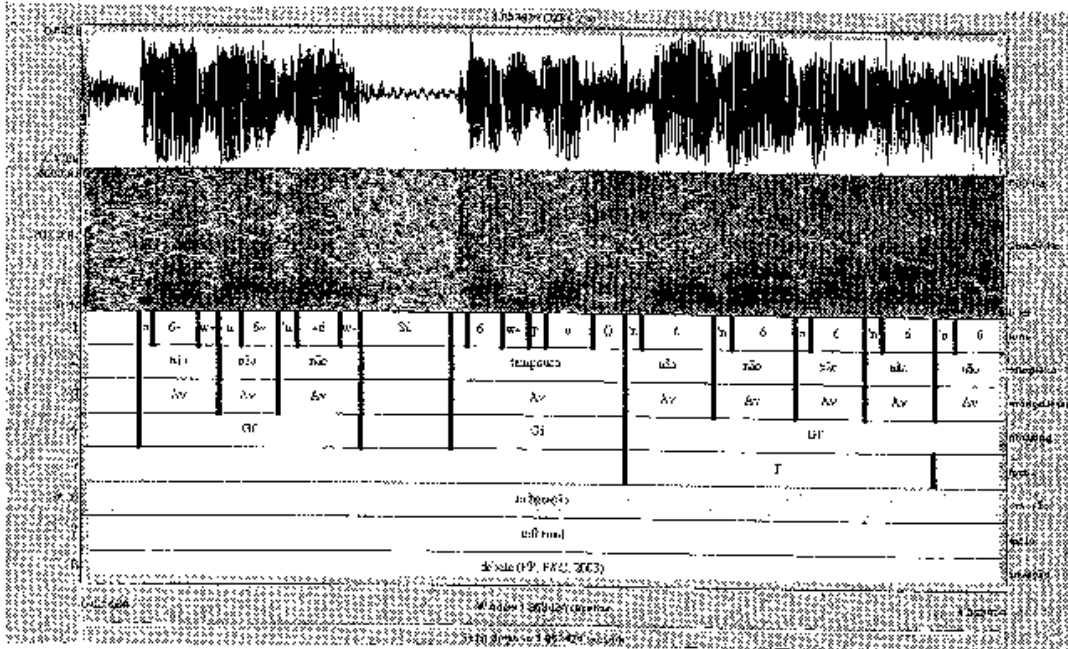


Figura 3. Etiquetagem de uma pergunta total
“O país não tem uma linha de rumo, Dr. Rui Rio?”.

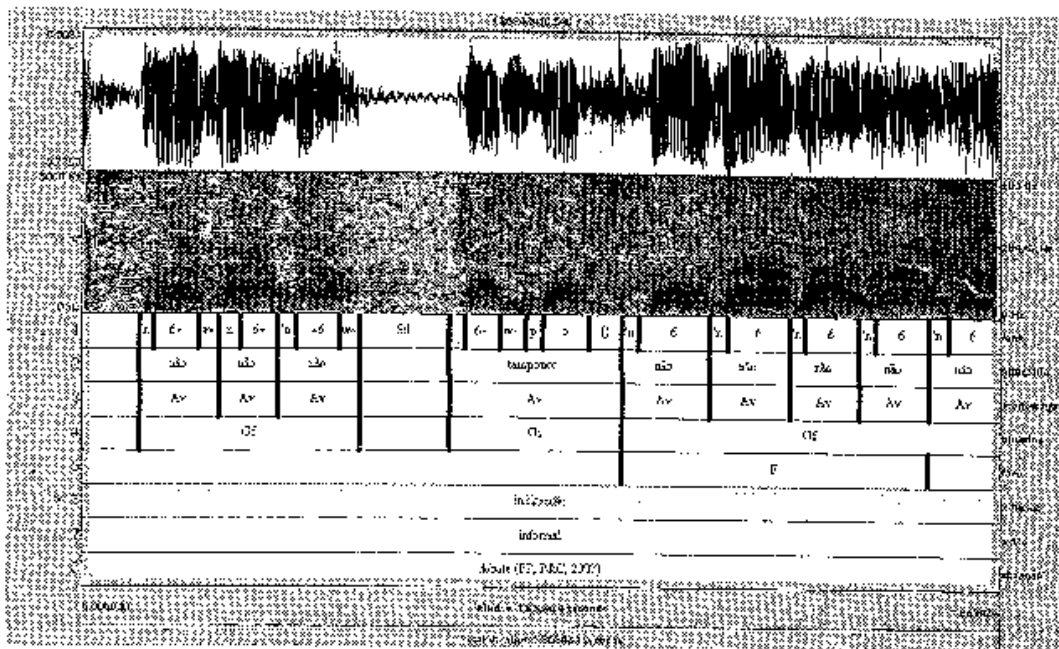


Figura 4. Etiquetagem de uma refutação “Não, não não!
Tampouco... Não Não Não Não Não!”.

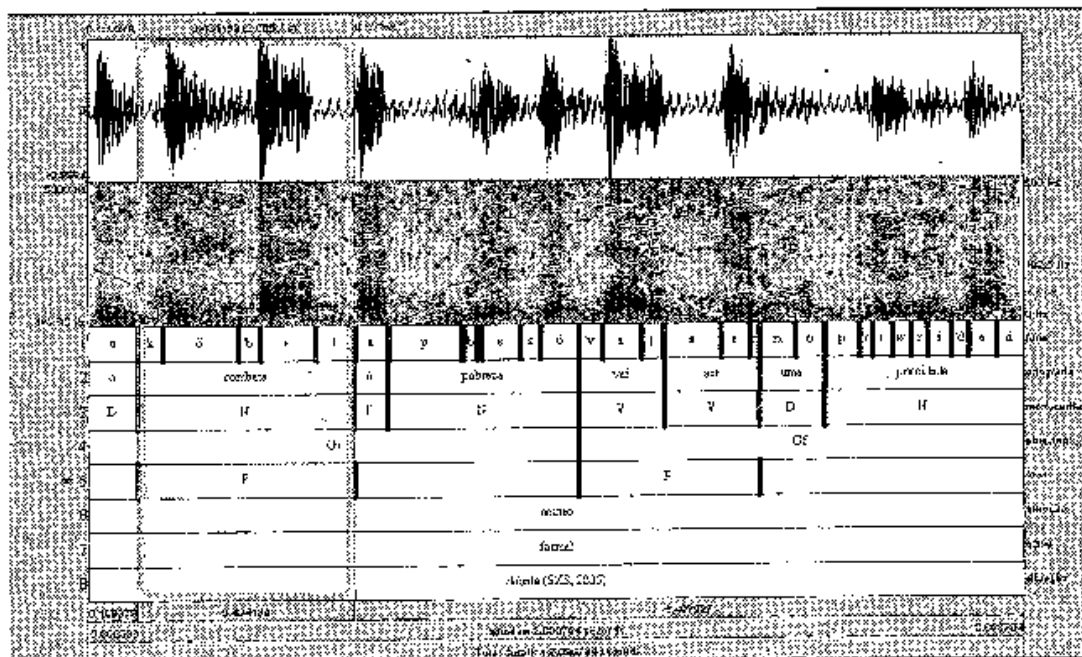


Figura 5. Etiquetagem de uma promessa
 “O combate à pobreza vai ser uma prioridade”.

4.2. Extração dos Parâmetros Acústicos

Os parâmetros acústicos prosodicamente relevantes são a frequência fundamental (F0), as durações segmentais e, menos relevante perceptualmente, a intensidade.

As durações são automaticamente retiradas dos ficheiros resultantes da etiquetagem. As durações dos fonemas são determinadas pela diferença dos instantes de início do fonema seguinte e do actual. A duração de unidades maiores será a soma das durações dos fonemas contidos na referida unidade. Além das durações dos fonemas, são ainda registadas as durações das pausas quer intra-frases, quer entre frases, dada a sua relevância para o presente estudo.

Em relação à intensidade foi registado para cada fonema o valor da intensidade média.

No que concerne ao parâmetro frequência fundamental, há diversas abordagens que procuram fazer uma representação das curvas de F0, sendo as mais usadas:

- **ToBI** (*Tone and Break Indices*). É baseado nos resultados de investigação científica em sistemas de entoação e na relação entre entoação e estrutura prosódica. Cada acento é representado por dois pontos que especificam o contraste relativo entre tons altos e baixos de forma abstracta. (Silverman, 1990).
- **Modelo Tilt**. Modelo que representa a entoação na forma de uma sequência linear de eventos (Taylor, 2000).

- **INTSINT (International Transcription System for INTonation)** proposto por Hirst e Di Cristo (1998). É um sistema de transcrição da entoação que codifica os padrões de F0 usando um conjunto de símbolos para os tons. Esses símbolos podem ser relativos ou absolutos.
- **Modelo Fujisaki.** É um modelo de geração dos contornos de F0 com fundamentos fisiológicos apresentado por Fujisaki em diversas publicações (1997). Trata-se de um modelo quantitativo que representa F0 pela adição de três componentes no domínio logarítmico. Apresentam-se de seguida as equações matemáticas resultantes do modelo fisiológico, com mais detalhe já que é o modelo usado neste trabalho.

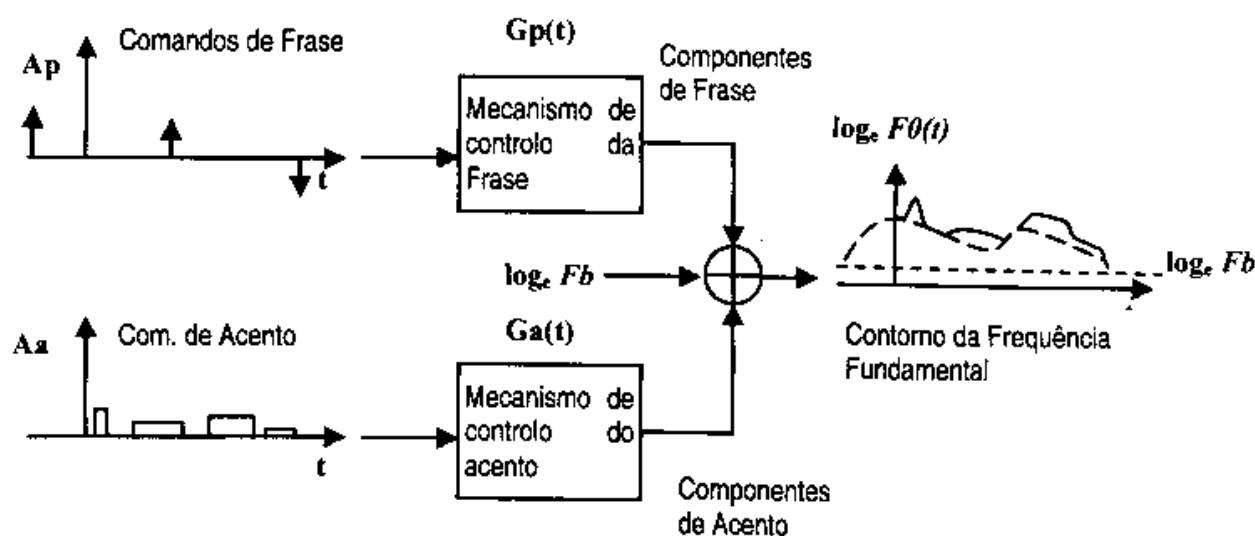


Figura 6. Modelo funcional de geração dos contornos de F0.

A figura 6 representa o processo de geração do contorno de F0 a partir dos Comandos de Frase (CF) e Comandos de Acento (CA) do modelo de Fujisaki. Os CF são um conjunto de impulsos e os CA um conjunto de funções em degrau. O contorno de F0 é expresso pela Equação 1 em que $G_p(t)$ da Equação 2 representa o mecanismo de controlo de frase e $G_a(t)$ da Equação 3 representa o mecanismo de controlo do acento.

Nas equações, F_b representa um valor de frequência fundamental base (dependente de cada frase). i e j são o número de CF e CA, respectivamente. A_{pi} e A_{aj} , são a amplitude dos CF e CA, de ordem i e j , respectivamente. T_{0i} , T_{1j} e T_{2j} , são os instantes dos CF i , e os instantes de início e fim do CA j , respectivamente. α , β e γ são a frequência angular natural dos CF e CA, e o valor máximo das componentes de acento, respectivamente.

$$\log_e F_0(t) = \log_e F_b + \sum_{i=1}^I A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^J A_{aj} \{G_a(t - T_{1j}) - G_a(t - T_{2j})\} \quad (\text{Eq. 1})$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (\text{Eq. 2})$$

$$G_a(t) = \begin{cases} \min[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\ 0, & t < 0 \end{cases} \quad (\text{Eq. 3})$$

Apresenta-se na figura 7 um exemplo de codificação das curvas de F0 com o modelo de Fujisaki. De cima para baixo é apresentado o sinal de fala, os valores de F0 medidos (representados com '+'), as curvas de F0 produzidas pelo modelo de Fujisaki (Fb + componentes de frase e Fb + componentes de frase + componentes de acento), comandos de frase (impulsos) e comandos de acento (degraus), sequência de sílabas alinhadas com o sinal de fala (sílabas tónicas a vermelho), marcas de frase (a vermelho), texto correspondente ao sinal de fala, sequência de fonemas alinhadas com o sinal de fala. É visível a aproximação entre as curvas de F0 produzidas pelo modelo e os valores medidos. No caso concreto o erro quadrático médio é de 2.96 Hz.

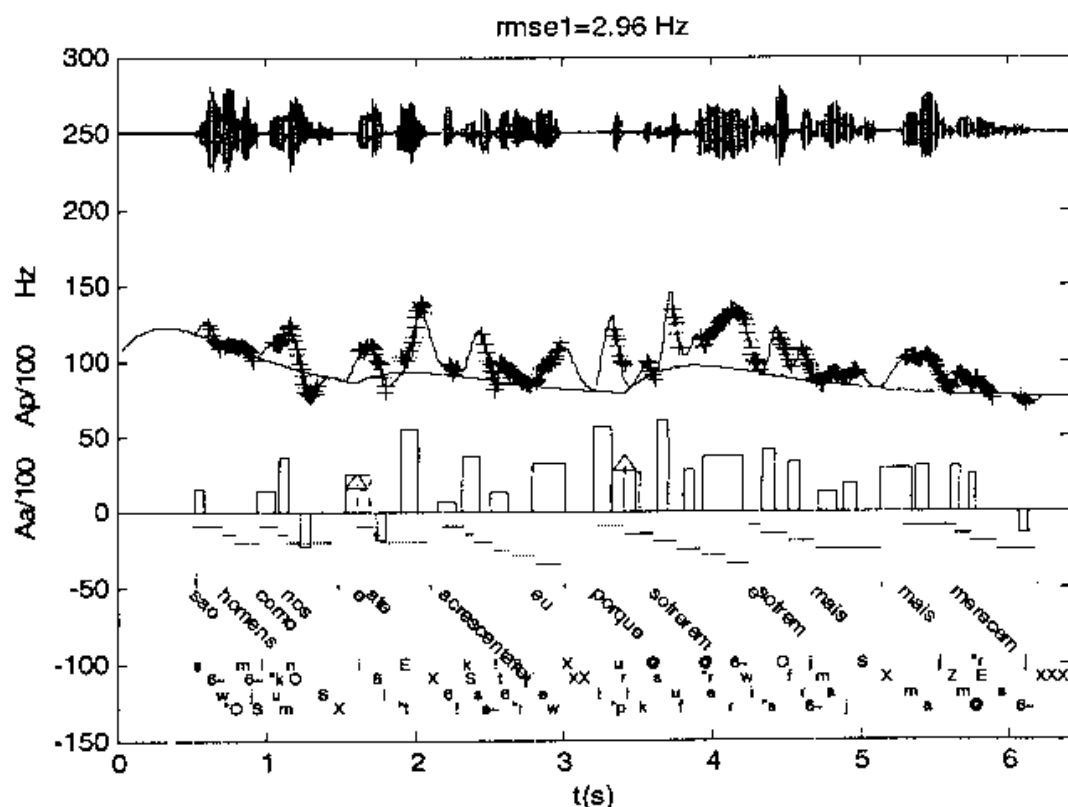


Figura 7. Exemplo de codificação do contorno de F0 com o modelo de Fujisaki.

4.3. Predição dos parâmetros prosódicos

Para cada acto de fala será criado um modelo que permita sintetizar os parâmetros prosódicos a partir do texto e eventuais marcas linguísticas.

Os modelos consistem na adaptação do modelo desenvolvido Teixeira (2004) para as durações segmentais e para o contorno de F0 a cada acto de fala.

O modelo de durações consiste numa rede neuronal artificial (RNA) cuidadosamente seleccionada no que concerne à sua arquitectura e tipo, bem como no conjunto de características a usar no vector de entrada. A arquitectura e vector de características de

entrada foram já definidos pelo autor. A adaptação do modelo a cada acto de fala consiste em determinar o vector de características para cada segmento e re-treinar a rede neuronal.

No caso do modelo de predição do contorno de F0, este baseia-se no modelo de Fujisaki e divide-se em dois sub-modelos. Um para predição dos parâmetros dos comandos de frase (amplitude – A_i , e localização – $T0_i$) e outro para predição dos parâmetros dos comandos de acento (amplitude – A_j , instantes de início e fim – $T1_j$ e $T2_j$). Os parâmetros destes sub-modelos são também preditos com RNA, pelo que necessitam de ser re-treinados para cada tipo de acto de fala e com os vectores de características de entrada definidos em (Teixeira, 2004).

Os vectores de características referidos para cada modelo são retirados a partir do texto.

Finalmente, sendo conhecidos os parâmetros do modelo de Fujisaki, resta aplicar as expressões matemáticas do modelo para ser conhecido o contorno de F0.

Conclusões

Neste artigo, fez-se a descrição de uma base de dados de fala espontânea em PE estruturada em actos de fala e etiquetada segundo níveis de informação linguística (fonética, ortográfica, morfossintáctica), prosódica (incluindo acento, f0 e durações) e pragmática (emoção, estilo e situação comunicativa).

O nosso objectivo é recolher um número suficiente de amostras de enunciados que nos permitam fazer uma descrição dos eventos prosódicos que permitem caracterizar cada acto de fala considerado neste artigo. Para isso, servir-nos-emos do modelo de Fujisaki para predizer a prosódia dos vários actos de fala produzidos em fala espontânea, uma vez que este método já foi testado e adaptado por Teixeira (2005) para a predição da prosódia da fala lida.

As principais dificuldades encontradas durante a construção desta base de dados tiveram a ver com problemas de definição dos tipos de actos de fala devido a uma permanente sobreposição de objectivos comunicativos, o que torna difícil a classificação das amostras de fala. Outro problema prendeu-se com a classificação de sinais não-verbais como risos, grunhidos, interjeições, aspirações e expirações, tão naturais e frequentes na fala espontânea e por vezes eles próprios portadores de conteúdo emocional e até comunicativo (Campbell, 2005).

Como planos futuros, pretendemos recolher segmentos de fala espontânea que em que possamos analisar não só a delicadeza linguística, como também as dimensões da certeza/ incerteza ou a dúvida tão relevante para sistemas de diálogo homem-máquina com ou sem fala.

Referências

- Austin, J. (1962) *How to do things with words*. Oxford: Oxford University Press.
- Campbell, N. (2000) Databases of Emotional Speech. In. *Proceedings of ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework*, Newcastle, N. Ireland, pp. 34-39.
- Campbell, N. (2005) No Laughing Matter. In, *Proceedings of Eurospeech 2005*. Lisboa, Portugal.
- Coelho, L. (2005) *Etiquetagem Automática de Sinais de Fala*. Dissertação de Mestrado. Faculdade de Engenharia da Universidade do Porto.
- Ducrot, O.; Anscombe, J. C. (1997) *L'argumentation dans la langue*. Liège: Mardaga.
- Fujisaki, H. (1997) Prosody, models and spontaneous speech. In, Sagisaka, Y., Campbell, N., Higushi, N., *Computing Prosody*. Springer-Verlag: New York, inc. pp 27-42.
- Hirst, D. and Di Cristo, A. (1998) *Intonation Systems – A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Kerbrat-Orecchioni, C. (1990, 2002) *L'énonciation. De la subjectivité dans le langage*. Paris: Armand Colin.
- Kerbrat-Orecchioni, C. (1996). *La conversation*. Paris: Seuil.
- Mozziconacci, S. (2002) Prosody and Emotions. In. *Proceedings of Speech Prosody 2002*. Aix-en-Provence, France.
- Nascimento, M. F. Bacelar (2000) O corpus de referência do português contemporâneo e os projectos de investigação do Centro de Linguística da Universidade de Lisboa sobre variedades do português falado e escrito. In. Gärtner, E.; Hundt, C.(ogs.) *Estudos de Gramática Portuguesa I*. Frankfurt am Main: TFM, pp. 185-200.
- Nascimento, M. Fernanda Bacelar; Marques, M. Lúcia Garcia; da Cruz, M. Luísa Segura (1987) *Português Fundamental: métodos e documentos, inquérito de frequência*. Lisboa: INIC/CLUL.
- Ramilo, M. C.; Freitas, T. (2002) A linguística e a linguagem dos media em Portugal: descrição do projecto REDIP. In. *XIII Congreso Internacional da ALFAL*. San José, Costa Rica.
- Searle, J. (1969) *Speech Acts. An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Silverman, K. and Pierrehumbert, J. (1990) The Timing of Prenuclear High Accents in English. In. J. Kingston and M. Beckman (eds) *Papers in Laboratory Phonology I*. Cambridge University Press: Cambridge UK, pp. 72-106.
- Taylor, P. (2000) Analysis and Synthesis of Intonation using the Tilt Model. *Journal of the Acoustical Society of America*. vol 1073, pp. 1697-1714.
- Teixeira, J. P et al. (2001) Phonetic Events from the Labeling the European Portuguese Database for Speech Synthesis, FEUP/IPB-DB. In, *Proceedings of Eurospeech 2001*. Aalborg, Denmark, pp.1707-1710.
- Teixeira, J.P. (2004) *A Prosody Model to TTS Systems*. Dissertação de Doutoramento. Faculdade de Engenharia da Universidade do Porto.