

Análise e modelamento dinâmicos da prosódia do português brasileiro¹

Dynamical analysis and modelling of Brazilian Portuguese speech prosody

Plínio A. Barbosa
DL/IEL/Universidade Estadual de Campinas

Abstract

This work presents, from output to input, a dynamical model of speech rhythm able to become a model of speech prosody. The model is based on coupled-oscillator theory, a class of dynamical systems theory that brings temporal variability into language by the bias of speech. From segmental duration to syntax-prosody interface, the model simulates several phonological processes such as secondary stress, variation of the duration patterns undergone by speech rate change, segmental effects conditioned by prosodic structure, compensatory effects on duration, and so on. Regularity and structuring are linked together in a principled way from coupling of both speech production and perception, highlighting the importance of a dynamical cyclic attractor. The model can deal with duration perturbation as a result of paralinguistic information or speech pathology.

Keywords

Prosody; Phonetics; Phonology; Brazilian Portuguese; Dynamical systems

Resumo

O trabalho apresenta, de forma retrospectiva, o funcionamento de um modelo dinâmico do ritmo da fala que pode *a posteriori* se constituir em modelo da prosódia da fala. Esse modelo fundamenta-

se na teoria de osciladores acoplados, uma classe de teoria de sistemas dinâmicos que leva o variável no tempo para o interior da língua, a partir da fala. Da duração segmental à interface sintaxe-prosódica, o modelo dá conta de diversos processos fônicos como acentuação secundária, variação na estrutura duracional por variação na taxa de elocução, efeitos segmentais decorrentes da estrutura prosódica, efeitos compensatórios de duração, entre outros. A regularidade acoplada à necessária estruturação na produção e percepção de proeminências prosódicas são tratadas de forma não-dicotômica pelo princípio do acoplamento e de atratores cíclicos. O modelo tem ainda a vocação de lidar com diversas alterações decorrentes de informação paralingüística ou mesmo de patologias da fala.

Palavras-chave

Prosódia; Fonética; Fonologia; Português brasileiro; Sistemas dinâmicos

1 Introdução

O principal objetivo deste artigo é o de justificar, teórica e experimentalmente, o emprego de construtos-chave das teorias de sistemas dinâmicos (doravante TSD), particularmente as teorias voltadas para o estudo do acoplamento entre osciladores, para o estudo científico da prosódia da fala. O emprego de uma TSD nas ciências da fala (PORT; GELDER, 1995) e na psicologia experimental (KUGLER; TURVEY, 1987; KELSO, 1995) tem raízes nos estudos da motricidade em geral, a partir da investigação experimental da coordenação entre membros em humanos e não-humanos adultos, para atividades tão diversas quanto a locomoção, mastigação, respiração, fala, e a partir do aprendizado dessas mesmas coordenações, como nas pesquisas de Thelen e Smith (1994) e Kelso (1995, p. 159-185). Para uma introdução à dinâmica da coordenação motora, ver Kelso (1994), e para uma introdução geral às TSD, ver Hirsch (1984).

Numa TSD, o indivíduo e seu ambiente constituem um sistema dinâmico, isto é, um objeto de estudo formado por partes que se relacionam de alguma maneira e que evoluem no tempo. Teorizar a estruturação prosódica da fala como produto do comportamento de um sistema dinâmico subjacente pressupõe que se aceite que a explicação científica está intimamente relacionada à inclusão de um modelo qualitativo ou matemático-computacional na teoria proposta. Esse modelo permite prever o comportamento do sistema no decorrer do tempo, ou seja, prever suas mudanças de estado para situações comportamentais diversas. A constituição desse sistema numa TSD se dá por auto-organização (CLARK, 1997), propriedade que caracteriza um sistema em estado de mudança, cujo comportamento macroscópico é resultante da interação entre seus componentes. Essa interação é provocada ou disparada por um estímulo externo que age como parâmetro de controle com conseqüências comportamentais descritas de forma menos invariante por um parâmetro de ordem ou variável coletiva (KELSO, 1995, p. 16-17). Uma característica singular da auto-organização é a influência que os

componentes do sistema exercem uns sobre os outros bem como recebem do e determinam no ambiente em que o sistema se encontra. Essa influência recebe o nome de *acoplamento*. Um exemplo histórico ilustra o que se entende por isso no caso de dois relógios de pêndulo.

Conta-se que o físico holandês Christiaan Huygens teve que ficar longo tempo acamado por motivo de doença em 1665, tendo a ocasião de observar por longos períodos os movimentos dos pêndulos dos relógios que inventara dez anos antes (STROGATZ; STEWART, 1993). Notou que os pêndulos dos relógios na parede que lhe fazia face estavam oscilando em sincronismo de fase, isto é, os instantes de tempo em que atingiam os extremos de deslocamento à esquerda e à direita eram os mesmos. Após perturbá-los, retornavam ao sincronismo de antes, decorrida cerca de meia hora. O acoplamento entre eles era assegurado por pequenas perturbações transmitidas pela parede de um para outro, tanto que, quando colocados em paredes distintas, não mais oscilavam simultaneamente, como um todo funcional.

O acoplamento também existe na fala, sendo a propriedade ou mecanismo que a constitui crucialmente, pela influência entre falantes. A longo prazo, os falantes de uma ou mais comunidades lingüísticas fazem emergir o falar do indivíduo que se constitui como um sujeito-outro durante o processo de aquisição da linguagem. A curto prazo, o acoplamento é evidente num indivíduo imerso numa comunidade lingüística diversa da sua, por morar um certo tempo em nova região ou novo país. O sincronismo de fase também é observado entre falantes que têm uma língua comum, durante a recitação de qualquer texto, falado ou cantado, aprendido de cor ou numa leitura em simultaneidade (CUMMINS, 2002).

Também há acoplamento entre duas tendências com efeitos prosódicos em oposição: uma tendência à regularidade na extensão temporal e uma tendência à diferenciação temporal tanto de unidades do tamanho da sílaba (que caracteriza a silabicidade) quanto de unidades do tamanho do grupo acentual (*stress group*). A regularidade serve o falante, pois torna a produção do enunciado mais simples: se os eventos são regulares, basta planejar sua frequência, e não os eventos isoladamente. A diferenciação gera estrutura temporal e serve o ouvinte, pois assinala, chama a atenção, para uma determinada informação lingüística e paralingüística (para uma exposição completa dessa complementaridade ver os trabalhos de FRAISSE, 1974 e de BARBOSA, 2006a, caps. 2 e 3). Essas duas tendências agem como um cabo de guerra, prevalecendo ora uma, ora outra,

e, nesse jogo em que mecanismos regular e diferenciador agem como atratores cíclicos, dá-se a enunciação. Um atrator é um padrão de comportamento ou movimento estacionário para o qual um sistema dinâmico tende. Se esse padrão é oscilatório, tem-se um atrator cíclico. Um oscilador como o relógio de Huygens ou a mandíbula são exemplos típicos de atratores cíclicos.

Desde o balbucio, a mandíbula em oscilação atua como máscara silábica por meio da qual a criança vai trabalhando, em sua interação com o meio, os segmentos que constituem a(s) língua(s) que aprende. Esse balbucio é um mecanismo de enunciação a um tempo contínuo, pois se estende no tempo, e discreto, pois pode-se identificar e isolar ciclos individuais de produção silábica (MACNEILAGE e DAVIS, 1990; MACNEILAGE, 1998).

O trabalho de pesquisa que temos conduzido procura mostrar a importância e os avanços conseguidos no estudo experimental da prosódia quando se trabalha com uma TSD, particularmente uma teoria de osciladores acoplados (BARBOSA, 2006a). Neste artigo, escolho a simulação passo a passo, através de um modelo matemático-computacional da enunciação, para mostrar as vantagens trazidas pelo abandono conseqüente de dicotomias como *langue e parole*, corpo e mente. A enunciação ilustrada é oriunda da leitura de um texto de Monteiro Lobato, de *A Menina do Narizinho Arrebitado* (1920).

2 Geração (retrospectiva) da estruturação prosódica do português brasileiro

A seguir ilustramos, etapa por etapa, uma proposta de geração automática da organização temporal do texto de Lobato escolhido, descrevendo simulações produzidas por um modelo dinâmico do ritmo da fala (doravante MDR) que leva em conta a variabilidade prosódica intra- e interindividual. A apresentação que proponho é retrospectiva, isto é, começo pelas durações dos fones do texto (segmentos isomorfos ao fonema) para terminar na interação entre sintaxe e prosódia. Comento, a cada etapa, as diferenças encontradas entre leituras naturais e simulações efetuadas pela técnica da perturbação. No caso do MDR, perturbar equivale a modificar de diversas maneiras os parâmetros desse modelo, de forma a reproduzir uma variabilidade na organização temporal que, desejavelmente, seja semelhante à variabilidade natural. Já no caso de dados naturais, perturbar experimentalmente esses dados equivale a introduzir diversas situações para eliciação deles, a fim de gerar diferentes organizações prosódicas,

tais quando o sujeito lê em diferentes taxas de elocução, quando enuncia em diferentes situações comunicacionais, quando narra ou descreve situações familiares ou novas (cf. MERLO, 2006). Os componentes do MDR (BARBOSA, 2001, 2002, 2006a) responsáveis por uma operação em particular são justificados em cada etapa até a entrada do modelo para que, ao final da apresentação, se tenha um retrato do modelo dinâmico de ritmo como um todo. Apesar de me concentrar nos dados de duração, também mostro como conjugar a geração da estuturação rítmica com a entoacional *stricto sensu*.

2.1. Corpus usado nas análises e simulações e medida da duração da unidade VV

O texto escolhido tem uma extensão que proporciona uma leitura de cerca de um minuto. Por tratar-se de um texto de autor de histórias infantis que faz parte do imaginário de todo brasileiro, as leituras foram feitas com muita naturalidade, que se traduziu por diferentes estratégias de colocação de proeminência e ênfase, de segmentação de constituintes prosódicos menores e maiores, de acelerações e decelerações diversas ao longo da produção do texto. Quatro locutores paulistas do sexo masculino na faixa de 20 a 30 anos na época da gravação leram o texto, começando em sua taxa de elocução espontânea, seguida de uma leitura em taxa lenta, à escolha do locutor, e outra o mais rapidamente possível, sem introduzir disfluências na fala (em caso de haver, a leitura inteira foi repetida até que não houvesse mais, o que aconteceu para um locutor, apenas uma vez). O texto conta o episódio do discurso do papagaio, do vexame da Lagartixa e do número de magia dos vagalumes, como se lê abaixo.

Em seguida apareceu um papagaio real que tinha fama de orador. Subiu a tribuna de um poleiro de ouro e fez um belo discurso a respeito da arte de falar. Nesse discurso provou que os homens tinham aprendido a falar com os papagaios, e não os papagaios com os homens, como diz a ciência destes. Uma chuva de palmas acolheu suas palavras.

O mesmo não aconteceu, porém, com a poetisa Lagartixa, que principiou a recitar uma longa poesia e engasgou no meio, acabando o recitativo em choro e faniquito. Para destruir essa má impressão vieram três vagalumes mágicos que fizeram várias sortes, sendo muito apreciada a sorte de comer fogo.

Os sinais acústicos correspondentes às doze leituras (quatro locutores, três taxas de elocução) foram segmentados e etiquetados manualmente em unidades VV usando o *software* Praat (BOERSMA; WEENINK, 2005). A unidade VV é compreendida entre o *onset* acústico de uma vogal e o *onset* da vogal imediatamente seguinte. A razão para o emprego dessa unidade foi apresentada várias vezes, sendo ele tão antigo quanto a década de 1930 (CLASSE, 1939). Essa unidade integra fones de duas sílabas distintas (da rima da primeira, do ataque da segunda) e evidencia a cadência da posição da vogal, que caracteriza essencialmente o ritmo da fala. Esse aspecto de fluxo carreador tem sido discutido com dados experimentais em várias línguas desde a década de 1930 até os dias atuais (BARBOSA, 2006a, p. 29-98). A seqüência de sílabas organizada dessa forma revela a estrutura rítmica do enunciado, tanto no que diz respeito à produção quanto à percepção da fala.

As durações das unidades VV ao longo das leituras de cada locutor, em cada taxa de elocução, foram normalizadas pela técnica estatística clássica do z-score e suavizadas com um filtro de média móvel de cinco pontos, a fim de fazer ressaltar uma duração que reflita o efeito da realização do acento frasal (*phrase stress*) ao longo da leitura. Cada pico local de duração suavizada é considerado, para efeitos de produção, como reflexo de um acento frasal, independentemente do valor da magnitude do pico. A duração é, em PB, o principal correlato acústico desse acento, pelo menos na fala lida (MASSINI, 1991; BARBOSA, 1996; FERNANDES, 1976).

A figura 1 ilustra o efeito da normalização e suavização da duração. Os cinco picos locais de duração ao longo da duração bruta, devidos à duração intrínseca de unidades VV específicas (por conter vários segmentos ou por conter um ou mais segmentos intrinsecamente longos) ou à implementação do acento lexical, reduzem-se a três picos: no início do enunciado e no final, além do pico central na palavra “tribuna”. Todas as três posições são prosódica e sintaticamente relevantes na pronúncia dessa frase. Alguns locutores não realizaram o acento frasal em “tribuna” e outros, se o realizaram na taxa lenta, não o realizaram na taxa rápida. Na figura, a duração suavizada, que é um número sem unidade física assinalando o grau médio de esticamento/encurtamento da unidade VV, foi convertida em milissegundos para fins de comparação no mesmo gráfico. Para tanto, desnormalizou-se a duração pelo uso de valores-referência de média e do desvio-padrão das durações da unidade VV, conforme indicado adiante. Esse procedimento é análogo ao usado para obter as durações dos fones.

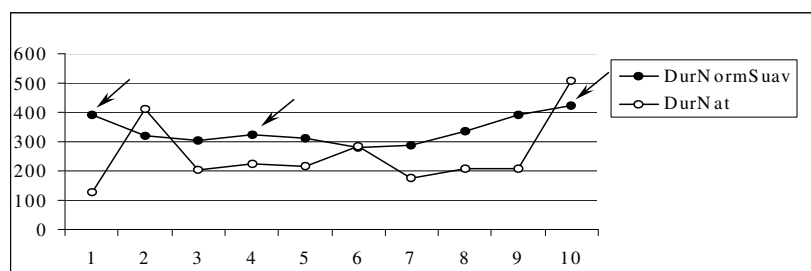


FIGURA 1 – Evolução da duração bruta (DurNat) e normalizada/suavizada (DurNormSuav) em milissegundos das unidades VV do enunciado “Subiu a tribuna de um poleiro de ouro.”, lido pelo locutor AP na taxa lenta. As setas indicam os locais de pico, em “-ub-” (subiu), “-un-” (tribuna) e “-our-” (ouro).

2.2. Obtendo durações dos fones a partir da duração da unidade prosódica mínima que os contém

O MDR, em seu estágio atual, gera a duração abstrata das unidades VV ao longo de enunciados, conectados ou não, a partir de informação sintática local introduzida manualmente. Com seu desenvolvimento, passaremos a incluir a geração da organização entoacional *stricto sensu*, bem como a interação entre prosódia e segmentos de maneira biomecanicamente plausível. A duração da unidade VV gerada no momento é abstrata por não integrar a influência da duração intrínseca e da implementação do acento lexical, justamente os efeitos atenuados pela técnica de cálculo do z-score apresentada anteriormente. Tal como fiz num trabalho anterior para o francês (BARBOSA; BAILLY, 1994), a reintrodução desses graus de liberdade da organização temporal, também chamados de microprosódicos, pode ser feita em duas etapas:

1. Obtenção do valor de z-score (z_{abst}) da duração abstrata da unidade VV em milissegundos (dur_{abstVV}), utilizando tanto o valor do parâmetro T_0 como média, que é o especificador subjacente da taxa de elocução, quanto 40 % de seu valor como desvio-padrão da duração das unidades VV abstratas. A percentagem de 40 % representa o coeficiente de variação das durações das unidades VV para as simulações com o MDR. A fórmula é dada então por: $z_{abst} = (dur_{abstVV} - T_0) / (0,4.T_0)$.

2. Uso do valor de z-score apresentado para cálculo da duração de cada fone $f(\text{dur}_f)$, compondo a respectiva unidade VV fazendo: $\text{dur}_f = m_f + z_{\text{abst}} \cdot s_f$.

O par de descritores estatísticos de duração do fone, (m_f, s_f) , é obtido de um *corpus ad hoc* para um locutor-referência do PB. A tabela dessas durações encontra-se no apêndice B do livro de Barbosa (2006a). As características duracionais específicas do locutor que se simula são especificadas pela seqüência de valores de z_{abst} , pois os descritores da duração intrínseca dos fones do locutor-referência não afetam essa informação, apenas restituem efeitos de contraste duracional lingüísticos, isto é, comuns aos locutores de uma mesma comunidade.

Para uma simulação que mantenha a mesma taxa lenta do locutor AP (para $T_0 = 225 \text{ ms}^2$), o erro médio (desvio-padrão) entre a duração da unidade VV produzida pelo locutor e aquela obtida com o MDR foi de 26 (75) ms, desconsiderando as pausas silenciosas, problema a parte ainda a ser tratado no MDR (para uma solução inicial para a geração da pausa silenciosa ver BARBOSA; BAILLY, 1997). Os maiores desvios ocorreram em palavras com ênfase, onde há emprego simultâneo de um movimento de subida de freqüência fundamental (f_0) e de alongamento da sílaba fonológica lexicalmente acentuada. Essa situação era esperada, tendo em vista que a ênfase pressupõe conhecimento semântico, que não foi incorporado ao MDR até o presente momento. Observe na figura 2 a semelhança entre durações de unidades VV naturais e simuladas para o mesmo enunciado da figura anterior.

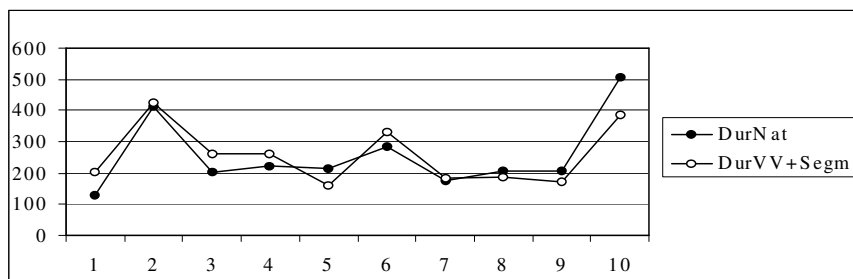


FIGURA 2 – Evolução da duração bruta natural (DurNat) e simulada (DurVV+Segm) em milissegundos das unidades VV do enunciado “Subiu a tribuna de um poleiro de ouro.” O enunciado natural foi lido pelo locutor AP na taxa lenta.

O mecanismo que explica o sucesso do modelo em gerar a estruturação prosódica dos enunciados reside na adoção de um modelo de osciladores acoplados contendo dois osciladores universais que se influenciam mutuamente, o oscilador silábico e o oscilador acentual. O primeiro gera as durações de unidades VV, o segundo gera a segmentação e proeminência prosódicas.

2.3. Obtendo durações de unidades VV pelo acoplamento entre dois osciladores universais

A duração abstrata de cada unidade VV ao longo do enunciado depende de dois mecanismos de ação, um de indução, responsável por aumentar a duração ao longo do grupo acentual, que culmina num acento frasal. O segundo, de decaimento, responsável por fazer um reinício (*reset*) de duração após a realização do acento frasal, para começar novo aumento duracional no novo grupo acentual a sua direita, caso esse exista. Quanto à forma desse aumento duracional, demonstrei em outro lugar que é exponencial, e não linear (BARBOSA, 2002). A taxa ou grau do aumento duracional depende, por sua vez, da magnitude da proeminência prosódica por vir, além de depender da distância dessa última à unidade VV corrente. O mecanismo de decaimento depende essencialmente da magnitude da proeminência prosódica que acabou de ser realizada, quanto maior ela for, maior a gama de *reset* de duração. Esses princípios de funcionamento estão reunidos na fórmula 1, em que T_n é a duração abstrata da unidade VV corrente, DT_n é a modificação de duração de unidade VV para a unidade seguinte, T_0 o período do oscilador silábico em repouso (sem acoplamento), $i(m)$ a magnitude do pulso do oscilador acentual que domina o grupo acentual m corrente (a magnitude e a posição desse pulso são geradas automaticamente, como demonstrado na seção seguinte) e $s(n)$ é a função de sincronismo exponencial (mais detalhes em BARBOSA, 2002), cuja taxa de subida é parametrizada pela força de acoplamento relativa entre os osciladores, w_0 . Os parâmetros a e b são as taxas de indução e decaimento, respectivamente. A forma geral de evolução da duração pela aplicação dessa fórmula pode ser vista na figura 3. Observe como no grupo acentual menor, que culmina na quarta posição, a duração da unidade VV cresce mais rapidamente, devido à proximidade do pulso do oscilador acentual (atribuído à quarta posição com magnitude inferior àquele atribuído à décima posição), do que a duração no segundo grupo acentual, que culmina na décima posição. Há dez posições nesse enunciado simulado porque, para

compará-lo com o natural, impomos a realização dos trechos entre parênteses na legenda da figura 3 como unidades VV únicas, pois ou não foi possível separar a vogal [a] do ditongo anterior, ou a vogal [i] da preposição foi realizada como semivogal.

$$\Delta T_n = \alpha \cdot T_n \cdot s(n) \cdot i(m) - \beta \cdot (T_n - T_0) \cdot i(m-1) \quad (1)$$

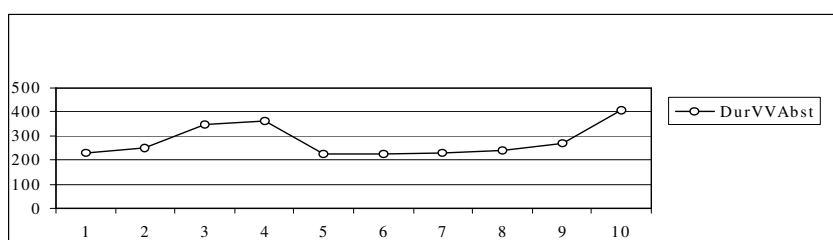


FIGURA 3 – Evolução da duração abstrata (DurVVAbst) em milissegundos das unidades VV do enunciado “Sub(iu a) tribuna d(e um) poleiro d(e our)o.” Cf nota 2 para os parâmetros do MDR.

Convém salientar que a modificação dos parâmetros da equação 1 gera uma grande variabilidade de padrões de duração, de acordo com as possibilidades reais de enunciação, a partir dos seguintes mecanismos: (1) um controle global da taxa de elocução afetando todos os segmentos do enunciado, através do parâmetro T_0 ; (2) um controle fino da taxa de aumento e redução da duração da unidade ao longo do grupo acentual, que afeta essencialmente as unidades VV frasalmente acentuadas e em sua vizinhança imediata, através das taxas de indução e decaimento; (3) um controle da influência do mecanismo de acentuação frasal sobre as durações silábicas simulando tendências a ritmo silábico ou acentual, através do parâmetro w_0 ; (4) um controle da posição e magnitude da proeminência prosódica, que afeta retrospectivamente a duração da unidade VV ao longo do grupo acentual, através do parâmetro $i(\cdot)$, a magnitude do pulso do oscilador acentual.

Explicada a maneira de obter as durações abstratas de unidades VV ao longo do grupo acentual, é necessário mostrar como obter automaticamente a posição e magnitude das proeminências prosódicas, especificadas abstratamente por pulsos de um oscilador acentual, que tende a oscilar a uma frequência regular, embora perturbável por informação sintática no MDR.

2.4. Obtendo de forma integrada a posição e magnitude das proeminências frasais pelo acoplamento entre sintaxe e restrições de produção/percepção da fala

A atribuição automática da posição do pulso do oscilador acentual, e, conseqüentemente, da posição do acento frasal, considerado no MDR como um pico local de duração abstrata da unidade VV, é realizada no interior de uma janela de *lookahead* expressa em número de palavras fonológicas, por questão de plausibilidade psicolinguística (NOOTEBOOM, 1995). A extensão dessa janela é calculada segundo a equação 2:

$$(\mu_{DurGA} + \sigma_{DurGA})/T_0 / \mu_{nVVporPalFon} \quad (2)$$

Na equação, (m_{DurGA}, s_{DurGA}) são a média e o desvio-padrão do período em repouso do oscilador acentual, isto é, da duração típica do grupo acentual (menos variável que o número de unidades VV no grupo) com valores fixados em (1300, 600) ms, calculados pela média dos quatro locutores paulistas para as três taxas de elocução. O parâmetro T_0 é o mesmo apresentado anteriormente. Já $\mu_{nVVporPalFon}$ é o número médio de unidades VV por palavra fonológica, que é 3,1, calculado a partir de um *corpus* de 353 palavras, o *corpus* Pantanal (cf. BARBOSA, 2006a, p. 494-495). O resultado da aplicação da equação 2 é arredondado para o número inteiro mais próximo. Para as simulações realizadas com o MDR, a extensão da janela de *lookahead* variou de duas a quatro palavras fonológicas, o que corresponde à frequência de atribuição de acentos frasais em enunciados naturais, visto que um único acento frasal é atribuído no interior da janela.

Para obter posição e proeminência dos pulsos do oscilador acentual, pulsos síncronos ao *onset* de vogal da sílaba tônica da palavra lexical de uma determinada palavra fonológica, começa-se a partir da primeira unidade VV do texto. Em seguida, atribui-se um pulso do oscilador acentual a uma das unidades VV lexicalmente acentuadas compreendidas na janela de *lookahead*. A posição da unidade VV lexicalmente acentuada a que se atribui o pulso é aquela que tem maior verossimilhança na referida janela. O valor dessa verossimilhança de acento frasal reflete a magnitude da proeminência prosódica. Essa verossimilhança v é calculada a partir da equação 3 a seguir, em que $\text{logit } p = \ln [p/(1-p)]$.

$$v(\text{acento frasal}) = \text{logit}(p_s) \cdot r_p + (1 - r_p) \cdot \text{logit}(p_b) \quad (3)$$

O parâmetro r_p , limitado entre 0 e 1, é o grau de acoplamento relativo entre restrições sintáticas e biomecânicas, que é modificado para simular estilos distintos de elocução, incluindo taxas de elocução que exigem uma maior participação da sintaxe. Assim, quando o valor se aproxima de 1, maior a influência da sintaxe na atribuição de proeminência prosódica. Quando se aproxima de 0, cada vez menor a influência da sintaxe e maior a da regularidade das proeminências. As duas probabilidades condicionais p_s e p_b são, respectivamente, (1) a probabilidade condicional de ocorrência de acento frasal, dada a fronteira sintática em que se encontra a posição candidata a ser posição de acento frasal, chamada de probabilidade condicional de natureza sintática, e (2) a probabilidade condicional de ocorrência de acento frasal, dada a distância em número de unidades VV desde o último acento frasal atribuído, chamada de probabilidade condicional de natureza biomecânica. A vantagem dos três parâmetros da equação 3 é a possibilidade de variação intra- e interindividual na atribuição de acento frasal.

Mostrei em Barbosa (2006a, p. 255-271) que é possível estimar probabilidades condicionais da ocorrência de acento frasal em uma palavra fonológica, dado um marcador sintático de uma gramática de dependência (cf. TESNIÈRE, 1967). Para tanto é necessário dispor de um texto relativamente longo (superior a 350 palavras), em estilo e gênero semelhantes àquele que se simula. A extensão do texto garante a relativa robustez dos valores de probabilidade encontrados. Para ilustrar, a análise de um texto de 353 palavras (BARBOSA, 2006a, p. 268), lido por um locutor brasileiro, revelou, para uma fronteira como aquela entre “pagagaio” e “real” no texto de Lobato, uma probabilidade condicional de 26 %, significativamente inferior à probabilidade *a priori* de acento frasal de 48 %, o que significa que o sujeito evita colocar acento frasal entre palavras em que a primeira é um regente não-verbal e a segunda é seu dependente. Já para uma fronteira entre frases (separadas por ponto final no texto), a probabilidade condicional é de 86 %, significativamente superior à probabilidade *a priori* de acento frasal de 48 %.

Para cálculo da probabilidade condicional de natureza biomecânica, p_b , parte-se do histograma da extensão dos grupos acentuais em número de unidades VV. A função lognormal é a que melhor caracterizou a distribuição para os quatro locutores paulistas e, portanto, dois descritores estatísticos, médio e desvio-padrão da extensão, bastaram para caracterizar a distribuição teórica da extensão do grupo acentual. Embora os locutores difiram quanto aos valores desses descritores, a variabilidade intra-individual é maior que a interindividual, pois as

maiores alterações se dão para leituras mais rápidas ou mais lentas no mesmo locutor. Valores típicos são 1,9 (0,4), que correspondem aproximadamente a 7 (2) unidades VV por grupo acentual (cf. MILLER, 1956, para limites de armazenamento de itens na memória de curto termo).

A obtenção da posição e magnitude do acento frasal e, isomorficamente, do pulso do oscilador acentual é a primeira etapa para a geração da organização temporal do enunciado e, portanto, encerra a apresentação do funcionamento do MDR. Cabe agora apresentar uma simulação em taxa mais rápida que a apresentada nas simulações anteriores, discutindo juntamente com isso a necessidade de se falar de sinergia de parâmetros no MDR e de sinergia de estratégias prosódicas na enunciação quando se fala mais rapidamente.

3 Falando mais rapidamente no MDR

Falar mais rapidamente não implica uma redução homogênea e global das durações dos fones ao longo dos enunciados naturais. Não apenas porque os fones não são regidos pelas mesmas leis de compressibilidade (a vibrante simples, [ʔ], é o menos variável dos segmentos, tendo sua duração praticamente inalterada mesmo quando falamos bem rapidamente). Sobretudo, porque alteramos o acoplamento entre sintaxe e restrições do sistema de produção (BARBOSA, 2006a), porque aumentamos o número de unidades VV subjacentes no grupo acentual ao mesmo tempo em que diminuimos a duração média da unidade VV (MEIRELES, 2007), sem contar no aumento considerável dos processos fônicos de redução (ditongações, coalescências, coarticulações extremas). A conjugação dessas estratégias com o fim da realização de uma meta recebe o nome de sinergia nas TSD.

No MDR, portanto, se quisermos simular uma taxa de elocução mais elevada, não basta reduzir o valor do especificador subjacente da taxa de elocução (T_0); é preciso também modificar, de acordo, os demais parâmetros do modelo. A simulação mais próxima à taxa de elocução rápida do locutor AP foi obtida com três dos parâmetros do modelo modificados para $m, s(\ln_{nVV}) = (1,9; 0,4)$, $r_p = 0,9$ e $T_0 = 150$ ms. Observe que, além da redução de T_0 , uma maior participação da sintaxe ($r_p = 0,9$ em vez de 0,6) e uma maior extensão do grupo acentual em valor logarítmico ($m \ln_{nVV} = 1,9$ em vez de 1,8) se impõem. Essas mudanças indicam a presença da sinergia.

4 O modelo dinâmico de ritmo da fala: vantagens e desafios

O MDR, que apresentamos em outros momentos (BARBOSA, 2002, 2006a), fundamenta-se em vários pressupostos das TSD, fios de Ariadne da pesquisa que apresentam visíveis vantagens teórico-experimentais para a compreensão do mecanismo de enunciação, da sua constituição na criança até sua fase adulta, incluindo todos os processos fônicos e as formas de alteração desse mesmo mecanismo, tanto em situações corriqueiras (ditado, narração de jogos, entrevistas, palestras, leitura) quanto em situações inesperadas, como as decorrentes de patologias congênitas ou adquiridas.

Parte desse trabalho está sendo conduzido pelo Grupo de Estudos de Prosódia da Fala,³ no Instituto de Estudos da Linguagem, Unicamp, abrangendo temas complexos como o acento secundário, a variação rítmica sob o efeito da taxa de elocução, o estudo da organização entoacional, bem como a estruturação rítmica na disartria e a laringalização na produção do portador de fissura palatina. Todos esses estudos pressupõem o MDR apresentado na figura 4 e cujos princípios de funcionamento foram descritos nas seções anteriores.

O acento secundário e o mecanismo prosódico pelo qual ele se manifesta em PB, a partir de um movimento de acentuação frasal, está sendo estudado à luz das TSD por Pablo Arantes, tanto na produção quanto na percepção da fala, com resultados que confirmam a hipótese de apenas dois osciladores acoplados no MDR, visto que não houve evidência de acento atribuído binariamente (ARANTES, 2005; ARANTES; BARBOSA, 2002, 2006). O estudo de percepção, conduzido de forma original ao ressaltar o vínculo com a produção (BERTHOZ, 1997) avaliando o comportamento do sujeito sem impor-lhe questões metalingüísticas, apresenta experimentos que exploram as características dinâmicas da percepção da fala.

A variação da estruturação rítmica do PB foi estudada por Alexsandro Meireles, tanto a partir de dados acústicos quanto, de forma inédita no Brasil, a partir de dados articulatórios para fins de estudos, no quadro da prosódia articulatória, do efeito do aumento da taxa de elocução para a organização temporal do enunciado (MEIRELES, 2007). O trabalho mostra uma tendência à conservação da extensão temporal do grupo acentual em todas as taxas, além de conduzir estudos de variabilidade entre indivíduos, gêneros e dois dialetos do PB. Esse trabalho aponta ainda que o aumento da taxa de elocução pode ser um mecanismo que conduza a uma mudança lingüística no caso de paroxitonização de proparoxítonas.

A organização entoacional de uma língua tem sido estudada mais recentemente no quadro das TSD, como mostram os trabalhos de Yi Xu (2005). É também no interior desse quadro que procuramos propor (1) um sistema notacional para a entoação do PB, tentando estabelecer um elo entre especificação abstrata e contornos de f_0 , bem como (2) uma análise dinâmica desses mesmos contornos para fins de síntese da entoação. O primeiro estudo tem sido conduzido por Luciana Lucente, que propôs uma série de oficinas para treinamento na marcação entoacional⁴, enquanto o segundo foi começado por Leandro Silveira (BARBOSA *et al.*, 2005). No MDR já esboçamos a inserção da geração automática da curva entoacional (BARBOSA, 2006a, p. 436-440).

A tese de doutorado de Jussara Vieira caracteriza a estruturação rítmica da fala, tanto longitudinal quanto vertical, de uma locutora disártrica por traumatismo craniano falante do PB a partir de cerca de nove descritores, incluindo a taxa de elocução, articulação, número e duração de pausas silenciosas e a distribuição de proeminências prosódicas. Esses descritores permitiram a definição de uma distância rítmica entre suas leituras ao longo de nove anos. O estudo da interface prosódia-sintaxe é também empreendido e comparado entre a locutora disártrica brasileira e uma locutora disártrica francesa. A despatologização de diversos aspectos da disartria é evidente (VIEIRA, 2007).

Aveliny Lima (2006) iniciou um projeto de doutorado que pretende dar conta dos fenômenos de glotalização (laringalização e oclusiva glotal) na fala de fissurados palatinos, a partir de uma redefinição dos gestos articulatórios glotais da Fonologia Articulatória (BROWMAN; GOLDSTEIN, 1989, 1992) segundo as propostas de Gafos (2002) e da Fonética Cognitiva (TATHAM, 1994).

A regularidade e a estruturação, que são características complementares na enunciação, são características imanentes do modelo, não terminam ao nível do grupo acentual, inferior ao nível da frase. A análise da fala semi-espontânea revelou que estão presentes na narrativa e na descrição e que se conjugam com a hesitação, uma das características mais comuns da fala espontânea, como demonstrou o trabalho de mestrado de Sandra Merlo (2006). Nesse trabalho, Merlo mostra que a regularidade dessas hesitações é variável entre os indivíduos e que ciclos hierárquicos relacionados a diversos níveis do planejamento do dizer estão presentes (LEVELT, 1989).

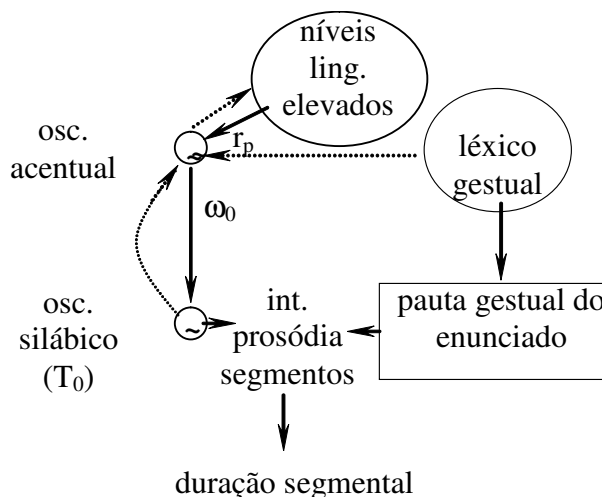


FIGURA 4 – O modelo dinâmico do ritmo da fala

O diagrama do MDR na figura 4 mostra, à esquerda, o modelo de osciladores acoplados e seu vínculo com (1) a sintaxe e a semântica a partir da relação do oscilador acentual com esses níveis lingüísticos superiores, (2) o léxico, considerado como gestual segundo a proposta de Browman e Goldstein (1989), em que cada item dá informações de posição de acento primário e de seu papel nas relações sintáticas, (3) os gestos no léxico dispostos numa pauta (*gestural score*) para dar conta da interação entre prosódia e segmentos (cf BARBOSA, 2006a, capítulo 5).

No MDR, as vogais constituem o fluxo portador (*carrier flow*) da enunciação, informando identidade de vogais no movimento de abertura, e de consoantes no movimento de fechamento das oscilações mandibulares e linguais. Esse fluxo, ancorado nos *onsets* das vogais, identifica-se com uma redefinição de silabidade, ligada à produção do enunciado por ciclos do tamanho da sílaba. Os *onsets* de vogais lexicalmente acentuadas são atratores dos pulsos do oscilador acentual.

O padrão sistemático de aumento, seguido de queda brusca (*reset*) da duração de unidades VV, evidencia a deceleração progressiva para a realização do acento frasal. Esse mecanismo, que chamamos de acentuação, serve o ouvinte ao assinalar a coesão entre constituintes prosódicos adjacentes, bem como a proeminência prosódica (via magnitude dos picos locais de duração).

Pelo MDR mostra-se claramente que não há e não pode haver isomorfismo entre sintaxe e prosódia, pois essa última, apesar de hierárquica, tem uma arborescência quase-simétrica com constituintes de extensão semelhante, por conta das restrições do sistema de produção. A prosódia é, assim, produto de uma espécie de negociação, sinergia, entre a informação sintática e as imposições do sistema de produção/percepção da fala.

A análise da organização temporal demonstrou que a duração normalizada da unidade VV é, para o PB ao menos, uma variável dependente quase-isomórfica à coesão entre constituintes prosódicos e à proeminência prosódica: quanto maior essa duração, maior a proeminência do trecho contendo a unidade longa e menor a coesão prosódica com o constituinte à sua direita.

O sucesso das predições do MDR em atribuir acentos frasais (BARBOSA, 2006b) mostra o acerto em teorizar a interface sintaxe-prosódia como acoplamento probabilístico entre a regularidade da extensão do grupo acentual e a coesão sintática local entre palavras fonológicas consecutivas (dadas por uma gramática de dependência).

O MDR dá conta da organização temporal em pelo menos três escalas temporais, a da duração segmental isomórfica ao gesto articulatório, a da duração do tamanho da sílaba e a da duração do grupo acentual, em diversas hierarquias prosódicas, suscetíveis de serem reveladas por análise estatística, como no enunciado natural.

O MDR introduz a necessidade de uma duração subjacente, abstrata ou prosódica que separa a duração prosódica *stricto sensu* da segmental e, por isso, é abstrato. Por trabalhar no quadro de uma TSD, traz a convivência do discreto com o contínuo, da regularidade e da estruturação de forma não dicotômica. Para isso, no entanto, impõe a necessidade de redefinir acento, acentuação, métrica, silabidade e tipo rítmico (acentual e silábico). O primeiro é ponto de culminância da acentuação. Os demais são variáveis e gradientes, não-dicotômicos.

Por fim, no MDR, a prosódia é uma solução funcional (sinérgica) do sistema comunicativo entre a regularidade na produção, que serve quem fala pela facilitação do controle (menos graus de liberdade), e a estruturação na percepção, que serve quem ouve pela transmissão da mensagem apropriada ao contexto. O discreto e o contínuo estão presentes o tempo todo e o comportamento do sujeito depende da tarefa. Um dos desafios do MDR é justamente iluminar a forma pela qual conhecimento discreto e contínuo se acoplam e conduzem a enunciação.

Outros desafios para o MDR são (1) o tratamento dinâmico da entoação pela inclusão de um oscilador glotal e da forma de acoplamento com os demais componentes do modelo. Esse oscilador deve reproduzir a curva de f_0 , bem como simular os fenômenos glotais condicionados prosodicamente, (2) a perturbação do funcionamento do modelo para dar conta de alterações na enunciação, como nas patologias que temos estudado, (3) o tratamento do componente semântico de forma a dar conta de efeitos particulares no enunciado, como a ênfase, acelerações e decelerações de curto termo, entre outros, e (4) a introdução de informações discursivas, para dar conta de diálogos, e efeitos situacionais ligados a atitude, e outros componentes paralingüísticos.

Notas

¹ Texto correspondente à apresentação na mesa redonda *Aspectos teóricos, metodológicos e empíricos em prosódia experimental*, realizada em Belo Horizonte no dia 28 de novembro de 2006, por ocasião do *IXº Congresso Nacional/IIIº Congresso Internacional de Fonética e Fonologia*. A mesa foi coordenada por mim e contou ainda com a participação da Profa. Dra. Sandra Madureira, da PUC-SP. O trabalho recebeu apoio da Fapesp, projeto *Análise e modelamento dinâmicos da prosódia da fala* (05/02525-7), e da bolsa de produtividade em pesquisa do CNPq, projeto *Modelamento dinâmico da produção do ritmo e da entoação da fala* (300296/2005-3).

² Os demais parâmetros do MDR para essa simulação foram $a = 0,4$; $b = 0,7$; $w_0 = 0,8$; $\text{catalexe} = 1$, $(m,s)_{GA} = (1300,600)$ ms; $m,s(\ln_{NV}) = (1,8;0,4)$; $r_p = 0,6$.

³ O site do grupo está localizado na URL: <http://www.experimentalprosodybrazil.org>.

⁴ O site das oficinas está localizado na URL: http://br.geocities.com/bp_tobi/.

Referências Bibliográficas

ARANTES, P. *Acentuação secundária em português brasileiro: fundamentos de produção e percepção da fala*. Qualificação de Mestrado. Universidade Estadual de Campinas, Campinas, 2005.

ARANTES, P.; BARBOSA, P. A. Acentuação secundária em português brasileiro à luz de um modelo dinâmico do ritmo: um estudo piloto. *I Congresso Internacional de Fonética e Fonologia/VII Congresso Nacional de Fonética e Fonologia*. Belo Horizonte, 2002.

ARANTES, P. Secondary stress in Brazilian Portuguese: the interplay between production and perception studies. In: *Speech Prosody 2006 Conference*. 2006, Dresden, Alemanha. p. 73-76.

BARBOSA, P. A. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. In: *First ESCA Tutorial Research Workshop on Speech Production Modeling and Fourth Speech Production Seminar*. 1996, Autrans, France. *Proceedings*. p. 85-88.

_____. Generating duration from a cognitively plausible model of rhythm production. In: *Seventh European Conference on Speech Communication and Technology*, 2. 2001, Ålborg, Dinamarca. *Proceedings*. p. 967-970.

_____. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. *Cadernos de Estudos Lingüísticos* 43, p. 71-92, 2002.

_____. *Incursões em torno do ritmo da fala*. Campinas: Pontes. 2006a. 540 p.

_____. A Dynamical Model for Generating Prosodic Structure In: *Speech Prosody 2006 Conference*. Dresden, Alemanha, 2006b. p. 366-369.

BARBOSA, P. A.; BAILLY, G. Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication* 15 (1-2), p. 127-137, 1994.

_____. Generation of pauses within the z-score model, In: VAN SANTEN, J.P.H.; SPROAT, R.W.; OLIVE, J.P.; HIRSCHBERG, J. (Eds.) *Progress in Speech Synthesis*. New York: Springer-Verlag, 1997. p 365-381.

BARBOSA, P.A.; MEIRELES, A. R.; VIEIRA, J. M. Abstractness in speech-metronome synchronisation: p-centres as cyclic attractors. In: Proc. Ninth European Conf. on Speech Communication and Tecnology. Lisboa, Portugal, 2005. p. 1441-1444.

BERTHOZ, A. *Le Sens du mouvement*. Paris: Éditions Odile Jacob, 1997.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer*. 2005. Disponível em: <<http://www.praat.org/>>, versão 4.4.

BROWMAN, C. P.; GOLDSTEIN, L. M. Articulatory gestures as phonological units. *Phonology* 6, p. 201-251, 1989.

_____. Articulatory phonology: an overview. *Phonetica* 49, p.155-180, 1992.

CLARK, A. *Being there. Putting brain, body, and world together again*. Cambridge, Estados Unidos: The MIT Press, 1997.

CLASSE, A. *The Rhythm of English Prose*. Oxford: Blackwell, 1939.

CUMMINS, F. Entraining speech with speech and metronomes. *Cadernos de Estudos Lingüísticos* 43, p.55-70, 2002.

FERNANDES N. H. *Contribuições para uma análise instrumental da acentuação e da intonação do português*. 1976. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo.

FRAISSE, P. *La psychologie du rythme*. Paris: Presses Universitaires de France, 1974.

GAFOS, A. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20, p. 269-337, 2002.

HIRSCH, M. W. The dynamical systems approach to differential equations. *Bulletin of the American Mathematical Society* 11, p. 1-64, 1984.

KELSO, J. A. S. Elementary coordination dynamics, In: SWINNEN, S. *et al.* (Ed.) *Interlimb coordination: neural dynamical and cognitive constraints*. San Diego, Estados Unidos: Academic Press, 1994. p. 301-318.

_____. *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, Estados Unidos: The MIT Press, 1995.

KUGLER, P. N.; TURVEY, M. T. *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, Estados Unidos: Lawrence Erlbaum Associates, 1987.

LEVELT, W. J. M. *Speaking: from Intention to Articulation*. Cambridge, Estados Unidos: The MIT Press. 1989.

LIMA, A. M. A glotalização em sujeitos com fissura palatina, segundo a abordagem dinamicista. *Projeto de doutorado*. Universidade Estadual de Campinas, Campinas, 2006.

LOBATO, J. B. M. *A Menina do Narizinho Arrebitado*. São Paulo: Revista do Brasil, 1920.

MACNEILAGE, P. F. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, p. 499-511, 1998.

MACNEILAGE, P. F.; DAVIS, B. L. Acquisition of speech production: Frame, then content, In: JEANNEROD, M. (Ed.) *Attention and Performance XIII*. Hillsdale, Estados Unidos: Lawrence Erlbaum, 1990. p. 453-475.

MASSINI G. *A duração no estudo do acento e do ritmo em português*. 1991. Dissertação (Mestrado em Linguística) – IEL, Universidade Estadual de Campinas, Campinas.

MEIRELES, A. R. *Reestruturas rítmicas da fala no português brasileiro*. 2007. Tese (Doutorado em Linguística) – IEL, Universidade Estadual de Campinas, Campinas.

MERLO, S. *Hesitações na fala semi-espontânea: análise por séries temporais*. 2006. Dissertação (Mestrado em Linguística) – IEL, Universidade Estadual de Campinas, Campinas.

MILLER, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review* 63, p. 81-97, 1956.

NOOTEBOOM, S. G. How far do we look ahead while speaking? In: *XIIIth International Congress of Phonetic Sciences*, 4. Estocolmo, Suécia, 1995. *Proceedings*. p. 578-581.

PORT, R.; GELDER, T. van (Ed.). *Mind as motion. Explorations in the dynamics of cognition*. Cambridge, Estados Unidos: The MIT Press, 1995.

STROGATZ, S. H.; STEWART, I. Coupled Oscillators and Biological Synchronization. *Scientific American Magazine*. December 1993.

TATHAM, M. The supervision of speech production – an issue in speech theory. *Institute of Acoustics*, 6. 1994. *Proceedings*.

TESNIÈRE, L. *Éléments de syntaxe structurale*. Paris: Éditions Klincksieck, 1965.

THELEN, E.; SMITH, L. B. *A dynamical systems approach to the development of cognition and action*. Cambridge, Estados Unidos: The MIT Press, 1994.

VIEIRA, J. M. *Para um estudo da estruturação rítmica na fala disártrica*. 2007. Tese (Doutorado em Linguística) – IEL, Universidade Estadual de Campinas, Campinas.

XU, Y. Speech melody as articulatorily implemented communicative functions. *Speech Communication*. v. 46, p. 220-251, 2005.