

MODELO DINÂMICO DA FALA: ENTOAÇÃO, RITMO E DISCURSO NO PORTUGUÊS BRASILEIRO

Lucente, Luciana*

Universidade Federal de Minas Gerais

Abstract

This article explores the relationship between intonational patterns and its relationship with speech rhythm and discourse, according to the dynamic systems research program. The study of these relationships were based on Barbosa's (2006) Dynamic Model of Speech Rhythm; on Dato intonational annotation system proposed by Lucente (2008); and on the Computational Model of the Structure of Discourse, proposed by Grosz & Sidner (1986). The Dynamic Model of Rhythm suggests that the speech rhythm is the result of the action of two oscillators – accentual and syllabic - which receive as input linguistic and gestural information, and give the output as gestural duration. This article hypothesis is that in addition to these oscillators, a glottal oscillator can act controlling the intonation patterns of speech. These patterns, or intonational cycles, which organize the BP intonation, emerge when related to the spontaneous discourse segmentation. For each discourse segment classified as spontaneous according to a criteria proposed in this article, the speech is segmented into the DaTo system in linguistically structured units, which contains the purposes of communication and attention. Each of these segments is aligned to the speech intonation pattern delimited by a rising contour (LH or > HL) at the beginning and by a falling contour (LHL), or a boundary level (L), at the end. The speech rhythm is also aligned to the pattern formed between intonation and discourse. By the inclusion of a new layer for the stress groups segmentation into DaTo system was possible to observe that the alignment between the stress groups segmentation and the intonational annotation coincide with discourse segments boundaries. The alignment between intonation, rhythm and discourse, having the stress groups as attractors, allowed us to propose the insertion of a glottal oscillator into the Dynamic Model of Rhythm.

Key-words: *Phonetics; Speech Prosody; Discourse and dialogue; Intonation*

Introdução

O objetivo deste artigo é mostrar como os processos de estruturação do discurso e da entoação estão ligados por meio um modelo que compreende não somente a entoação, mas também o ritmo da fala.

É comum a pesquisa em análise do discurso se concentrar em relações históricas-sociais e na produção de sentido por parte do sujeito. Também são conhecidas análises que se referem à prosódia da língua para explicar a realização de foco e diferentes intenções no discurso. No entanto, estas pesquisas muitas vezes se referem à prosódia ou à entoação de forma impressionística, sem que haja uma análise fonética que evidencie o que ocorre na relação prosódia-discurso.

O percurso deste artigo mostra como através da notação entoacional e da segmentação discursiva de um corpus de fala espontânea foram encontradas evidências para o fato de que as estruturas discursivas, rítmicas e entoacionais do português brasileiro (doravante PB) estão organizadas em torno de um atrator comum e tem seu funcionamento explicado por um modelo dinâmico de produção da fala.

1. Entoação e Modelo Dinâmico da Fala

A entoação comumente aparece relacionada a aspectos físicos do contorno entoacional. Em um sentido estreito ela “*se refere ao uso de características fonéticas suprasegmentais para expressar significados pragmáticos no nível da sentença de forma linguisticamente estruturada*” (Ladd, 1996, p. 6) e é um dos componentes da prosódia, termo mais amplo que inclui também a descrição de características dinâmicas e temporais, associadas aos aspectos formais (Barbosa, 2009).

As características suprasegmentais a que se refere Ladd (1996) são a frequência fundamental (doravante f_0), intensidade e duração, os correspondentes de ordem fonética dos fenômenos psicofísicos, de *pitch*, volume (*loudness*) e duração (*length*) (Lehiste, 1970), embora não numa relação de uma para um. Dentre essas características, f_0 é o correlato direto do aspecto fonético que a entoação assume nos estudos prosódicos. A f_0 é medida em Hz (*Hertz*) e corresponde ao número de vezes por segundo em que as pregas vocais completam um ciclo de vibração. Esses ciclos são controlados pelos músculos da laringe que determinam a tensão nas pregas vocais, como também por forças aerodinâmicas do sistema respiratório sublaríngeo (Botinis et al, 2001). Essa definição de entoação que envolve elementos físicos e psicofísicos, ou fonéticos e fonológicos, na sua produção faz com que esta seja vista como um componente dinâmico da fala.

Na concepção clássica de Saussure (2002), a fala se contrapõe ao conceito de língua. Enquanto a língua é um produto de natureza social não constituindo uma função do falante, a fala se constitui como um ato individual de vontade e inteligência. Como um ato individual, a fala possui caráter sincrônico, refletindo fenômenos e usos em um dado momento, enquanto a língua se volta predominantemente para a diacronia se pensarmos em seu caráter evolutivo. O conceito de entoação,

portanto, se liga à fala e à língua, expressando as funções comunicativas presentes na fala do indivíduo que estabelecem os padrões entoacionais que dizem respeito à língua.

Na produção da fala a entoação se manifesta como a forma mais produtiva de expressão das intenções do falante. Através da manipulação de f_0 é possível atribuir ênfase a um determinado item, determinar se o que está sendo dito é uma asserção, uma interrogação, uma interrogação parcial, determinar foco, pressuposição e status da informação, além de várias outras funções pragmáticas e comunicativas.

1.2. O programa dinâmico

A perspectiva dinâmica nas ciências teve seu início a partir dos estudos da motricidade humana e animal. As pesquisas de Kugler & Turvey (1987), Thelen & Smith (1994) e Kelso (1994; 1995) sobre motricidade em geral, revelaram a existência de uma coordenação funcional e auto-organizada de movimentos que pressupõe o abandono das dicotomias mente-corpo e planejamento-execução em detrimento da existência de sistemas dinâmicos atuando no controle desses movimentos.

Um exemplo típico de coordenação entre movimentos e sistemas sensoriais é o movimento executado pelos gatos para saltar de uma árvore ou sobre alguma plataforma. Da mesma forma, a capacidade humana de saltar, andar e aprender a falar são exemplos dessa coordenação. Ao observarmos a figura 1 a seguir, podemos acompanhar os sucessivos movimentos que um gato¹ executa para saltar e cair seguramente sobre as quatro patas. Os movimentos registrados pelo fotógrafo Eadweard Muybridge² em sua clássica obra *Animal Locomotion* (1887) também são exemplos da organização envolvida na movimentação dos homens e dos animais em um espaço de tempo, como se observa na figura 2.

¹ Fotografia estroboscópica de Kim Taylor.

² Fotógrafo inglês nascido em 1830, conhecido por seus experimentos fotográficos utilizando múltiplas câmeras a fim de captar a movimentação de homens e animais em diferentes situações. Seu trabalho é considerado precursor na criação da película de celulóide usada no cinema.

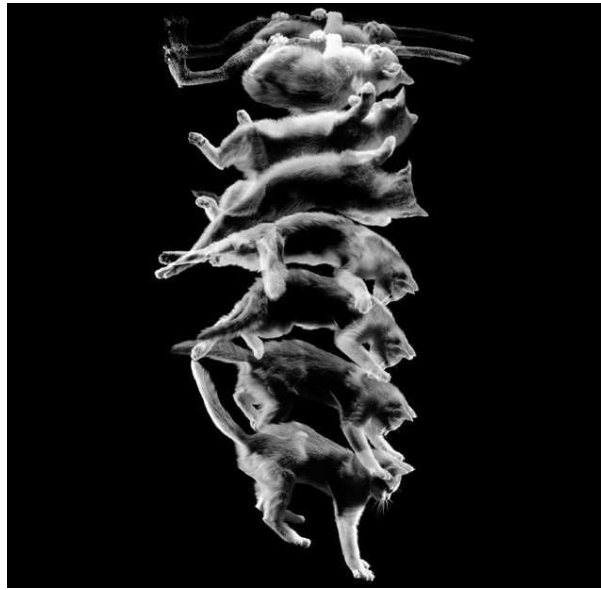


Figura 1: gato saltando; durante o salto o gato ajusta seu movimento para atingir a tarefa de cair em pé.

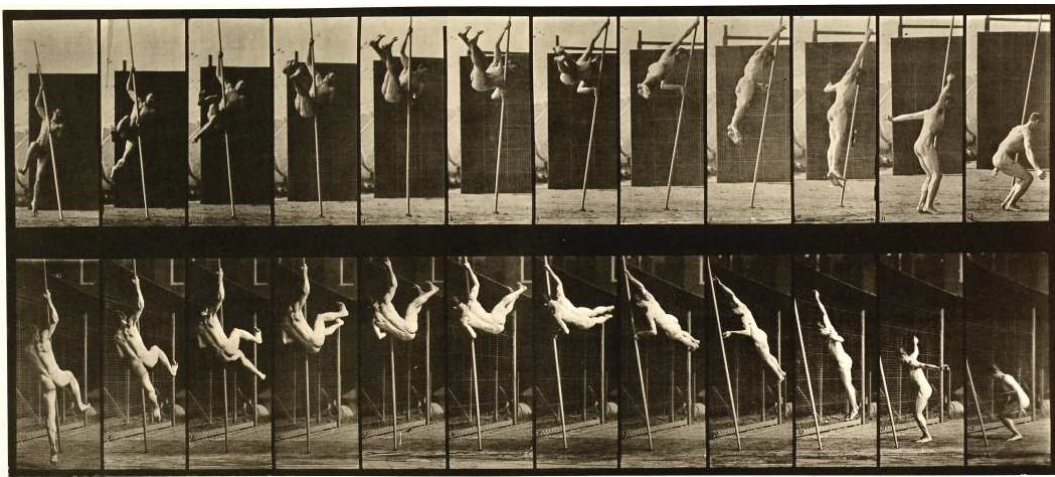


Figura 2: (*Animal Locomotion. Plate 165*) da mesma forma que o gato, o homem coordena seus movimentos para a realização de uma tarefa específica.

A execução perfeita de movimentos no espaço é possível pela ação de um sistema dinâmico, que compreende um sistema físico composto por um número de diferentes componentes que interagem entre si e com o ambiente de uma forma auto-organizada, e que muda seu estado inicial com a passagem do tempo (Port e Van Gelder, 1995; Barbosa, 2006).

O conceito de auto-organização se refere a uma tendência espontânea não linear de ordenamento observada em sistemas complexos, que podem ser naturais ou artificiais. Esse ordenamento, ou interação, deve-se a características intrínsecas ao próprio processo desses sistemas complexos. Tais sistemas são sensíveis às suas condições iniciais de partida, quando estão imersos em

um regime caótico e à sua relação com o meio ambiente. Exemplos de auto-organização podem ser encontrados na natureza na forma de sistemas químicos, físicos, biológicos e sociais, em certos tipos de magnetismo, na estrutura de um furacão, na organização de colmeias, de movimentos sociais, na coloração de espécies animais (Lucente, 2007) ou mesmo na organização dos sons e da entoação da fala.

A relação de auto-organização pressupõe o abandono das dicotomias mente - cérebro e planejamento - execução em detrimento de ações coordenadas. No entanto, para o controle dessas ações o número de variáveis a serem controladas deve ser reduzido e se correlacionar por meio de uma sinergia³ (Kelso, 1995), permitindo que o sistema exiba propriedade de adaptação. Dessa relação resulta um padrão de funcionamento que age de acordo com uma frequência interna e que é capaz de alterar seu estado inicial ao longo do tempo. Esse comportamento possibilita ao observador a capacidade de previsão de eventos futuros a partir da observação do estado inicial desse sistema e a previsão deste em um estado futuro, como por exemplo, os saltos do gato e do homem nas figuras 1 e 2.

A coordenação entre as diferentes partes desse sistema dinâmico auto-organizado possibilita a emergência de parâmetros de controle que podem influenciar o sistema como um todo. Quando ocorre algum tipo de modificação, mesmo que suave, em algum destes parâmetros, pode ocorrer em seguida o que chamamos de bifurcação no sistema. Em termos de sistemas dinâmicos, uma bifurcação é uma transição brusca em relação ao estado anterior em que se apresentava o sistema. Transições desse tipo podem ser de origem intrínseca ou a partir de uma perturbação imposta externamente ao sistema. Quando há uma perturbação no sistema é que se torna possível determinar a relação funcional entre os parâmetros de controle, os quais podem ser identificados nesses momentos de transição (Kelso, 1995).

Por apresentar uma relação estreita com estados temporais, um sistema dinâmico, além de ter bifurcações, como propriedades que ocorrem ao longo do tempo, pode dispor também de atratores cíclicos (caso seja um sistema dinâmico periódico) que são padrões de movimento oscilatório para os quais o sistema tende com o passar do tempo ou com a realização de uma tarefa específica, ou ainda, quando sujeito a alguma perturbação externa (Barbosa, 2006).

Segundo essa perspectiva dinâmica, além da coordenação de movimentos que possibilitam a um sujeito a realização de uma tarefa específica, existe ainda uma tendência biológica à coordenação de movimentos entre sujeitos ou entre sujeito e ambiente, fato que pode ser exemplificado pelos experimentos de Kelso (1984) e de Barbosa et al (2005).

³ Do grego *sunergia*, 'cooperação, ajuda, assistência', o termo designa para a fisiologia a ação associada de dois ou mais órgãos, sistemas ou elementos anatômicos ou biológicos, cujo resultado seja a execução de um movimento ou a realização de uma função orgânica

Se esta coordenação possibilita um mecanismo de controle do movimento como abordado até o momento, na produção da fala, mais especificamente na entoação, esta hipótese também pode ser aplicada, pois esta sincronia entre elementos auto-organizados pode forçar os movimentos articulatórios que possibilitam o alcance de um alvo específico na curva entoacional a coincidirem com certos ciclos articulatórios recorrentes. Existem evidências, por exemplo, de que a sílaba sirva como estrutura coordenativa com a qual vários movimentos articulatórios estão alinhados (Fujimura, 2000). Segundo Xu (2005), isso significa que pode ser difícil para um falante manter uma relação de fase entre movimentos de f_0 e sílaba senão por uma sincronia que revela um mecanismo atrator.

1.3. Modelo Dinâmico do Ritmo

Dentro do Programa Dinâmico de Investigação Científica (doravante PDI) (cf., Lakatos, 1978), o modelo que se ajusta à análise entoacional deste trabalho é o modelo dinâmico do ritmo da fala (doravante MDR). O MDR, proposto por Barbosa (2006), é uma implementação matemático-computacional de um sistema dinâmico do ritmo da fala que exhibe todas as propriedades de um sistema auto-organizado (Barbosa, 2006) como (i) mudanças de estado dos sistemas com o passar do tempo, (ii) previsões sobre mudanças de estado em momentos futuros, (iii) sinergia e acoplamento funcional entre os componentes do sistema, e (iv) a presença de atratores cíclicos.

De acordo com esses pressupostos, o MDR recebe como entrada níveis linguísticos elevados (sintaxe, semântica, discurso) e também um léxico gestual, segundo a definição da Fonologia Articulatória (Browman & Goldstein, 1986), e gera como saída a duração gestual (Barbosa, 2006). Ou seja, o modelo em si é composto pela (i), interação entre a informação linguística contida nos níveis superiores e um oscilador acentual, (ii), pelo acoplamento de dois osciladores linguísticos abstratos (acentual e silábico) regidos por uma força de acoplamento, e (iii), por um mecanismo de acoplamento entre o oscilador silábico e a pauta gestual (originada do léxico gestual). A interação entre oscilador silábico e pauta gestual gera a duração gestual. A figura 3 mostra o diagrama do MDR.

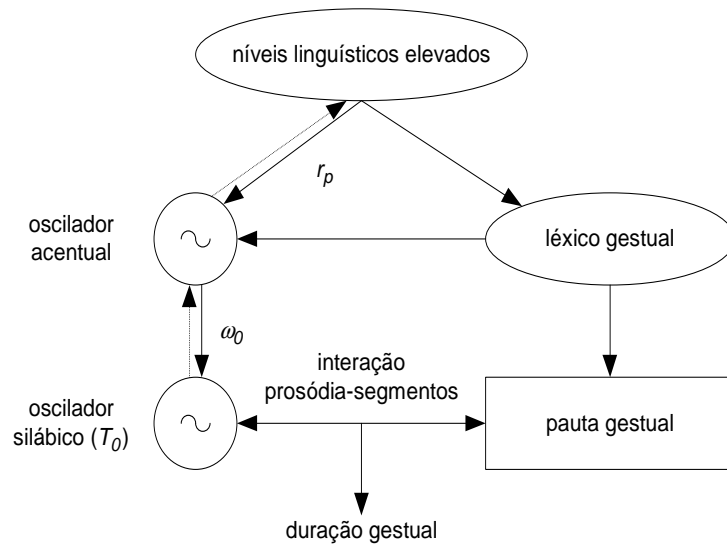


Figura 3: Diagrama do Modelo Dinâmico do Ritmo da Fala

Nesse modelo, o oscilador acentual recebe informações especificadas de forma abstrata no léxico gestual, que determina a posição dos acentos lexicais e faz com que os pulsos do oscilador acentual ocorram nas sílabas marcadas lexicalmente como tônicas (Barbosa, 2006). Essa sucessão de pulsos especifica os locais onde ocorrem os acentos frasais isomorficamente, ou seja, obedecendo a um mesmo padrão. Desta forma, o oscilador acentual gatilha no oscilador silábico um mecanismo de acentuação universal ao longo dos ciclos do oscilador silábico, que gera os padrões de duração de unidades V-V (vogal a vogal)⁴ ao longo do enunciado, assinalando os acentos frasais (Barbosa, 2006).

O acento frasal, definido por um pico de duração local no enunciado, assinala uma proeminência do domínio da produção da fala, baseada apenas na duração dos segmentos. A interação entre esta proeminência, enquanto um fator prosódico, e a pauta gestual, gera a duração gestual. Barbosa (2006, p.11) destaca que a ênfase entoacional, obtida com a excursão de f_0 em determinado local, acarreta aumento de duração nos segmentos vocálicos e é mais saliente perceptivelmente do que apenas o aumento das durações. Esse alinhamento entre f_0 e duração dá pistas sobre como o acento frasal pode influenciar não somente a estrutura rítmica dos enunciados do PB, mas também a estrutura entoacional.

⁴ Uma unidade V-V, ou vogal-a-vogal, se refere à segmentação de unidades acústicas entre os *onsets* de duas vogais seguidas.

Com base no alinhamento entre fenômenos entoacionais, duração e estrutura discursiva do PB observados em dados de fala espontânea, este trabalho propõe a inserção de um novo oscilador ao MDR: o oscilador glotal, como mostra a figura 4. A hipótese da existência de um oscilador glotal acoplado aos osciladores silábico e acentual oferece uma resposta sobre como elementos prosódicos se organizam no discurso, ou como o discurso se organiza sobre características prosódicas da língua.

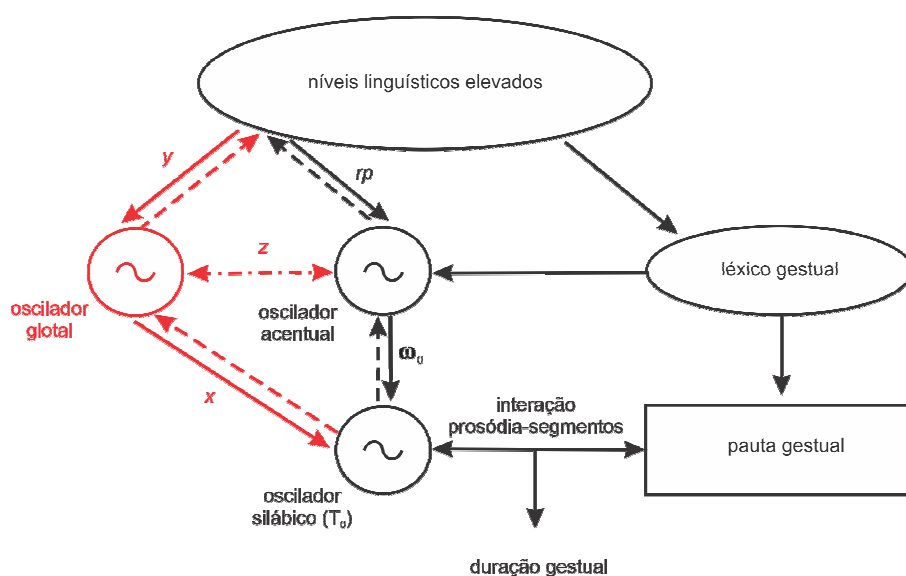


Figura 4: Diagrama do Modelo Dinâmico do Ritmo da Fala com o acréscimo de um oscilador glotal.

2. CORPUS DE FALA

O corpus de fala utilizado neste trabalho foi o *Corpus VoCE*, acrônimo formado pelos nomes da rádio **V**ocê AM, do site **C**iência Hoje **O**nline, e do programa de rádio **F**im de **E**xpediente⁵, dos quais foram obtidas as gravações diretas dos programas de rádio e de *podcasts*.

O Corpus VoCE está em composição desde 2006 (Lucente e Barbosa, 2007) e conta com aproximadamente 10 horas de gravações no total, contendo exemplos da fala de profissionais (jornalistas ou radialistas) e de participantes dos programas. Os 46 trechos selecionados para análise a partir do corpus somam aproximadamente 32 minutos de duração e são exemplos de fala de 23 sujeitos diferentes, sendo estes nove mulheres e 14 homens, com faixa etária entre 20 e 60 anos⁶. Desses exemplos foram excluídas as falas dos profissionais, por apresentarem registro com marcas de locução profissional, que não é o foco do presente trabalho.

⁵ Respectivamente: <http://www.radiovoce.com.br>; <http://cienciahoje.uol.com.br/podcasts>; <http://cbn.globoradio.globo.com/programas/fim-de-expediente/FIM-DE-EXPEDIENTE.htm>

⁶ Ver detalhes em Lucente (2012, p. 17).

Os trechos de fala selecionados apresentam em sua grande maioria exemplos da fala paulista e carioca, pois as rádios e a revista de onde foram extraídas as gravações estão localizadas em São Paulo (Rádio Você AM, Americana - SP; Programa Fim de Expediente gravado em São Paulo, capital) e Rio de Janeiro (Revista Ciência Hoje *On Line*, produto da Fundação Oswaldo Cruz, Rio de Janeiro, capital). Informações sobre os sujeitos, como sexo, faixa etária e escolaridade foram obtidas a partir do reconhecimento dos falantes, que são em grande parte figuras públicas.

A estrutura e a composição do Corpus VoCE se assemelham às de outros corpora compostos por gravações radiofônicas, como por exemplo o *Boston University Radio News Corpus* (Ostendorf et al, 1996), que foi bastante utilizado em pesquisas fonético-entoacionais do inglês (Hasegawa-Johnson et al, 2005; Abner, 2009; Rosenberg & Hirschberg, 2009). Ambos os corpora classificam seus dados como sendo exemplos de fala espontânea, devido ao fato destes não terem sido obtidos em laboratório. No entanto, tal classificação dos dados de fala como sendo espontâneos merece especial atenção, como se verá a seguir na proposta de classificação adotada.

2.1. Classificação da fala espontânea: gênero e controle

As definições adotadas por Xu (2010) e Beckman (1996) sobre fala espontânea e de laboratório deixam de abordar, ou abordam sem desenvolver suas implicações, dois pontos essenciais para a definição da fala espontânea, que são (i), o grau de controle do experimentador sobre a obtenção dos dados, e (ii), o papel dos gêneros discursivos.

A fala espontânea não pode ser considerada sinônimo de conversa livre, pois um ditado, uma leitura, uma narrativa ou uma entrevista podem ser considerados espontâneos se ocorrerem em situação natural de comunicação. Da mesma forma, fala de laboratório não é sinônimo de total ausência de espontaneidade, pois dependendo da forma como os sujeitos são eliciados estes podem alcançar bastante espontaneidade na fala. O que irá definir a dimensão que caracteriza a espontaneidade é o grau de intervenção do experimentador na obtenção dos dados. Essa perspectiva de que graus de controle podem determinar a naturalidade da fala já foi abordada por Lehiste (1963), onde a autora coloca que os dois requisitos, naturalidade e controle, ocupam dois extremos de uma escala gradual de exclusão mútua, ou seja, em um extremo podemos encontrar a completa naturalidade da fala somada a completa ausência de controle, e no outro extremo o controle absoluto sobre a produção da fala à custa da naturalidade e da relevância linguística.

Sendo assim, a intervenção, ou controle do experimentador sobre a obtenção dos dados vai de um controle extremo – situação em que o experimentador diz como o sujeito deve produzir a fala – até a ausência total de controle – fala dita sem nenhuma indicação de produção, em qualquer que seja o registro.

Dessa maneira, a partir de uma relação ortogonal entre esses dois fatores – gênero e controle –, foi possível propor uma forma efetiva de classificação dos dados espontâneos e aplicá-la ao Corpus VoCE.

De acordo com a classificação proposta na figura 5, no eixo gênero, estão classificados os gêneros discursivos de forma não ordenada, pois não é possível ordená-los segundo uma hierarquia em que um gênero apresente mais naturalidade em relação a outro, pois cada um pode ser realizado de acordo com uma escala de naturalidade própria ao gênero.

A escala de naturalidade se realiza no eixo do controle que classifica o grau de intervenção de um experimentador na obtenção dos dados, sendo as extremidades desse eixo marcadas pela ausência completa de controle e pelo controle extremo do experimentador.

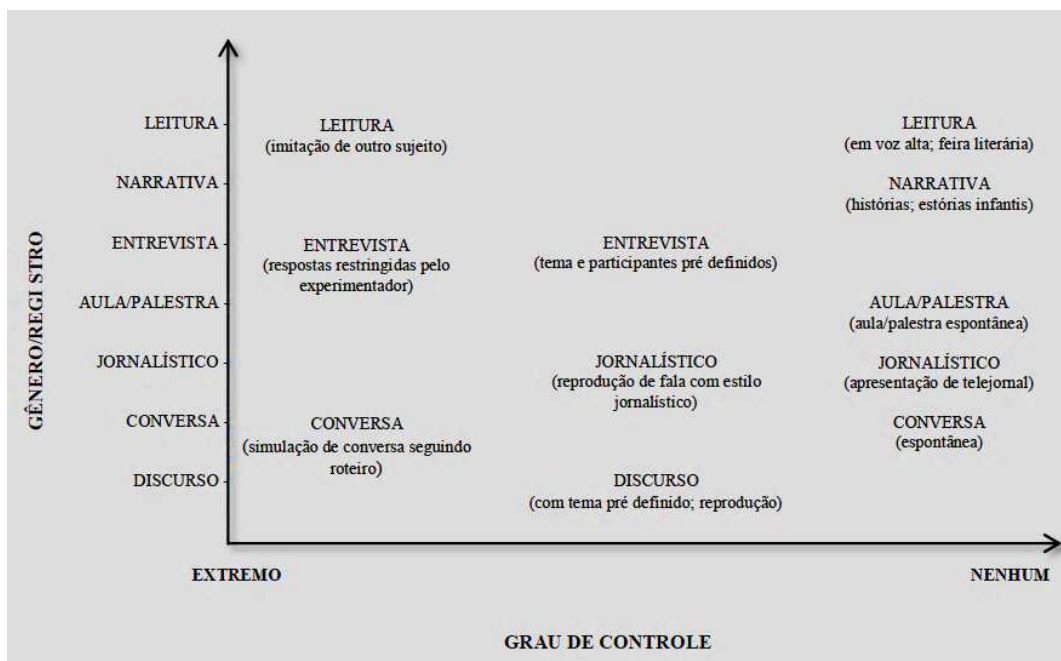


Figura 5: classificação dos gêneros discursivos em relação ao eixo que mede o grau de controle

De acordo com essa classificação, o gênero leitura, que é comumente tido como não espontâneo, pode ser classificado como espontâneo se a leitura em questão for realizada em um evento, uma feira literária, etc., como classificada na figura 5.

Essa proposta de classificação tem o objetivo de evitar interpretações diferentes acerca da espontaneidade da fala se baseando apenas nos gêneros e não considerando a intervenção externa no momento de sua produção.

Tendo analisado o Corpus VoCE e classificado suas produções como espontâneas, passamos a analisar o conteúdo discursivo dos exemplos contidos no corpus.

3. Estrutura do Discurso

A análise da fala e da entoação presentes neste trabalho parte de uma abordagem discursiva que compreende as partes que compõe o discurso⁷. Esta análise, que segue uma teoria computacional sobre a estrutura discursiva (Grosz & Sidner, 1986), se aproxima também de uma perspectiva textual-interativa da conversação (Jubran & Koch, 2006), na qual se assume uma inter-relação cognitiva entre os participantes de uma conversa, tendo como objeto de estudo o texto, nesse caso, o texto falado, o qual será denominado, daqui por diante, por *discurso* apenas.

A análise textual-interativa da conversação, inserida na linguística textual, leva em consideração, além do aspecto cognitivo, aspectos recorrentes da subjetividade dos falantes, como a produção de sentido, por exemplo. A abordagem do discurso utilizada aqui procura explorar apenas os aspectos formais e funcionais que levam à produção do discurso e seu entendimento por parte dos interlocutores, ou seja, um processo cognitivo que se encerra quando a comunicação acontece, não sendo relevante para a análise discursiva a interpretação da produção de sentido por parte do sujeito.

3.1. Estrutura do Discurso - Modelo de Grosz e Sidner

A segmentação discursiva dos dados do Corpus VoCE foi feita de acordo com a Teoria da Estrutura do Discurso ou Modelo de Grosz e Sidner (*G&S's model* – modelo G&S em PB), proposta em Grosz & Sidner (1986). O modelo G&S foi desenvolvido a princípio como uma teoria computacional que, ao descrever a estrutura do discurso, oferece as bases necessárias para a descrição e seu significado. De acordo com as proponentes desse modelo, a descrição da estrutura do discurso desempenha um papel central no processamento da linguagem à medida que estipula restrições nas porções do discurso (Grosz & Sidner, 1986). Essa descrição é intimamente relacionada a duas questões: *o que distingue o discurso*, e *o que o faz coerente*? A tentativa de resposta a essas questões leva a dois aspectos não linguísticos fundamentais no desenvolvimento desse modelo, que são a *atenção* e a *intenção* dos participantes de uma conversa. A atenção é um fator essencial no processamento de enunciados em um dado discurso, enquanto a intenção desempenha um papel importante na explicação de como se estrutura o discurso, proporcionando coerência a este e ao próprio termo “discurso” (Grosz & Sidner, 1986, p. 175).

A hipótese defendida nesse modelo é de que qualquer discurso é composto por três elementos essenciais e distintos, mas que interagem entre si a todo momento, que são: i) a estrutura sequencial dos enunciados do discurso em um dado momento; ii) a estrutura das intenções envolvidas no discurso; e iii) o estado de atenção dos participantes envolvidos no discurso. Paralelamente, existem ainda dois fatores importantes na constituição desse modelo, que são os papéis dos participantes em uma conversa e o significado contido no discurso.

⁷ Será adotado aqui o termo “análise discursiva” e não “análise do discurso”, pois a análise apresentada não se insere na perspectiva da Análise do Discurso enquanto área do saber.

Assumindo que o discurso seja uma peça do comportamento linguístico que envolve múltiplos enunciados e participantes, ele pode ser produzido por um ou mais participantes desse evento, pois em uma conversa, por exemplo, mais de um dos envolvidos pode falar e compor diferentes enunciados do discurso. Portanto falar de falante e ouvinte, ou locutor e interlocutor, é inviável, pois em uma situação de conversação não é claro o papel que cada participante desempenha. Dessa forma, a modelo G&S utiliza os termos ICP – *initiating conversational participant* – e OCP – *other conversational participant* – para distinguir esses participantes. Em PB usaremos respectivamente participante iniciador da conversa (PI) e participante não iniciador da conversa (PNI).

Sobre o significado ou sentido do discurso é importante ressaltar que este é um problema ainda não resolvido, e que não é o foco desse modelo. O objetivo dessa proposta é oferecer um modelo adequado sobre a *estrutura* do discurso, para que se possa desenvolver uma teoria adequada sobre o *significado* do discurso baseada nessa estrutura.

Sendo assim, podemos dizer que o discurso possui três componentes responsáveis pela sua estruturação e interação: i) estrutura da sequência dos enunciados, ou estrutura linguística (*linguistic structure*); ii) a estrutura dos propósitos, ou estrutura das intenções (*intentional structure*); e iii) o estado do foco de atenção, ou estado de atenção (*attentional state*). Juntos, esses três constituintes da estrutura do discurso suprem a informação necessária para que os participantes da conversa (PI e PNI) possam determinar como um determinado enunciado se encaixa com as outras partes do discurso, possibilitando que os participantes entendam por que algo foi dito e o que isso significa (Grosz & Sidner, 1986, p. 177), sem que se toque em aspectos sobre o significado do discurso como um todo.

3.1.1. Estrutura Linguística

A estrutura linguística, entendida no modelo G&S como a estrutura dos enunciados que compõem o discurso, é responsável por agregar tais enunciados em segmentos de discurso (doravante DS – *discourse segment*⁸).

O desafio desse modelo é a forma de delimitar os segmentos de discurso, pois embora diferentes pesquisas em diferentes teorias (Cohen & Levesque, 1980; Polanyi & Scha, 1986) tenham analisado uma variedade de gêneros de discurso, não houve concordância suficiente sobre como delimitar as fronteiras dos segmentos de discurso. Existem várias pistas linguísticas que apontam para tais fronteiras, como pistas prosódicas tais como pausas, alongamentos (Chafe, 1979, 1980), taxa de elocução (Butterworth, 1975) e pistas textuais, como o uso de subordinações, coordenações, pronomes, referentes e marcadores discursivos (Hirshberg & Litman, 1993).¹ Serão usadas siglas de alguns dos termos em inglês por uma facilidade de assimilação com o material sobre o tema já publicado em inglês.

⁸ Serão usadas siglas de alguns dos termos em inglês por uma facilidade de assimilação com o material sobre o tema já publicado em inglês.

Essas pistas textuais podem indicar inclusive as mudanças na estrutura das intenções ou o estado atencional do discurso. O exemplo de segmentação dos enunciados no quadro 1 a seguir, extraído do Corpus VoCE, é ilustrativo de como tais expressões linguísticas (em itálico colorido) funcionam como marcadores de fronteiras discursivas.

- 0 - 1 - (1) *mas hoje em dia* por exemplo eu to lançando o u::: um meu disco novo no formato dum livro também
- (2) *então* vem o livro e o CD juntos
- (3) *que* é o Sambazz :::
- 2 - (4) *e o Samba/* e o livro explica a produção do disco
- (5) *então* além de ter o disco ainda tem o livro :
- 3 - (6) *que* eu escrevi todo ::: falando da produção do disco desd' a::: das composições até a finl/ finalização dele :::
- (7) *também* pra despertar um interesse na pessoa [assim] não simplesmente em baixar as o disco na in/ na internet de graça e :: ouvir as músicas ::
- (8) *mas* também pro cara : ter o:: ter a vontade de ler o livro ::: e:: e saber da história do disco *né?*

Quadro 1: Exemplo de expressões linguísticas como marcadores de fronteiras discursivas

3.1.2. Estrutura das Intenções

A estrutura das intenções compreende os propósitos que subjazem ao discurso e seus componentes e suporta a distinção entre os propósitos fundamentais ao discurso.

Entre os participantes de um dado discurso existe mais do que um único objetivo que os leva a participar de uma conversa. A distinção desses objetivos, ou intenções, é fundamental para o entendimento do discurso. Cada intenção que subjaz a um discurso em particular é chamada de propósito do discurso (doravante DP - *discourse purpose*).

De acordo com as proponentes do modelo, o propósito do discurso seria a intenção geradora do discurso. Essa intenção proporciona a razão do discurso ser produzido, assim como a razão do conteúdo particular do discurso ser sobreposta a qualquer outra informação (Grosz & Sidner, 1986). Sendo assim, cada segmento de discurso (DS) carrega uma intenção, que é chamada de propósito do segmento de discurso (doravante DSP - *discourse segment purpose*). O DSP especifica como cada segmento contribui para alcançar o propósito final do discurso.

Na estrutura das intenções, um PI pode ter diferentes tipos de intenções que o levam a iniciar um discurso, no entanto essas intenções que servem como DPs ou DSPs se distinguem de outras intenções por terem o intuito de serem reconhecidas, enquanto outras intenções são privadas, e não explicitadas nos DPs ou DSPs. O reconhecimento dos propósitos (DPs e DSPs), por parte dos participantes, é essencial para alcançar o efeito pretendido pelo PI. Dentro dessa estrutura existem

ainda duas relações estruturais que desempenham um papel importante na estrutura do discurso. A essas relações o modelo G&S denomina: dominância e satisfação.

Por exemplo, uma ação que satisfaz uma intenção em um determinado DSPx pode ser destinada a prover parte da intenção de outra ação em um outro DSPy, precedente ou posterior. Nesse caso diríamos que DSPx *contribui* ou *satisfaz* com DSPy, e que portanto, DSPy *domina* DSPx, o que gera uma hierarquia de dominância na estrutura dos DSs. A relação de satisfação entre os DSs pode ser significativa e também observável em casos em que o DSPx satisfaz o DSPy, porém o DSPx tem que ser satisfeito antes de satisfazer o DSPy – a satisfação é um tipo de precedência que não exerce domínio sobre o DSP seguinte. O trecho segmentado em diferentes DSs no quadro 2 estabelece as relações entre seus DSPs descritas no quadro 3.

DS1 - (1) porque : tem uma con/ existia uma convicção de que o mundo ia acabar ::: né?

(2) no dia que concid/ coincidisse isso :::

DS2 - (3) só que assim eh ::: a superstição o que que a superstição é?

(4) ela nada mais é do que aquela velh' insegurança né?

(5) então eu acredito em tudo :: eu todo mundo o que todo mundo me fala eu eu : coloco em :::

DS3 - (6) porque a a superstição ela vem :: de uma pessoa que num pensa ::: né?

(7) se voce for parar e pensar como é que um número :: pode ter poder e força em cima de você? :::

DS4 - (8) agora o que qu' ele tem poder e força?

(9) porque na bíblia fica/ na bíblia sagrada católica apostólica romana fala :: que o número meia meia

meia : é o número da besta

DS5 - (10) só que o número da best' ele num tem zero meia zero meia de dois mil e seis :

(11) é o número meia meia meia ::

(12) é diferente :: né? :::

DS6 - (13) então é por causa disso que acredito que há :: essas superstições

Quadro 2: segmentação hierárquica do discurso em diferentes DS.

DSP1 SATISFAZ DSP2	Existe uma superstição
DSP2 DOMINA DSP3 DSP2 DOMINA DSP4	A superstição é uma crença, um tipo de insegurança
DSP3 SATISFAZ DSP4	Se você pensar vai entender que um número não tem poder sobre você
DSP4 DOMINA DSP5	A bíblia cita o número 666 e não 6 do 6 de 2006
DSP2 SATISFAZ DSP6 DSP3 SATISFAZ DSP6 DSP4 SATISFAZ DSP6 DSP5 SATISFAZ DSP6	A superstição existe por esses motivos

Quadro 3: relação de dominância e satisfação entre os DSPs do quadro 2.

O modelo G&S faz uma distinção importante entre a *determinação* e o *reconhecimento* de um DSP. A determinação se refere a uma relação semântica, que é a especificação completa do que é pretendido por alguém. O reconhecimento se refere a um processamento que permite a um participante do discurso identificar qual a intenção de outro (Grosz & Sidner, 1986). Esses conceitos são interrelacionados, pois a mesma informação que determina um DSP pode ser usada por um participante para reconhecer a intenção deste DSP (Grosz & Sidner, 1986, p.179).

3.1.3. Estado de Atenção

E por fim, o estado de atenção é definido como uma propriedade intrínseca do discurso e não dos participantes do discurso. Esse estado é (i) inerentemente dinâmico, fazendo um registro de objetos, propriedades e relações salientes em cada ponto do discurso (Grosz & Sidner, 1986) e (ii) modelado por um conjunto de espaços focais (doravante FS – *focus space*), que são mudanças que ocorrem no estado de atenção motivadas por um conjunto de regras de transição que especificam as condições para se acrescentar ou excluir esses espaços. De acordo com a modelo G&S, a soma de todos os espaços focais em um discurso determina a estrutura focal e o processo de manipulação desses espaços focais é chamado focalização. Os espaços focais incluem DSPs em seu escopo. Essa inclusão de propósitos no espaço focal mostra que os participantes da conversa focam não apenas o que está sendo dito, mas também porque isso está sendo dito (Grosz & Sidner, 1986).

O estado de atenção pode ser confundido muito facilmente com o estado cognitivo do participante, porém não é o caso. É importante esclarecer esta relação: o estado de atenção é apenas um dos componentes do que chamamos de estado cognitivo. O estado cognitivo compreende o estado interno do falante, que inclui conhecimento, crenças, desejos, emoções (raiva, medo, alegria), intenções, atitudes (paciência, impaciência, amizade) e atenção (*stress*, interesse, tédio). O estado cognitivo também inclui informações sobre o tema da conversa que pode ser expresso por meio do afeto (Scherer, 1984).

Dessa forma, o estado de atenção, por meio da estrutura focal, age de forma coordenada com a estrutura linguística e com a estrutura de intenções, ou seja, os DSs se ligam a seus respectivos espaços focais (FSs). Sendo assim, os primeiros DSs do discurso têm seus FSs menos acessíveis aos participantes, pois estes estão sendo sobrepostos pelo FSs dos últimos DSs a aparecerem no discurso por apresentarem propriedades, objetos e relações mais salientes aos participantes em comparação com os atributos dos primeiros DSs.

A sobreposição dos espaços focais reflete a saliência relativa das entidades em cada espaço durante a porção correspondente do segmento do discurso. Sendo assim, quando um DSP oferece uma nova intenção que se sobrepõe a outras na estrutura focal, espaços focais anteriores saltam da pilha de sobreposições quando esse novo FS é inserido. Portanto a estrutura focal – que compõe o estado de atenção – possui duas propriedades essenciais: i) ela parasita a estrutura de intenções, no sentido de que as relações entre os DSPs determinam a sobreposição dos FSs; e ii) assim como a estrutura intencional e a estrutura linguística, ela se desenvolve com o desenrolar do discurso, pois a estrutura das intenções, com raras exceções, se constrói pelos participantes com o progresso do discurso (Grosz & Sidner, 1986, p. 180).

Resumindo, a estrutura focal é o repositório para a informação contextual necessária para o processamento dos enunciados a cada parte do discurso. Ela distingue objetos, propriedades e relações que são salientes a cada momento e ainda oferece ligação para partes relevantes de ambas as estruturas linguística e de intenção. Durante o discurso uma crescente quantidade de informações é discutida, por isso é cada vez mais necessário ser possível identificar os segmentos relevantes para o discurso, as entidades que se fazem salientes e seus DSPs. O papel do estado de atenção em delinear a informação necessária para o entendimento é, portanto, central no processamento do discurso.

3.2. Aplicação do Modelo

O modelo G&S se aplica diretamente à análise da estrutura do discurso por meio da segmentação dos DSs e da construção de uma relação hierárquica entre esses segmentos.

O quadro 4 mostra a segmentação de um trecho de um discurso em 12 enunciados distintos, cujas fronteiras são determinadas pela presença de marcadores linguísticos (em itálico colorido). Cada um desses enunciados faz parte de um DS, que por sua vez possui um DSP, e este estabelece uma relação de satisfação ou dominância com outro DSP, por exemplo: DS2 e DS4 estabelecem relações

de dominância com DS3 e DS5, respectivamente. No entanto DS3 estabelece relação de satisfação com DS4.

- DS1** - (1) as outras visões ::: de criação do aparecimento das espécies ::: não são teorias científicas :::
(2) *dessa forma* é uma perda de tempo ficar comparando :::
(3) *o que* seria uma teoria científica ::: com uma teoria ::: não científica
- DS2** - (4) *eu* conheço ::: vários pesquisadores importantes :::
(5) *que* ::: tem as suas crenças religiosas :::
- DS3** - (6) *contudo* eles aceitam completamente a teoria evolutiva do Darwin :::
(7) *porque* não :: consideram ::: qu'exista incompatibilidade entre as duas :::
- DS4** - (8) *as teorias* são basicamente uma forma de ver o mundo ::
(9) *e* as teorias científicas são :: interessantes do ponto de vista de que ::: elas tem em período de duração ::
- DS5** - (10) *nesse aspecto* ::: a teoria do Darwin :: já tem cent'e cinquenta anos :::
(11) *isso* é um verdadeiro record ::
(12) *ele* praticamente só perde para teoria da gravidade de Newton

Quadro 4: relação hierárquica entre os DS que compõem o discurso

A relação exposta na tabela 1 é ilustrativa de como a teoria se aplica ao segmento discursivo apresentado no quadro 4. Cada um dos cinco DSs, formado por um conjunto de enunciados, tem a intenção (DSP), por parte do iniciador da conversa, de que os outros participantes acreditem ou aceitem suas proposições.

Seguindo o modelo G&S, o espaço focal do DSP5 assume a posição mais alta na posição dos FSs no momento em que ocorre no discurso, enquanto o espaço focal de DSP1 fica sobreposto por todos os outros. Nesse exemplo, ele inicia o discurso com foco sobre a validade da teoria do surgimento das espécies, e termina com foco direcionado para a robustez da teoria, ou seja, a intenção do discurso até atingir o FS5 vai sendo construída pela sobreposição dos espaços focais na atenção do PNI, e pela relação de dominância de cada uma das intenções dos DSPs.

DS	Enunciados	Intenções
DS1	as outras visões ::: de criação do aparecimento das espécies ::: não são teorias científicas ::: <i>dessa forma</i> é uma perda de tempo ficar comparando ::: <i>o que</i> seria uma teoria científica ::: com uma teoria ::: não científica	existe somente uma teoria científica válida para o aparecimento das espécies;
DS2	eu conheço ::: vários pesquisadores importantes ::: que ::: tem as suas crenças religiosas :::	existem pesquisadores com crenças religiosas;
DS3	contudo eles aceitam completamente a teoria evolutiva do Darwin ::: porque não :: consideram ::: qu'exista incompatibilidade entre as duas	as crenças religiosas não são incompatíveis em aceitar a teoria do Darwin como válida;
DS4	as teorias são basicamente uma forma de ver o mundo :: e as teorias científicas são :: interessantes do pon'de vista de que ::: elas tem em período de duração ::	teorias tem períodos de duração (e podem ser falseadas);
DS5	nesse aspecto ::: a teoria do Darwin :: já tem cent'e cinquenta anos ::: isso é um verdadeiro record :: ele praticamente só perde para teoria da gravidade de Newton	a teoria darwiniana perdura por 150 anos, é uma teoria forte e não foi falseada até hoje

Tabela 1: composição dos DSs e seus DSPs

3.3. Elementos de Construção do Discurso

O modelo G&S, em sua aplicação, se utiliza de estratégias de construção do discurso que são empregadas na segmentação dos enunciados que compõem os DPs e também no reconhecimento das intenções dos participantes. As estratégias mais exploradas em Grosz & Sidner (1986) são as *interrupções* no discurso e o uso de *marcadores linguísticos*, como *marcadores discursivos*, *pronomes*, *coordenações*, *subordinações*, *referentes*, *pausas* e *hesitações*. Na abordagem dos elementos de construção do discurso que se segue, essas estratégias são ampliadas segundo a perspectiva textual-interativa da conversação. De acordo com a gramática de Jubran e Koch (2006), as interrupções de modo geral contêm, ou expressam, estratégias de construção do discurso como as hesitações, repetições, correções, parafraseamentos e parentetizações. Paralelamente às estratégias relacionadas à interrupção, há estratégias independentes de interrupções como a tematização/rematização, a referenciação e o uso de marcadores discursivos, a última também explorada em Grosz & Sidner (1986).

Os elementos de construção do discurso são parte integrante e fundamental dos segmentos discursivos, eles marcam pontualmente trechos do discurso e têm sua função comunicativa clara para as partes envolvidas em uma conversa. No modelo G&S, os elementos de construção funcionam como indicadores dos pontos em que os DSPs se segmentam e se organizam hierarquicamente seguindo um critério semântico.

Sendo assim, os elementos que compõem a estrutura do discurso no modelo G&S se utilizam de elementos ou estratégias de construção do discurso, que podem agir interrompendo o fluxo do discurso ou não. De acordo com esse ponto de vista, no discurso ou na fala, as ideias se estruturam em unidades hierárquicas, não necessariamente longas ou completas do ponto de vista sintático, mas perfeitamente analisáveis, pois possuem elementos construtores que agem possibilitando o entendimento e a análise.

4. Notação e Descrição Entoacional

Para a notação e descrição dos padrões entoacionais do Corpus VoCE, foi utilizado o sistema DaTo (acrônimo de *Dynamic Tones of Brazilian Portuguese Intonation*) de notação entoacional, proposto por Lucente (2008). O sistema DaTo propõe para notação entoacional um conjunto de contornos dinâmicos e de níveis de fronteira que compõem o quadro entoacional para a notação do PB. Paralelamente aos fenômenos entoacionais, o sistema DaTo apresenta segmentação manual da fala em segmentos funcionais e a segmentação semiautomática dos enunciados em unidades rítmicas.

O sistema DaTo adota uma perspectiva dinâmica na análise da entoação (Xu, 1999; 2005; 2006; Barbosa, 2006). Porém, por ter como foco uma notação entoacional baseada na avaliação auditiva de proeminências entoacionais, na observação visual da curva de f_0 e na notação manual, o sistema remete inevitavelmente à teoria métrica-autossegmental (Lieberman, 1975; Pierrehumbert, 1980), ou teoria MA, que tem como expoente o sistema ToBI (*Tones and Break Indices*) de notação entoacional, baseado no trabalho de Pierrehumbert (1980), desenvolvido inicialmente para a notação do inglês americano (Beckman et al, 1994; Silverman et al, 1992)

Ao contrário do que propõe a teoria MA, a notação DaTo segue uma perspectiva dinâmica ao analisar os fenômenos da entoação. Essa perspectiva dinâmica contrapõe-se a uma notação métrico-linear, como em ToBI, na medida em que procura não representar a estrutura tonal por meio de eventos tonais isolados, mas por meio de contornos entoacionais dinâmicos, que se relacionam ao longo da estrutura, e são determinados por aspectos fonéticos como duração, intensidade (o que Xu (2005) chama de *strength*) gama de variação tonal e alinhamento com o material linguístico (Lucente, 2008).

4.1. Teoria MA

De acordo com Ladd (1996), a teoria MA propõe-se a caracterizar adequadamente na estrutura tonal, elementos categoricamente distintos e promover o mapeamento dos parâmetros acústicos contínuos a partir de elementos fonológicos, ou seja, a partir de representações fonológicas.

Os fundamentos da teoria MA são: i) a linearidade da estrutura tonal; ii) a distinção entre proeminência (*pitch accent*) e acento lexical (*stress*); iii) a análise dos *pitch accents*⁹ em relação à altura dos tons; iv) e a relação entre características locais e tendências gerais.

Partindo da hipótese da linearidade da estrutura tonal, a teoria MA representa fonologicamente os contornos entoacionais como sequências de eventos entoacionais discretos (Ladd, 1996, p. 43). Sendo assim, a teoria assume claramente a distinção entre eventos e transições que ocorrem na estrutura tonal, reconhecendo partes da estrutura como linguisticamente importantes e outras como apenas transições preenchedoras entre eventos localizados. Os eventos considerados importantes do ponto de vista linguístico são os *pitch accents*.

Para essa teoria, as relações de proeminência são em primeira instância abstrações fonológicas, que se manifestam por meio de dois aspectos fonéticos distintos, o acento lexical (*stress*) e a posição das proeminências (*pitch accents*) (Ladd, 1996). Os *pitch accents* são eventos locais que assinalam proeminência no nível do enunciado, e estão associados na teoria MA à altura da curva entoacional nos pontos em que esses eventos ocorrem. A literatura sobre o assunto, predominantemente em inglês, destaca a diferença entre *pitch accent* e *stress*, ou seja, entre proeminência e acento lexical. Diferentemente da proeminência, o acento é uma propriedade lexical de sílabas individuais, assinalada no nível da palavra. A teoria MA assume que proeminência e acento lexical são fenômenos distintos, e podem ocorrer alinhados ou não na estrutura métrica do enunciado.

Além da distinção dos *pitch accents* em relação aos acentos lexicais, esses ainda tem relação direta com a altura dos tons. Apesar de a teoria MA considerar os *pitch accents* como os fenômenos mais importantes da curva entoacional, estes são analisados à luz da teoria apenas como sequências ou combinações de tons altos e baixos, sendo representados por meio dos rótulos H (*high*) e L (*low*), e pela combinação destes.

Seguindo esses pressupostos, o sistema ToBI (Beckman et al, 1994) apresenta uma sistematização mais econômica da proposta de Pierrehumbert (1980), ainda propondo uma notação que reflete a linearidade da estrutura tonal e a marcação dos *pitch accents*. Tal sistema foi desenvolvido por um grupo de pesquisadores com o intuito inicial de produzir uma ferramenta de transcrição que pudesse auxiliar no entendimento e no desenvolvimento de modelos computacionais de síntese e reconhecimento de fala que requerem grande quantidade de enunciados transcritos prosodicamente (Silverman et al 1992).

⁹ Para uma melhor adequação com a literatura será usado o termo *pitch accent*, em inglês.

4.2. Perspectiva Dinâmica

Enquanto a teoria MA se baseia em aspectos lineares da estrutura tonal, na identificação dos *pitch accents* e no alinhamento abstrato dessa estrutura com o material linguístico, o paradigma dinâmico-entoacional se concentra na convergência de aspectos fonéticos – velocidade, intensidade, altura, duração – da curva entoacional a fim de atingir um alvo ou desempenhar uma tarefa linguística por meio dos contornos entoacionais da gama de variação tonal e do alinhamento específico com o material linguístico.

Os contornos entoacionais, enquanto um dos primitivos melódicos desse paradigma é definido por uma *trajetória ideal da curva entoacional especificada por um alvo a ser atingido (pitch target) associado a uma unidade segmental, provavelmente uma sílaba* (Xu, 2005). Na especificação dessa trajetória da curva até alcançar seu alvo estão envolvidas, além dos parâmetros físicos, funções comunicativas.

O modelo PENTA (*Parallel Encoding and Target Approximation*) proposto por Xu (2005, 2006) ilustra como diferentes funções e parâmetros estão codificados nos contornos entoacionais. Nessa proposta, funções comunicativas de ordem lexical, frasal, focal, de tópico, etc, são codificadas em parâmetros físicos, como duração, gama de variação tonal e intensidade, que possibilitam o alcance do alvo e seu alinhamento com a sílaba na curva entoacional a uma altura e taxa de aproximação específicas (Xu, 2005; 2006).

Considerando a definição de contorno dinâmico e a proposta do modelo de Xu 2005;2006), a figura 6 ilustra o conceito de contorno dinâmico, que é definido aqui *como uma unidade tonal que contém elementos comunicativos expressos em uma trajetória ideal da curva entoacional, especificada por um alvo a ser atingido e associada a uma unidade segmental linguística*.

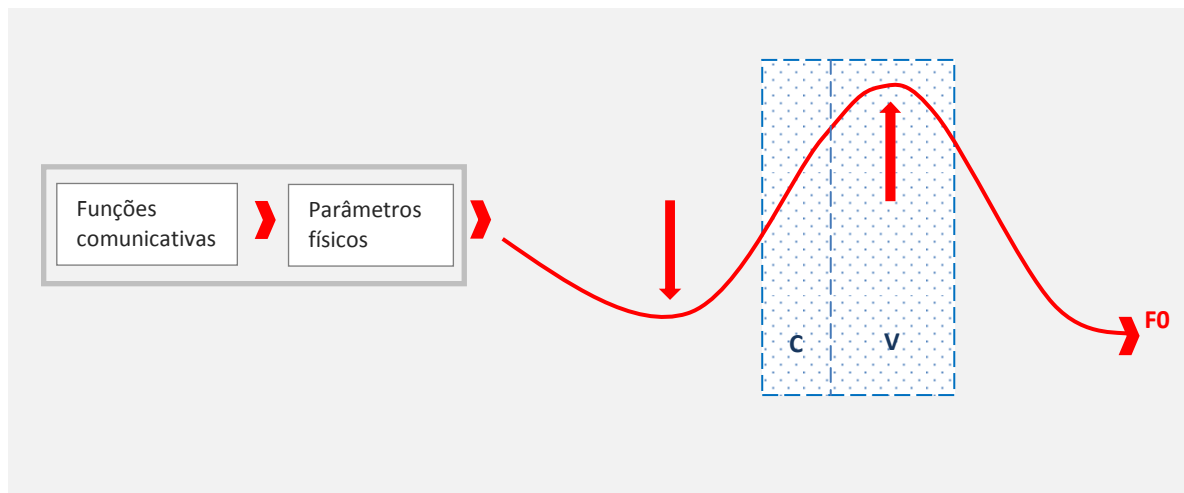


Figura 6: Em um contorno entoacional, as (i) funções comunicativas são codificadas em parâmetros físicos; expressas em uma (ii) trajetória ideal de f_0 com um alvo a ser atingido (setas); e (iii) alinhado a uma unidade segmental (sílabas tônicas)

A gama de variação tonal delimita o intervalo de *pitch* no qual um alvo é implementado, e pode ser definido pelos parâmetros do valor do alvo e intervalo de variação. A gama de variação tonal

é de grande importância para a especificação de foco, pois é de acordo com as variações do valor do alvo de *pitch* que antecedem e seguem o foco é que é possível percebê-lo. Essas variações de *pitch* podem compor um esquema de codificação do foco, ou mesmo constituírem propriedades intrínsecas a este (Xu, 2005). Da mesma forma, por meio das variações no valor do alvo de *pitch*, podemos determinar fenômenos como mudanças de tópicos na fala e distinções entre tipos de enunciados, como declarativas e interrogativas (Lucente, 2012).

Cada contorno dinâmico, marcador de proeminência, tem movimento e altura relacionados à altura e ao movimento de f_0 que o antecede e segue, e sua posição na estrutura tonal é determinada pelo alinhamento com o material linguístico, mais precisamente com as sílabas tônicas, ou seja, com o acento. O contorno dinâmico na notação DaTo não se refere à adição de movimentos com diferentes alturas de f_0 , mas a um movimento da curva entoacional ao qual se atribui marcação de proeminência. A figura 7 é ilustrativa de como cada movimento na curva se relaciona com o anterior, pois para cada movimento de subida ou descida na curva, o movimento anterior tem que proporcionar condições para que isso ocorra, ou seja, uma descida para que ocorra subida, e vice-versa.

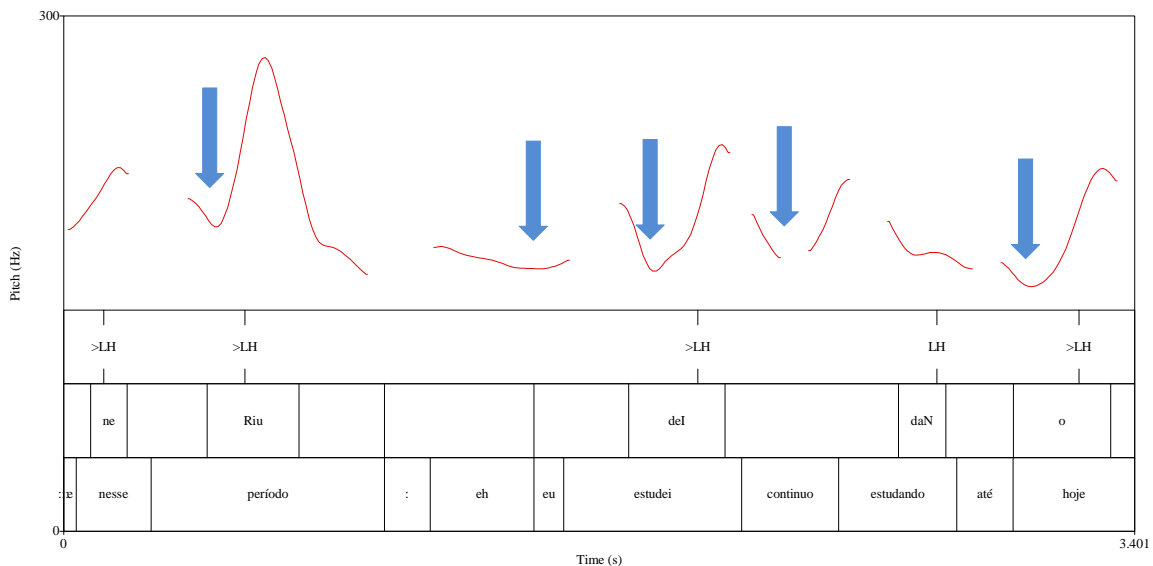


Figura 7: movimentos dinâmicos dos contornos entoacionais com as setas indicando os vales antes das subidas; na primeira camada a notação entoacional, na segunda a marcação das sílabas tônicas, na terceira a transcrição ortográfica.

Essa inter-relação entre os movimentos da curva entoacional defendida pela interpretação dinâmica dos fenômenos que compõem a estrutura entoacional fica evidente ao se extrair a função derivada da curva de f_0 . Ao realizar a extração da função derivada de f_0 por meio do *script F0_extrema.praat* escrito por Arantes (2009) pode-se observar a existência de um controle na produção de f_0 que faz com que esta coordene os movimentos da curva de acordo com a velocidade de subida e descida de f_0 e com o alvo a ser atingido. A figura 8 mostra essa relação: em cinza a curva

suavizada de f_0 , em vermelho estão indicados os pontos de mínima altura de f_0 , em verde indicados os pontos de máxima altura. Os pontos que indicam os picos de máxima velocidade de subida (R) e descida (F) de f_0 estão alinhados com a notação na segunda camada de notação da figura.

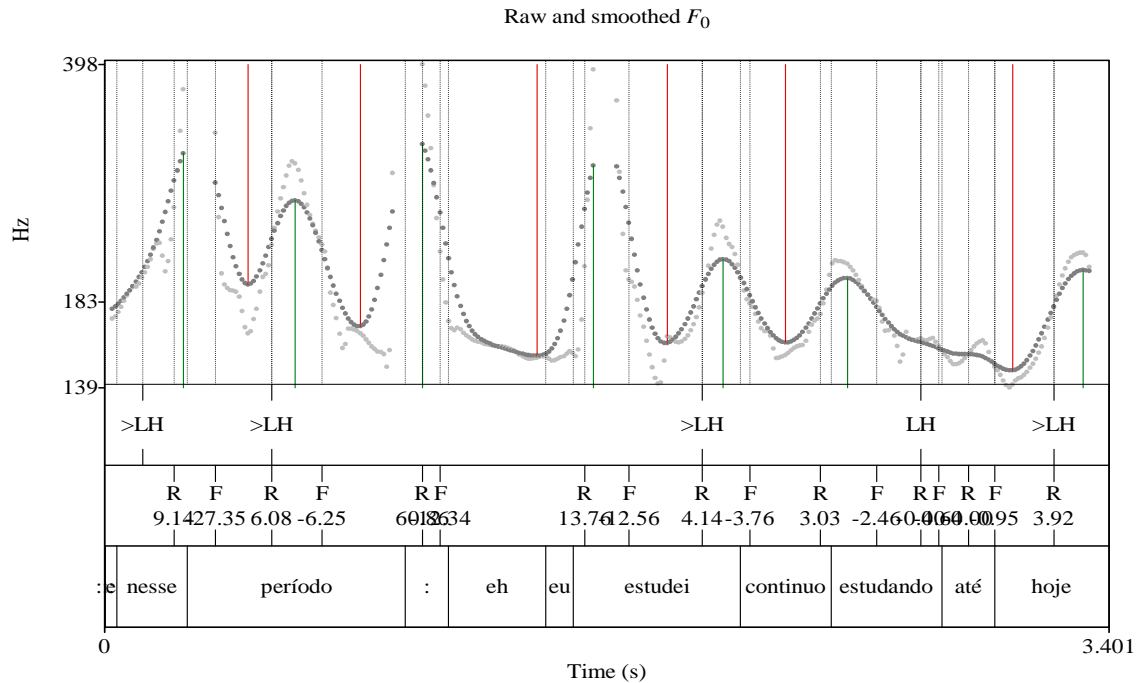


Figura 8: curva entoacional de enunciado; na primeira camada indicando os contornos entoacionais, na segunda os pontos de máxima velocidade de subida e descida de f_0 (linhas azuis) e na terceira a transcrição ortográfica.

O alinhamento específico entre contornos dinâmicos e o material linguístico compreende uma questão central na defesa desta abordagem, pois enquanto a teoria MA estabelece um alinhamento abstrato da curva entoacional cujo ponto específico é decidido por regras de implementação fonética (Pierrehumbert, 1980; 2000), a perspectiva dinâmico-entoacional estabelece um alinhamento específico (Atterer e Ladd, 2004) entre esses elementos, a ponto de este evidenciar a existência de uma sincronia entre os movimentos que produzem f_0 e movimentos articulatorios que geram os padrões espectrais, apesar de esses movimentos serem controlados separadamente. O fato de existir uma sincronia entre movimentos independentes não significa que haja uma total interdependência entre eles, pois estes podem estar limitados por determinados graus de liberdade na execução de movimentos concomitantes (Xu, 2005).

Os pressupostos e funcionamento da notação DaTo, dentro de uma perspectiva dinâmica da entoação, como contornos dinâmicos e níveis de fronteira, assim como os outros componentes da notação serão apresentados a seguir.

4.3. Notação DaTo

O sistema DaTo, além de apresentar uma notação dos fenômenos entoacionais do PB a partir de uma perspectiva dinamicista, procura também apresentar em suas camadas de notação informações pragmático-comunicativas e aspectos fonéticos do ritmo da língua. Para tanto, a notação completa se divide em seis camadas, das quais três são fixas e servem para a notação de: i) contornos dinâmicos, ii) segmentação da fala em unidades V-V, e iii) transcrição ortográfica; e outras três camadas, que podem ser omitidas ou apresentadas de acordo com o objetivo de cada notação, que são: iv) detecção automática de pausas/grupos acentuais, v) informação pragmática e vi) segmentação em unidades discursivas (DS).

4.3.1. Contornos Dinâmicos

Os contornos dinâmicos, como definido anteriormente, são movimentos contínuos de f_0 , que percorrem a transição de um ponto a outro na curva entoacional até atingir o seu alvo. Este movimento de transição carrega a informação tonal que antecede e segue seu alvo. Nessa concepção, um contorno dinâmico torna relevantes as transições consideradas pela teoria MA como apenas transições preenchedoras entre eventos estáticos. Estes contornos se dividem no sistema DaTo entre ascendentes, descendentes e níveis de fronteira, e são representados por rótulos que ilustram o movimento de f_0 pelas letras H (*high*) e L (*low*), assim como na teoria MA, por julgar tais rótulos produtivos e ilustrativos e, portanto, de mais fácil compreensão.

A notação entoacional do sistema DaTo considera o padrão de movimento e alinhamento da curva entoacional com as sílabas tônicas partindo de uma inspeção visual da curva entoacional e da percepção de ênfase. Diferentemente, a teoria MA pressupõe inicialmente a percepção de outiva dos *pitch accents* e posteriormente sua verificação na curva. Um dos pressupostos do treinamento para notação nos sistemas que seguem a teoria MA é a concordância entre sujeitos sobre o tipo de movimento tonal que conseguem detectar.

Os *contornos ascendentes* são os mais frequentes na entoação do PB e estão constantemente associados à marcação de focos estreitos. Estes contornos têm como referência para o alinhamento da curva com a sílaba tônica da palavra focada a posição alta (H) do contorno, e apresentam um padrão de movimento determinado por características acentuais que antecedem e sucedem o foco.

Antecedendo o foco, estes contornos apresentam uma descida de f_0 durante as sílabas pretônicas, considerada obrigatória para a realização do movimento posterior, que é a subida de f_0 para a realização do movimento ascendente. Em posição pós-tônica verifica-se para todos os contornos ascendentes um *resetting*. Esse *resetting* mostra que a percepção do foco está associada, além do emprego de maior energia na sua realização, à oposição ocorrida com as sílabas não acentuadas posteriores ao foco. Estas características mostram estratégias do sistema fonador para produção e percepção do foco.

Os *contornos descendentes* seguem a mesma perspectiva dos ascendentes, porém tendo como referência para o alinhamento com a sílaba tônica a posição baixa (L). O movimento *default* de f_0 para estes contornos se inicia após uma subida suave de f_0 , que pode anteceder determinado contorno ou mesmo estar presente dentro dos limites deste, ocorrendo em seguida uma queda que se alinha à sílaba tônica da palavra focada. O *resetting* das sílabas pós-tônicas ocorre da mesma forma que nos contornos ascendentes, confirmando uma estratégia do PB para realização do foco.

Os níveis de fronteira no sistema DaTo são representados apenas pelas letras H e L, e notam respectivamente níveis de fronteiras entoacionais altas e baixas. As fronteiras assinaladas no sistema DaTo são fronteiras finais de enunciados ou fronteiras que precedem pausas no interior dos enunciados. Fronteiras intermediárias como assinaladas no sistema ToBI são detectadas no sistema DaTo automaticamente pelo programa SG Detector (Barbosa, 2006), que faz a segmentação dos enunciados em grupos acentuais pela detecção de acentos frasais, que delimitam os grupos acentuais por terem duração maior na sílaba tônica. Esses acentos frasais são detectados pela duração das unidades V-V, e pelo cálculo do *z-score* dessas durações encontrando assim picos de duração que delimitam os grupos acentuais (Barbosa, 2006).

A tabela 2 mostra uma lista dos contornos dinâmicos e níveis de fronteiras do sistema DaTo, acompanhados de suas descrições.

Contornos Dinâmicos		
LH	rising	contorno que parte de uma posição baixa na sílaba pré-tônica, alcança seu pico alinhado à vogal tônica e tem um resetting final.
>LH	late rising	contorno que parte de uma posição baixa na consoante da sílaba tônica e tem todo o movimento de subida alinhado à vogal tônica, alcançando seu pico após a vogal tônica, acompanhado de um resetting final
vLH	compressed rising	contorno formado entre dois picos em palavras distintas; não há espaço temporal para que ocorra um pico na vogal tônica nessas condições, fazendo com que se realize comprimido entre dois picos.
HLH	falling-rising	apresenta a mesma estrutura que lh, porém antecedido por um pico na mesma palavra; geralmente ocorre em palavras com mais de três sílabas, tendo o primeiro pico alinhado à primeira sílaba da palavra e o seguinte à sílaba tônica.
HL	falling	contorno que parte de uma posição alta na sílaba pré-tônica, alcança seu nível mais baixo alinhado à vogal tônica; ocorre em fronteiras de enunciados assertivos.
>HL	late falling	contorno que parte de uma posição alta na consoante da sílaba tônica e tem todo o movimento de descida alinhado à vogal tônica, alcançando seu nível mais baixo após a vogal tônica; também ocorre em fronteiras de enunciados assertivos.
vHL	compressed falling	contorno que parte de uma posição alta na sílaba pré-tônica, porém não alcança seu nível mais baixo alinhado à vogal tônica, pois não há espaço temporal para alcançar o nível mais baixo, tendo que realizar um movimento de subida logo em seguida.
LHL	rising-falling	contorno que se inicia em posição alta e realiza um downstepping durante a duração das palavras seguintes; ocorre em fronteiras de enunciados assertivos.
Níveis de Fronteira		
L	Low	nível de fronteira baixo.
H	High	nível de fronteira alto.

Tabela 2: conjunto de rótulos do sistema DaTo.

4.3.2. Transcrição Ortográfica

A transcrição ortográfica que ocupa uma das camadas de notação do sistema é tão simplesmente a transcrição de como os enunciados foram realizados pelos participantes, com o intuito de preservar idiosincrasias da fala e do dialeto do falante. A transcrição ortográfica feita no programa Praat é alinhada item por item ao sinal de fala, e oferece a possibilidade de extração do formato *.textgrid* para um arquivo de texto por meio de um *script* específico para essa finalidade. Dessa forma o conteúdo transcrito no sistema é idêntico à transcrição do corpus.

4.3.3. Notação Pragmática

A camada reservada para notação pragmática tem como função apontar fenômenos pragmáticos e comunicativos que possam facilitar a interpretação da ocorrência de fenômenos fonético-entoacionais, como realização de foco, interrupções (como pausas e hesitações) e alongamentos, por exemplo. A notação desses fenômenos não segue um formato específico, ficando a critério de o pesquisador atribuir um rótulo para cada evento destacado.

4.3.4. Unidades V-V

A camada da transcrição que apresenta a segmentação dos enunciados em unidades V-V compreende a divisão de todo enunciado em segmentos compreendidos entre os *onsets* de duas vogais. Essa segmentação é feita semi-automaticamente através do programa *Beat Extractor* (Barbosa, 2006, p.451), que gera no programa Praat uma camada contendo a segmentação dessas unidades.

Este tipo de segmentação no sistema DaTo, assim como proposto em Barbosa (2006) e Barbosa et al (2005), é usada para delimitar uma unidade cujas fronteiras são distintas das da sílaba fonológica e possui a vantagem de dinamizar a transcrição do sistema na medida em que a detecção de transições C-V por meio desse tipo de segmentação auxilia na observação de fenômenos de ordem física que podem estar relacionados à produção e percepção (Marcus, 1981; Dogil and Braun, 1988; Wong and Schreiner, 2003), e na determinação do ritmo (Barbosa, 2006) e da entoação.

De acordo com Barbosa (2006), as unidades V-V apresentam uma estabilidade duracional que oferece um efeito de compensação e uma homogeneidade sobre a duração dos segmentos. Devido a essas propriedades, unidades V-V são usadas como blocos de construção para delimitação de grupos acentuais.

4.3.5. Grupos Acentuais

Os grupos acentuais (doravante SG do inglês *stress group*)¹⁰ são unidades delimitadas por dois acentos frasais consecutivos. Os acentos frasais não se referem a aspectos sintáticos, mas sim a aspectos fonéticos, pois são definidos como proeminências no domínio da produção da fala, assinaladas por picos locais de duração ao longo dos enunciados (Barbosa, 2006, p.10). Por sua vez, os picos locais de duração que assinalam os acentos frasais são determinados pelos padrões de duração das unidades V-V.

No sistema DaTo, os grupos acentuais são determinados automaticamente através do programa *SG Detector* (Barbosa, 2006, p. 459) que, assim como no caso das unidades V-V, gera uma camada com a segmentação dos grupos acentuais no programa Praat.

A delimitação dos SGs através da detecção dos acentos frasais possibilita a localização e marcação automática das fronteiras entoacionais, as mesmas fronteiras que em ToBI são marcadas com os números 1, 2 e 3 baseando-se apenas na percepção (Beckman et al, 1994). A marcação dos SGs se mostrará bastante funcional no sistema DaTo, pois como se verá a seguir ela será responsável pela coordenação entre aspectos prosódicos e discursivos.

¹⁰ Aqui será adotada a terminologia em inglês.

4.3.6. Segmentos Discursivos

A camada reservada para a segmentação discursiva tem como objetivo marcar as fronteiras entre cada enunciado de acordo com a divisão proposta no modelo G&S (Grosz & Sidner, 1986), que utiliza as expressões linguísticas como delimitadores entre cada enunciado que compõe um segmento discursivo (DS). Em outras palavras, é a transposição da segmentação discursiva efetuada no Corpus VoCE para uma camada de notação.

5. Análise e Resultados: Interação Discurso – Ritmo - Entoação

O sistema de notação DaTo, por não se limitar a fazer apenas a notação dos aspectos entoacionais relacionados à curva entoacional e acrescentar ao sistema ferramentas que dão conta de aspectos do ritmo da fala e do discurso, possibilitou a observação da existência de uma interação entre esses fatores de forma dinâmica.

Para justificar essa hipótese foi usado como modelo teórico o MDR (ver capítulo inicial), apresentado em Barbosa (2006), que apesar de se tratar de um modelo do ritmo, oferece um modelo plausível para a interação entre este, o discurso e a entoação, pois a análise dos dados de acordo com o sistema DaTo, assim como no MDR, indica um sistema que tem como mecanismo atrator os grupos acentuais. Além disso, Barbosa (2006) propôs a inclusão de um oscilador glotal em interação com os demais osciladores, posicionamento que reformulamos aqui.

5.1. Interações entre contorno entoacional, ritmo e discurso

Os dados do Corpus VoCE transcritos pelo sistema DaTo mostraram uma coincidência consistente entre fronteiras de grupos acentuais, contornos entoacionais e segmentos discursivos. Tais coincidências entre entoação, ritmo e discurso, ou entre aspectos prosódicos e discursivos, sugere a hipótese de que no MDR, paralelamente ao oscilador silábico, que determina a proeminência baseando-se apenas na duração dos segmentos, exista um oscilador glotal, que determina proeminências e fronteiras entoacionais de acordo com a frequência de vibração de pregas vocais na glote, sinalizados pelo movimento de subida ou descida de f_0 .

Num primeiro momento, foi observada a ocorrência de fronteiras de SG alinhadas proximamente aos contornos dinâmicos LH e >LH, e de nível de fronteira L. Os contornos LH e >LH são mais frequentes na marcação de foco estreito no PB, de acordo com o sistema DaTo, o que indica que uma das estratégias para a marcação de proeminências é o alongamento na duração das vogais, indicado pelo alinhamento com o acento frasal.

Esse alinhamento não acontece de forma exata entre fronteiras de SG e contornos entoacionais. Pode-se dizer que se trata de um alinhamento aproximado, e isso se deve ao fato de que SGs e contornos entoacionais são marcados de forma distinta: os SGs são marcados com base na duração das unidades V-V e se utilizam da marcação dos *onsets* vocálicos para sua delimitação,

enquanto os contornos entoacionais são anotados, por convenção, coincidindo com a posição medial das vogais tônicas. Dessa forma, o que consideramos alinhamento ocorre quando o contorno entoacional está marcado na vogal tônica da palavra em que ocorre o alongamento necessário para a detecção de uma fronteira de SG, como se pode observar na figura 9, que assinala o alinhamento entre os contornos >LH e LH e as fronteiras dos SGs.

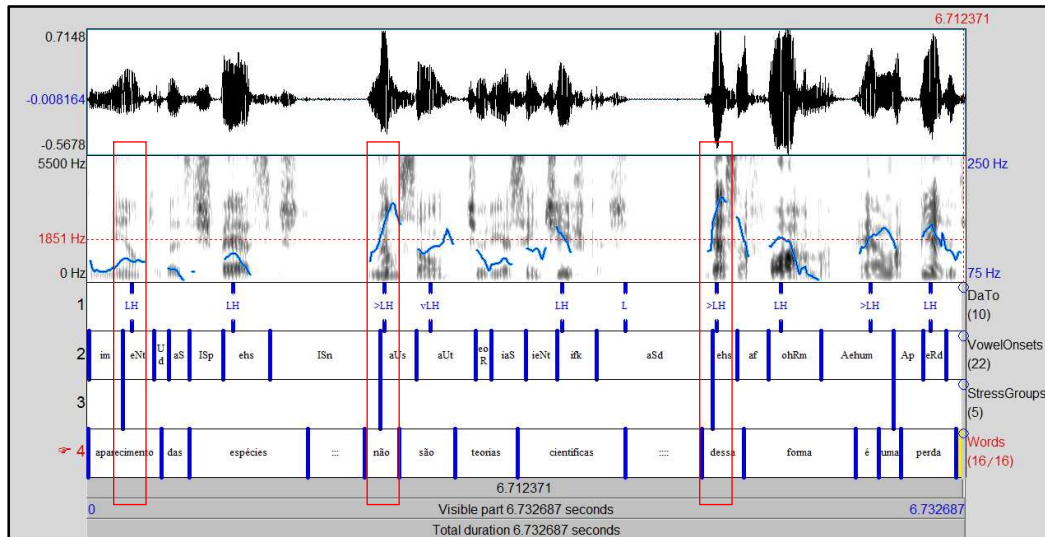


Figura 9: alinhamento entre fronteiras de grupos acentuais e contornos dinâmicos ascendentes

Em um segundo momento, após ter considerado pertinente para o estudo funcional dos aspectos entoacionais a análise da estrutura discursiva e levar a segmentação dos enunciados em unidades discursivas para as camadas de notação do sistema DaTo, foi possível observar que existe um alinhamento aproximado também entre as fronteiras de DSs e SGs. Em geral a diferença no alinhamento que foi desconsiderada é do valor de duração de um segmento vocálico ou consonantal, não chegando a ser do valor de uma unidade V-V, pois, pelo mesmo motivo que determina a diferença no alinhamento entre SGs e contornos dinâmicos, a segmentação das unidades discursivas obedece fronteiras lexicais, enquanto nos SGs a segmentação é feita com base nas unidades V-V. Os casos em que o alinhamento entre SG e contornos dinâmicos se dá a uma distância acima do valor de uma unidade V-V foram considerados desalinhados.

Dos 315 segmentos discursivos analisados a partir do Corpus VoCE, foram encontrados 530 SGs, dos quais 56,7% (178) se alinham às fronteiras dos DS, e 43,3% (136) não se alinham. Os DSs têm duração média de 4,06 s e contêm em média dois SGs. A tabela 3 mostra um resumo de algumas medidas realizadas nesses segmentos e a figura 10 mostra a distribuição do número de SGs por DS e a concentração de dois SGs por DS. Esses dados incluem os SGs que não se alinham exatamente aos DSs, mas que se inserem nestes.

	Média	Mediana	Desvio Padrão
Duração dos DS	4,06 s	3,49	2,49s
SG por DS	2,74	2,0	1,77

Tabela 3: dados sobre a relação entre segmentos discursivos e grupos acentuais.

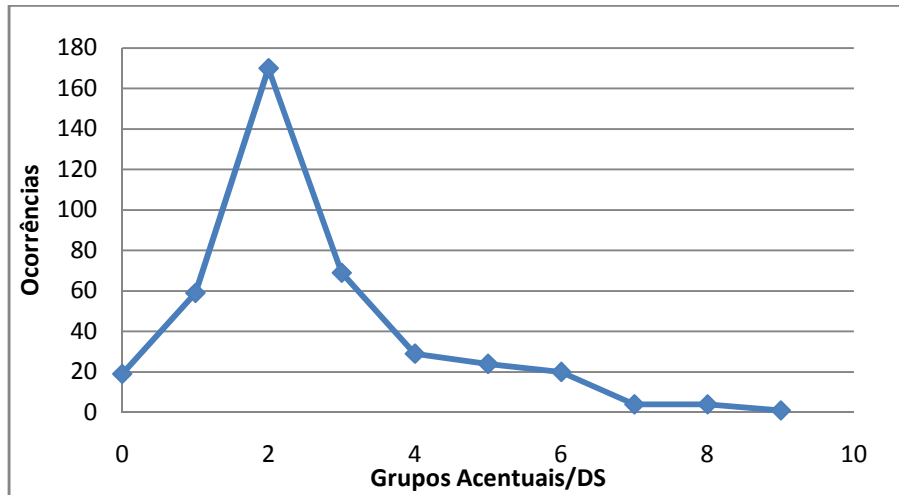


Figura 10: gráfico de distribuição e ocorrência de SG por DS

Além das interações entre ritmo e entoação e ritmo e discurso, os dados mostraram ainda a interação entre os três fenômenos. Como observado inicialmente, as fronteiras dos grupos acentuais coincidem em algum grau com a marcação dos contornos LH e >LH, assim como coincidem com a segmentação dos segmentos discursivos. Ao analisar o alinhamento entre essas três camadas foi possível observar que, no espaço de um segmento discursivo, a curva de f_0 apresenta um padrão de movimento já descrito para enunciados declarativos em PB (Moraes, 1998; Lucente, 2008) que é a subida inicial de f_0 , marcada por LH ou >LH, seguida por um *resetting* de f_0 até a fronteira final, marcada pelo nível L. O que foi observado nos dados de fala espontânea é que esse padrão se repete ao longo do discurso em alinhamento com os DSs, como mostra a figura 11.

que mostra a curva entoacional, SGs e DSs do trecho de discurso *usp* (ver anexo), os DSs assinalados ilustram:

- no DS8 existem marcador discursivo e pressuposição sobre a presença de vários portões a partir de um alongamento e subida de f_0 no final do enunciado (*porque né? a USP tem vários portões tem o portão principal :::*);
- no DS16 ocorre pausa e presença de marcador discursivo final (*e o pessoal fala pô e ó :::: governo deixou explorar isso aqui né? :::*), nesse segmento é interessante observar que ocorre um *reset* de f_0 após a pausa intermediária;
- no DS18 ocorrem pausa, hesitação e ironia sobre uma suposta solução para os atrasos devido ao fato dele não poder atravessar a USP de carro (*eu fiquei pensando né? cara eh :: os caras podiam dar pelo menos um atestado pra eu levar pro empregador*).

O mesmo ocorre na figura 13 com o trecho de discurso *trama* (ver anexo). Os DSs 3, 7 e 13 assinalados contêm, respectivamente:

- hesitação, pausa, enumeração (*porque eu lembro que ah ah na época que eu retornei ao Brasil ::: o que realmente interessava pras grandes gravadoras ou era o sertanejo :::: ou o axé : ou o pagode :: qualquer uma :*);
- marcador discursivo, pressuposição (*então :::: eh a Trama junto com outras gravadoras independentes começaram apostar em outras coisas*);
- pausa, hesitação (*eh mas eu acho que a importância ::: da n...da nossa geração e da Trama e dessa mentalidade do do mercado independente musical :: foi justamente abrir um espaço pra coisas que não tavam rolando*)

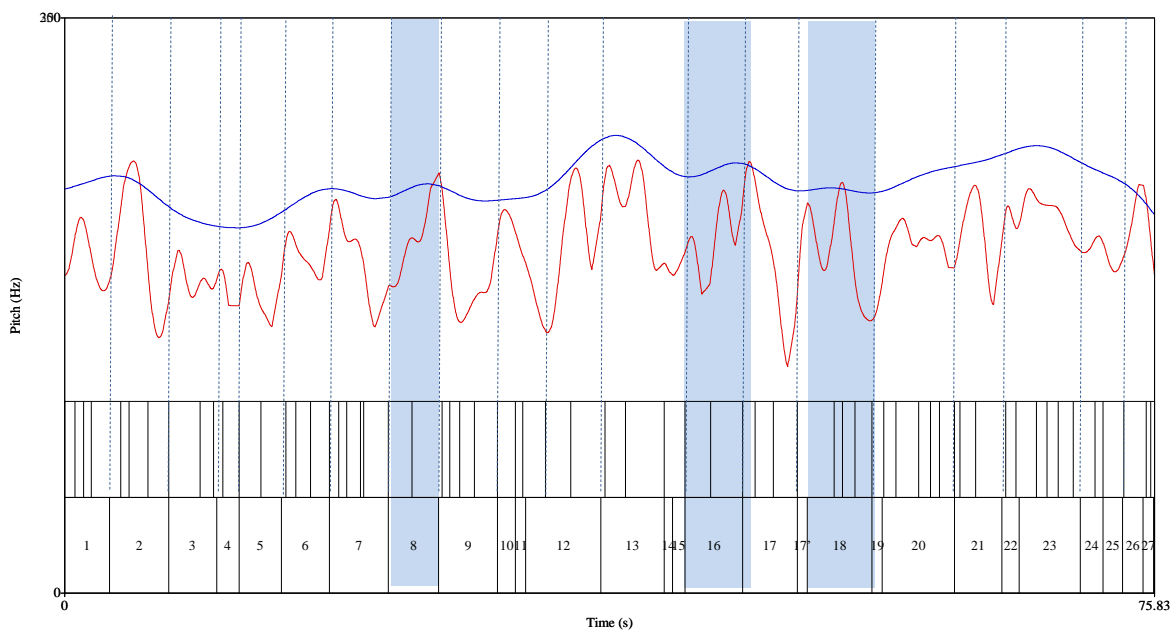


Figura 12: movimentos atípicos no padrão de movimento da curva entoacional (marcados em azul) e alinhamento DSs em trecho do discurso *USP*. A linha vermelha se refere à F_0 e a linha azul à sua suavização.

A correlação cruzada das notações dos DSs mostra, em números absolutos, que a combinação dos contornos dinâmicos LH e >LH em posição inicial e de L em fronteiras de enunciados também é predominante (36%), como mostra a tabela 5. A aplicação do teste chi-quadrado para variáveis independentes ($\chi^2 = 78.22$; $df = 54$; $p = 0.01724$) mostrou resultado significativo para variáveis dependentes.

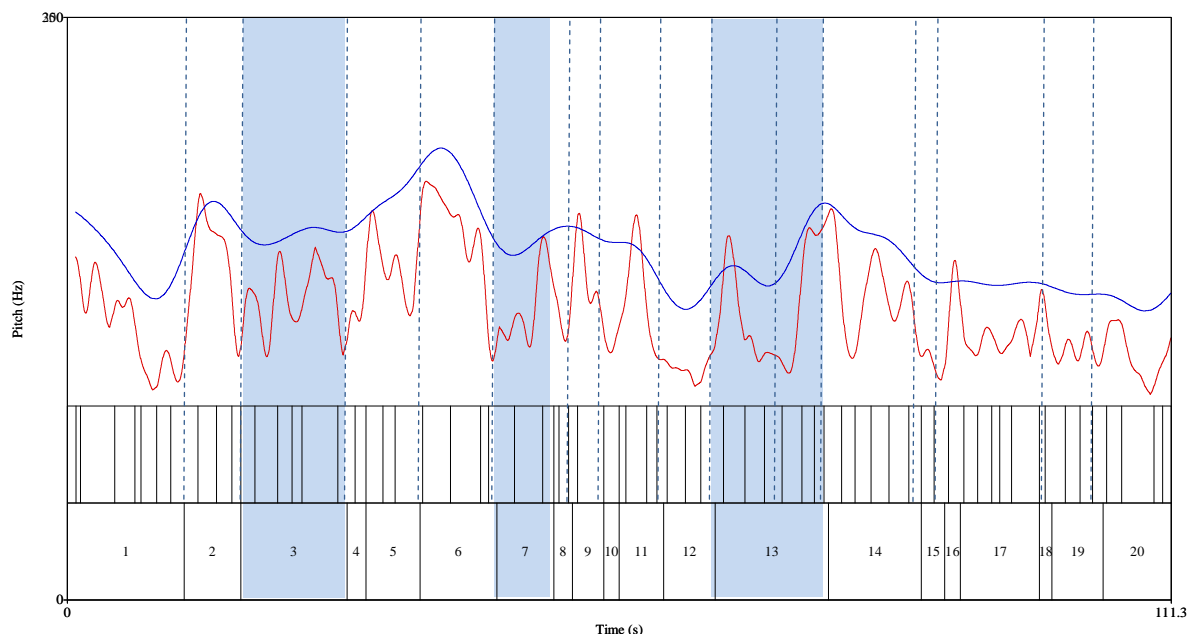


Figura 13: movimentos atípicos no padrão de movimento da curva entoacional (marcados em azul) e alinhamento DSs em trecho do discurso *trama*. A linha vermelha se refere à F0 e a linha azul à sua suavização.

Contorno Inicial	LH	>LH	N	HLH	vHL	vLH	LHL
Contorno Final							
L	80.0	62.0	10.0	4.0	2.0	2.0	1.0
>HL	2.0	3.0	1.0	1.0	1.0	1.0	1.0
>LH	9.0	6.0	6.0	1.0	1.0	1.0	1.0
H	26.0	15.0	8.0	3.0	2.0	1.0	1.0
HL	6.0	4.0	5.0	1.0	1.0	1.0	1.0
HLH	1.0	2.0	3.0	1.0	1.0	1.0	1.0
LH	13.0	8.0	5.0	2.0	1.0	1.0	1.0
LHL	15.0	11.0	9.0	1.0	1.0	1.0	1.0
N	5.0	4.0	11.0	1.0	1.0	1.0	2.0
vLH	2.0	1.0	2.0	1.0	1.0	1.0	1.0

Tabela 5: correlação cruzada das ocorrências de combinação de contornos iniciais e finais para cada um dos 315 enunciados que compõem os DSs.

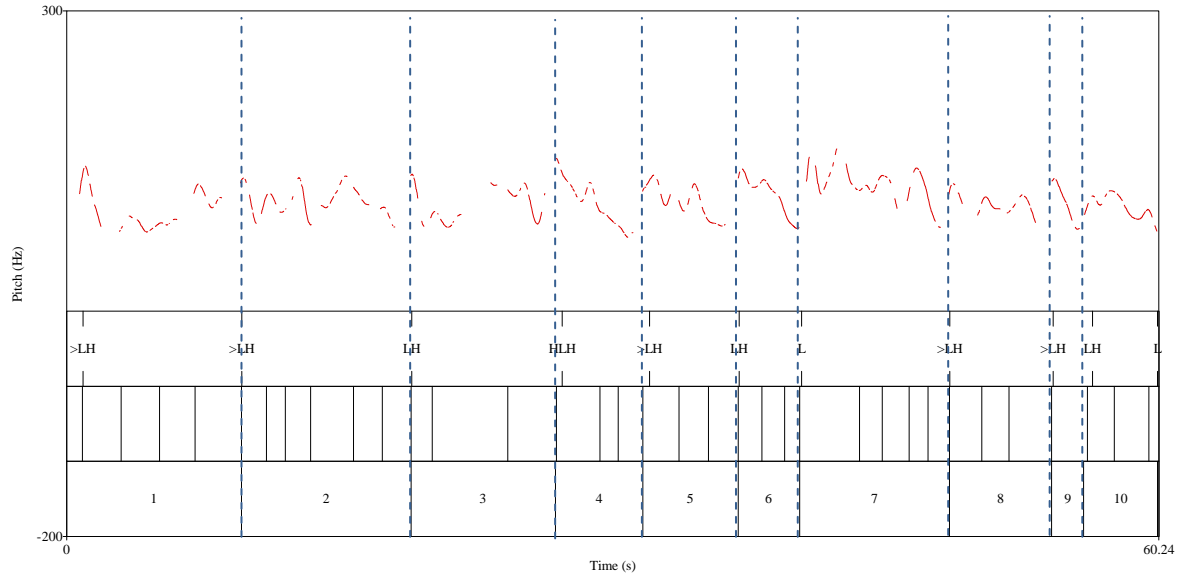


Figura 11: repetição de padrão de movimento ao longo da curva entoacional e alinhamento DSs; na primeira camada estão os contornos dinâmicos marcando o início de DS (exceto os níveis de fronteira L), na segunda os SGs e na terceira os DSs.

Podemos observar que a cada início de DS ocorre um pico de f_0 , que contrasta com a fronteira baixa do segmento anterior, organizando o discurso em segmentos que atendem aos requisitos do modelo G&S, apresentados na seção anterior. Em números, a notação dos contornos iniciais de cada segmento dos 315 DSs é predominantemente marcada por LH e >LH, e a fronteira do enunciado pelo nível L, como se pode ver na tabela 4.

		LH	>LH	vLH	HLH	LHL	>HL	vHL	HL	H	L	n
Posição	Nº	149	106	1	6	1		2				50
Inicial do	absolutos											
DS	%	47,30	33,65	0,31	1,90	0,31		0,63				15,87
Posição	Nº	24	18	2		32	3		11	49	153	18
Final do	absolutos											
DS	%	7,61	5,71	0,63		10,15	0,95		3,49	15,55	48,57	5,71

Tabela 4: ocorrência de contornos dinâmicos na notação das posições iniciais e finais de enunciados que compõem os DSs.

Os DSs que não seguem esse padrão são encontrados em diferentes posições nos trechos de discurso, como se podem ver destacados em azul nas figuras 12 e 13 a seguir. Essa diferença de padrão entoacional pode ser atribuída à presença de enunciados interrogativos, pausas internas aos DSs, hesitações, correções – ou qualquer fenômeno que quebre a sequência discursiva (ver segundo capítulo da parte II) – presença de aspectos pragmáticos (que alteram de alguma forma a configuração *default* de f_0 nos enunciados declarativos), presença de marcadores discursivos (*né?*). Na figura 12,

Esses números sugerem que a repetição do padrão do movimento da curva entoacional, iniciado por um pico e finalizado em posição baixa, alinhada aos segmentos discursivos, que por sua vez se alinham às fronteiras de grupos acentuais em grande parte dos casos, mantém uma relação consistente entre os três níveis de análise. De acordo com o MDR, a observação de como se estrutura essa relação sugere que o oscilador silábico pode estar em acoplamento com um mecanismo que gera padrões glotais, e conseqüentemente sugere também a hipótese de que no PB os falantes organizam a fala, ou o discurso, em ciclos temporais ancorados em padrões rítmico-entoacionais.

Em suma: os dados de segmentação discursiva, segmentação em grupos acentuais e notação entoacional revelam uma estrutura dinâmica em que concentra em grande maioria dois SG a cada DS com duração aproximada de 4s. As fronteiras dos DSs coincidem em 56% com fronteiras de SGs, ou seja, com os acentos frasais. Os acentos frasais, determinados pelo aumento na duração dos segmentos vocálicos também se alinham aos contornos entoacionais LH e >LH em posição inicial dos DSs e aos contornos L e LHL em posição final dos DSs. A correlação cruzada entre os contornos entoacionais mostra forte correlação entre LH – L, entre >LH – L e entre LH – LHL, que indica um segmento iniciado com contorno ascendente e finalizado com um *resetting* de f0, formando um padrão que se repete ao longo do discurso.

Esse alinhamento entre as estruturas discursiva, rítmica e entoacional é possível por uma coordenação de fatores que convergem para a determinação dos SGs como atratores de um sistema produtor da fala e da entoação.

5.2. Nova Hipótese para o Modelo Dinâmico do Ritmo

Da hipótese da organização do discurso em ciclos temporais ancorados em padrões rítmico-entoacionais, e da existência de um oscilador glotal em sincronia com o oscilador silábico no MDR, surge a proposta de acréscimo deste oscilador no MDR, já apresentado por Barbosa (2006).

Em nossa proposta de reformulação, o oscilador glotal se encaixa no MDR paralelamente ao oscilador acentual, interagindo com informações contidas em níveis linguísticos elevados – onde se inserem as informações sintática, semântica e pragmática do discurso – e acoplado ao oscilador silábico por uma força de acoplamento x . O oscilador silábico mantém sua relação com o oscilador acentual e com a pauta gestual da mesma forma, o que é proposto nessa inserção é que o acoplamento entre os osciladores glotal e silábico possa determinar o alinhamento entre informações discursivas contidas nos padrões entoacionais e grupos acentuais.

Nessa nova proposta o oscilador glotal mantém ainda uma relação de acoplamento z com o oscilador acentual. Essa relação determina o alinhamento entre o alvo dos contornos entoacionais e as sílabas tônicas. Ou seja, o oscilador glotal se relaciona com os osciladores silábico e acentual, num sincronismo que une duração, tonicidade e movimento de glote na determinação da ênfase e dos padrões entoacionais do discurso. O diagrama da figura 14 ilustra a nova proposta para o MDR.

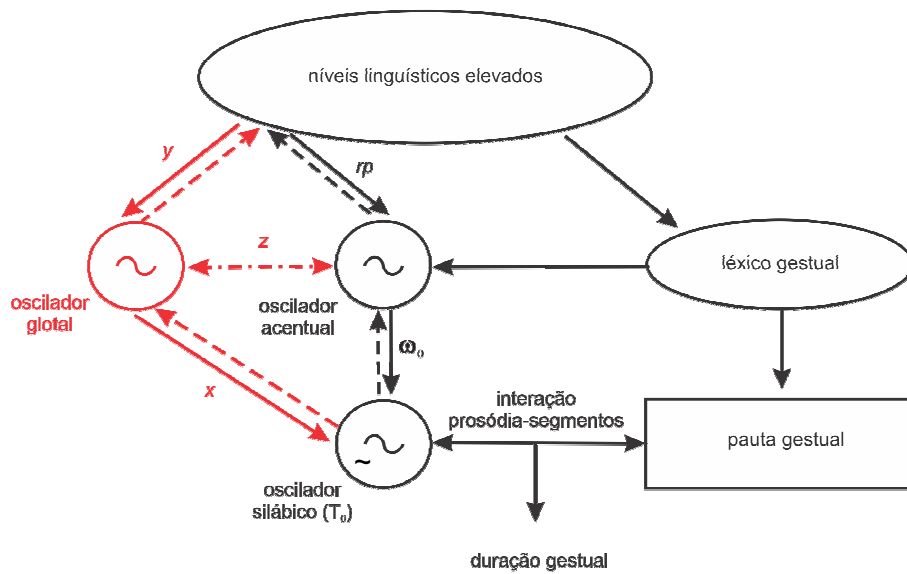


Figura 14: diagrama do Modelo Dinâmico do Ritmo da Fala

O sistema DaTo vem usando a segmentação dos enunciados em unidades V-V em sua notação desde sua implementação, e se valendo dessa segmentação para observações de como nos momentos de transição C-V se alinham a picos de velocidade na subida e/ou descida de f_0 , alinhamento de picos de f_0 com vogais tônicas, transições nos movimentos de f_0 ancoradas nas transições C-V e V-C, etc.

A observação de como se estrutura essa relação entre os padrões de duração das unidades V-V, como resposta à ação do oscilador silábico e a organização entoacional, que aponta para a existência de um oscilador glotal, sugere a hipótese de que, no PB, os falantes organizam a fala em ciclos ou grupos que seguem padrões entoacionais. Esse tipo de observação sobre o comportamento da curva entoacional do PB auxilia na delimitação de um programa de investigação científica dinamicista, pois corrobora a hipótese da ação de diferentes sistemas num único sistema que os contém.

De acordo com o que propõem as teorias que trabalham com sistemas dinâmicos, a entoação pode ser vista como resultante da interação entre diversos elementos em um sistema dinâmico-funcional, como propõe a fonologia articulatória (Browman e Goldstein, 1986; Saltzman, 1995), e o estudo da produção do ritmo da fala (Barbosa, 2006). Essa incursão no campo da fonologia articulatória ao ser adotado o modelo de Barbosa (2006) preenche uma lacuna que o modelo PENTA, proposto por Xu (2005) não preenchia e que se faz necessária no tipo de abordagem que o projeto faz dos aspectos entoacionais, que é a necessidade de representação dos fenômenos entoacionais. O modelo PENTA não pressupõe representação fonológica para os eventos entoacionais, considerando estes estritamente fonéticos. A perspectiva adotada aqui, apesar de tratar a entoação sob um ponto de vista fonético, pressupõe ainda a representação dos contornos entoacionais por meio do sistema DaTo, que é ferramenta central na pesquisa a que se propõe este projeto.

O MDR, apesar de não conter uma representação fonológica nos moldes de um sistema de notação, não torna excludente o papel da fonologia na representação prosódica, muito embora considere a relevância de um alinhamento gradiente entre eventos singulares dos contornos representados com o material linguístico.

6. Discussão

Este trabalho pode ser dividido em duas partes, uma dedicada à análise do discurso e outra dedicada à análise da entoação. Essas partes se complementam na medida em que a estrutura de análise discursiva e segmentação das unidades que compõem o discurso passam a serem analisadas por meio de uma ferramenta comum à análise entoacional: o sistema DaTo.

O fato de uma estrutura discursiva poder ser analisada de tal forma se deve à abordagem dinâmica adotada para o estudo e análise da entoação. A partir do momento em que se adota um programa de pesquisa que abandona a dicotomia entre mente e corpo é possível unir em um único sistema cognitivo-motor a elaboração do discurso e a produção da fala.

O sistema DaTo, em sua notação, agrega esse fator dinâmico na medida em que possibilita a observação da emergência de padrões entoacionais relacionados aos segmentos discursivos. Paralelamente a essa relação do discurso com a entoação, o ritmo age como um atrator e coordena as três estruturas: discursiva, rítmica e entoacional.

Se o sistema DaTo oferece condições para a observação da coordenação dessas estruturas e a notação dos elementos que fazem transparecer a economia dessa coordenação, o Modelo Dinâmico do Ritmo, proposto por Barbosa (2006), com a proposta de inclusão de um oscilador glotal, assume a tentativa de explicar o funcionamento dessa organização em um sistema dinâmico.

A hipótese da existência de um oscilador glotal, que funciona recebendo informações pragmáticas e discursivas de um nível superior – sistema cognitivo – e sua relação com os osciladores acentual e silábico – sistema motor –, pode explicar o fato de a informação contida no discurso ocorrer em um segmento específico, com acentuação e duração apropriadas para que tenha seu significado compreendido no discurso.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abner, N. 2009. Phrasing and Prominence: Extracting Prosodic Information from the BU Radio Corpus. Ms. UCLA.
- Atterer, M. and Ladd, D. R. (2004) On the phonetics and phonology of "segmental anchoring" of f_0 : evidence from German. *Journal of Phonetics* 32, 177-197.
- Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes.
- _____ (2007) From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Communication*. 49 (1-2), 725 - 742.
- _____ (2008) Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. *Proceedings of the Speech Prosody 2008 Conference*. Campinas.
- _____ (2010) Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech In *Proceedings of the Speech Prosody 2010*. Chicago.
- Barbosa, P. A., Arantes, P., Meireles, A. R., Vieira, J. M. (2005). Abstractness in Speech-Metronome Synchronisation: P-Centres as Cyclic Attractors. *Proceedings of the Ninth European Conference on Speech Communication and Technology (Interspeech 2005)* Lisbon, Portugal (1441-1444).
- Beckman, M. E. (1996). A typology of spontaneous speech. In Y. Sagisaka, W. N. Campbell & N. Higuchi, eds., *Computing Prosody*, pp. 7-26 . New York: Springer-Verlag.
- Beckman, M. E., Hirschberg, J., Pitrelli, John F., (1994). *Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework*. (Disponível em http://www.ling.ohio-state.edu/~tobi/ame_tobi).
- Biber, D., Conrad, S. (2009) *Register, genre and style*. Cambridge University Press.
- Boersma, P., Weenink, D. (2009): Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from <http://www.praat.org/>
- Botinis, A., Granström, B., Möbius, G., (2001) Developments and paradigms in intonation research. (4): 263-296.
- Brown, G. (1983) "Prosodic structure and the given_new distinction". In Ladd, D. R. & Cutler, A. (eds.) *Prosody: Models and Measurements*, Springer Verlag, Berlin, p.67-78.
- Browman, C.P. & Goldstein, L. (1986) Towards an articulatory phonology. In C. Ewan and J. Anderson (eds.) *Phonology Yearbook 3*. Cambridge: Cambridge University Press, p. 219-252.
- Butterworth, B. (1975) Hesitation and semantic planning in speech. In *Journal of Psycholinguistic Research* (4); p.143-178.

- Campbell, W.N. (1996) Synthesizing Spontaneous Speech. In Sagisaka, Y., Campbell, N., Higuchi, N. (eds) *Computing Prosody. Computational Models for Processing Spontaneous Speech*. New York: Springer-Verlag.
- Chafe, W. (1976) "Givenness contrastiveness de_niteness subjects topics and point of view". In Li, C., (ed.) *Subject and Topic*, Academic Press New York, p. 25-55.
- Chafe, Wallace L. (1979) The Flow of Thought and the Flow of Language. In Givon, T., Ed., *Syntax and Semantics, Vol. 12, Discourse and Syntax*. Academic Press, New York, New York: 159-182.
- Chafe, W.L. (1980) The Deployment of Consciousness in the Production of a Narrative. In Chafe, W.L., Ed., *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production. Vol. 3. Advances in Discourse Processes*. Ablex Publishing Corp, Norwood, New Jersey: 9-50.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Cohen, P. R., and Levesque, H. J. (1980). Speech Acts and the Recognition of Shared Plans. In *Proceedings of the 3rd Conference of the Canadian Society for Computational Studies of Intelligence*, Victoria, B.C., p. 263-271.
- Côrtes, P. O., Mittmann, M. M., Caetano, R. V. O., Mello, H. R.; Raso, T. (2011) A convergência entre anotadores na segmentação prosódica do corpus C-ORAL-BRASIL. In: *Anais do III Colóquio Brasileiro de Prosódia da Fala*. Belo Horizonte, p. 1-7.
- Dogil, G., Braun, G. (1988) The PIVOT model of speech parsing.. Vienna, Áustria: Verlag.
- Fernandes, F. (2007) Ordem, focalização e preenchimento em português: sintaxe e prosódia. . Ph.D. Thesis. Unicamp, Campinas.
- Fujimura, O. (2000) The C/D model and prosodic control of articulatory behavior. *Phonetica* 57, p.128-138.
- Gravano, A. and Hirschberg, J. (2006) "Effect of Genre, Speaker, and Word Class on the Realization of Given and New Information," *Proceedings of Interspeech 2006*, Pittsburgh.
- Grice, M., Ladd, D. R., Arvaniti, A. (2000) On the place of phrase accents in intonational phonology. *Phonology* 17: p. 143-185.
- Grosz, B.J., and Sidner, C.L., (1986) "Attention, Intentions, and the Structure of Discourse", *Computational Linguistics*, p. 12:3.
- Hirschberg, J., and Litman, D. (1993) Empirical Studies on the Disambiguation of Cue Phrases, *Computation Linguistics*, 19-3, p. 501–530.
- Hasegawa-Johnson, Mark, Chen, Ken, Cole, Jennifer, Borys, Sarah, Kim, Sung-Suk, Cohen, Aaron, Zhang, Tong, Choi, Jeung-Yoon, Kim, Heejin, Yoon, Tae-Jin, & Chavarría, Sandra. (2005)

- Simultaneous Recognition of Words and Prosody in the Boston University Radio Speech Corpus. *Speech Communication* 46, p. 418-439.
- Jubran, C. C. A. S. & Koch, I. G. V. (Org.). (2006). Gramática do português culto falado no Brasil. Volume 1: Construção do texto falado. Campinas: Editora da Unicamp.
- Kelso, S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative* 15.
- _____ (1995). *Dynamic Patterns*. Cambridge: MIT Press.
- Kohler, K. J. (1996) Modelling Prosody in Spontaneous Speech. In Sagisaka, Y., Campbell, N., Higuchi, N. (eds) *Computing Prosody. Computational Models for Processing Spontaneous Speech*. New York: Springer-Verlag.
- Kohler, K. J. (2005) Timing and Communicative Functions of Pitch Contours. *Phonetica* 62, p. 88-105.
- Kugler, P. N., Turvey, M. T. (1987). Information, natural law, and the self-assembly of rhythmic movement. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Ladd, D. R. (1996) *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lakatos, I. (1978) *Falsificação e Metodologia dos Programas de Investigação Científica*. Lisboa: Edições 70.
- Lehiste, I., (1970) *Suprasegmentals*. Cambridge: MIT Press.
- Lieberman, M. (1975) *The intonational system of English*. Ph.D. Thesis; MIT.
- Lucente, L., Hirschberg, J., Barbosa, P. A. (2012) “The role of discourse structures and intonational features on determining the information status prominence” (*to appear*).
- Lucente, L. (2008) *DaTo: Um sistema de notação entoacional do português brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea*. Dissertação de Mestrado. Unicamp.
- Lucente, L., Barbosa, P. A. (2010) The role of alignment and height in the perception of LH contours. *Proceedings of Fifth Conference on Speech Prosody*. Chicago.
- _____ (2008) Narrow focus in Brazilian Portuguese: spatial and temporal constraints. *Proceedings of Fourth Conference on Speech Prosody*. Campinas.
- _____ (2007) Notação Entoacional do Português Brasileiro em Corpora de fala Semi-Espontânea e Espontânea. *Revista Intercâmbio* 16.
- Lucente, L. (2007) Abelhas, pessoas, zebras e caracóis: auto-organização na fala. *Não publicado*.
- Marcus, S.M. (1981) Acoustic determinants of Perceptual-center (p-center) location. *Perception and Psychophysics* 30 (3), 247-256.
- Moraes, J. A. (1998) “Intonation in Brazilian Portuguese”. In Hirst, D., Di Cristo, A. (eds.) *Intonational Systems: a Survey of Twenty Languages*. Cambridge. MIT Press.
- Nakatani, C. H. (1996). Integrating prosodic and discourse modelling. In Y. Sagisaka & al. (Eds). *Computing Prosody*, New-York: Springer: p. 67-80.

- Nakatani, C. H., Hirschberg, J. (1994) A Corpus-based study of repair cues in spontaneous speech, *Journal of the Acoustical Society of America*, 95-3: p. 1603–1616.
- Ostendorf, M., Price, P., and Shattuck-Hufnagel, S. (1996) “Boston University Radio Speech Corpus”, Linguistic Data Consortium, Philadelphia.
- Pierrehumbert, J. (1980) The Phonology and Phonetics of English Intonation. Ph.D thesis, MIT.
- Polanyi, L. and Scha, R.J.H. (1986) “Discourse Syntax and Semantics”. In Polanyi, L., Ed., *The Structure of Discourse*. Ablex Publishing Co., Norwood, New Jersey.
- Port, R. & van Gelder. T. (1995). It’s About Time: An Overview of the Dynamical Approach to Cognition. In R. Port, & T. van Gelder, (Eds.). *Mind as motion: Dynamics, behavior, and cognition*. Cambridge, MA: MIT Press.
- Raso, T., Mello, H., (2012) C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal. Belo Horizonte, UFMG-CNPq.
- Rietveld, A. C. M., Gussenhoven, C. (1988) “On the relation between pitch excursion size and prominence”. *Journal of Phonetics*, Vol 13(3), 299-308.
- Rosenberg, A., Hirschberg, J. (2009) Detecting Pitch Accents at the Word, Syllable, and Vowel Level, NAACL/HLT, Boulder, CO.
- Saussure, F. et al. (2002) *Curso de linguística geral*. 24. ed. São Paulo, SP: Cultrix.
- Sagisaka, Y., Campbell, N., Higuchi, N. (1996) COMPUTING prosody: computational models for processing spontaneous speech. Coautoria de. New York, N.Y.: Springer, 1996.
- Saltzman, E. L. (1995). Dynamics and coordinate systems in skilled sensorimotor activity. In Port, R. & van Gelder, T., (Eds.). *Mind as motion: Dynamics, behavior, and cognition*. Cambridge, MA: MIT Press.
- Scherer, K. R. (1984) On the nature and function of emotion: a component process approach. In Scherer, K. R. Ekman, P. (Eds.) *Approaches to emotion*. Hillsdale, NJ: Lawrence Erlbaum. 1984. p. 293-318.
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, J. Pierrehumbert, J. Hirschberg, and P. Price (1992). TOBI: A Standard Scheme for Labeling Prosody. *Proceedings of the International Conference on Spoken Language*, Banff.
- Sonntag, G., P. & Portele, T. (1998) Comparative evaluation of synthetic prosody with the PURR method. *International Conference on Spoken Language Processing*.
- Tenani, L. E. (1996) Análise prosódica das inserções parentéticas no corpus do projeto da gramática do português falado. Dissertação de Mestrado. Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas.

- Thelen, E. and Smith, L.B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*, MIT Press.
- Wong, S.W., Schreiner, C.E., 2003. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Communication*. 41, p. 93-106.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27, p.55-105.
- _____(2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46, p. 220-251.
- _____(2006). Speech prosody as articulated communicative functions. In *Proceedings of Speech Prosody 2006*, Dresden, Germany.
- _____(2010) In defense of lab speech. *Journal of Phonetics* 38: p. 329-336.

LUCENTE, L.

ANEXO

tramaTR

- 0 - 1 - (1)CARA NA VERDADE ASSIM a a Trama ela ela ::: deu muito certo porque além d/ de ser uma gravador' independente ::: eh co' a mentalidade gravador' independente
(2)eh eh el' apostou em pessoas que o merca/ o grande mercado não tava querend' apostar
2 - (3)porque eu lembro que ah ah na época que eu retornei ao Brasil ::: o que realmente interessava pras grandes gravadoras ou era o sertanejo ::: ou o axé : ou o pagode :: qualquer uma :
(4)[qualquer onda né?] é
(5)qualq/ qualquer outra coisa fora desses três era muito difícil a gravadora apostar :::
3 - (6)ENTÃO o mercado independente começou a servir como u-uma uma válvula de escape pra quem quaria ouvir outras coisas e pra quem também tava fazendo outras coisas
4 - (7)ENTÃO ::: eh a Trama junto com outras gravadoras independentes começaram apostar em outras coisas
(8)não era só nessa coisa do filho de artista
5 - (9)porque ficou muito marcado na Trama isso :::
6 - (10)eu saí da Trama em dois mil e cinco
(11)mas ali no começo ficou muito marcado porque tinha eu tinha o Simoninha o João Marcelo
7 - (12)era e e continua sendo um dos donos ::: um dos donos da Trama e tal :::
8 - (13)eh mas eu acho que a importância ::: da n...da nossa geração e da Trama e dessa mentalidade do do mercado independente musical :: foi justamente abrir um espaço pra coisas que não tavam rolando
9 - (14)ai depois que esse mercad'independente abriu esse espaço ::: a e as gravadoras co/ eh começaram apostar em outras coisas também
(15)ENTÃO isso foi muito legal [elas faliram todas]
10 - (16)um/ a situação não tá fácil :::
(17)eu por exemplo quando saí da trama em dois mil e cinco ai você começa perceber como esse mercado realmente eh eh eh difícil
11 - (18)por conta da pirataria
(19)por conta ::: d'um atraso também das dos próprios executivos das gravadoras
(20)que demoram muito pra en/ entender certos assuntos como a internet ::: eh como a venda online essas coisas

usp

- 0 - 1 - (1)eh eh eu f/ eu fico eu passo ali perto da USP quase todo dia né? :
2 - (2)eh ::: os caras ali aconteceu uma coisa interessante ai do estado brasileiro né? :
- (3)pra/ eh p/ por conta do do Serra também :: a gente tem um governador novo né?
3 - (4)ninguém sabe quem ele é :::
(5)mas eh BASICAMENTE o que ele tem feito hoje em dia é não fazer nada :::
4 - (6)e a gent' também tem um reitor na USP que é mais ou menos o mesmo mesmo pacote ::
5 - (7)ai aconteceu uma coisa interessante essa semana que foi a privatização do portão da USP né? :::
(8)porque né? a USP tem vários portões tem o portão principal :::
6 - (9)o estado como não tá muito interessado com a USP e o reitor parece também não tá nem mais na USP ::
(10)eles pegaram tem um portão né?
(11)tem os grevistas
7 - (12)tem uns oitenta grevistas que representam a class' inteira ::: e os caras ficam abrindo e fechando portão né? :::
8 - (13)e ai TIPO ASSIM cê vai entrar na USP cê tem que meio que negociar c'os caras né? :::
(14)porque né? ::
(15)foi privatizado né?
9 - (16)e o pessoal fala pô e ó ::: governo deixou explorar isso aqui né? :::
10 - (17)é mais ou menos como se dessem as ruas pro pessoal pros flanelinhas explorarem assim né? cê vai lá ::
11 - (18)eu fiquei pensando né? cara eh :: os caras podiam dar pelo menos um atestado pra eu levar pro empregador
(19)ASSIM fala ó
(20)eu vim pe/ eu vim pela USP daí o pessoal não me deixou passar :: eles me deram esse atestado falando que eu vou levar duas horas a mais ::
(21)vou ter que dar volta inteira na USP pra chegar lá cara :::
12 - (22)que é que é uma bizarrice né?
13 - (22)tem oitenta pessoas que fecham o portão ninguém faz nada ninguém entra na USP faz ::
(24)isso t' acontecendo faz três semanas
(25)o estado não faz nada ::
(26)o reitor não tá na USP:::
(27)e tá tudo certo