

A convergência entre anotadores na segmentação prosódica do corpus C-ORAL-BRASIL

Priscila Côrtes

Universidade Federal de Minas Gerais
Laboratório de Estudos Empíricos e Experimentais da
Linguagem
Belo Horizonte, Brasil
pritico@gmail.com

Tommaso Raso

Universidade Federal de Minas Gerais
Laboratório de Estudos Empíricos e Experimentais da
Linguagem
Belo Horizonte, Brasil
tommaso.raso@gmail.com

Raíssa Caetano

Universidade Federal de Minas Gerais
Laboratório de Estudos Empíricos e Experimentais da
Linguagem
Belo Horizonte, Brasil
raissavoliveira@gmail.com

Maryualê Mittmann

Universidade Federal de Minas Gerais
Laboratório de Estudos Empíricos e Experimentais da
Linguagem
Belo Horizonte, Brasil
maryuale@gmail.com

Heliana Mello

Universidade Federal de Minas Gerais
Laboratório de Estudos Empíricos e Experimentais da
Linguagem
Belo Horizonte, Brasil
heliana.mello@gmail.com

Resumo—Nesse estudo, investigamos o acordo entre anotadores referente à anotação prosódica do corpus C-ORAL-BRASIL, uma amostra do português brasileiro de fala espontânea. A validação estatística do acordo aponta para um alto grau de concordância com relação à anotação de quebras prosódicas e a análise qualitativa dos fatores que possam ter gerado o desacordo revela fatores relacionados à prosódia do português brasileiro.

Palavras-chave-Fala espontânea, segmentação prosódica, validação estatística, fatores de desacordo

I. INTRODUÇÃO

Neste trabalho investigamos a segmentação da fala espontânea em entidades pragmáticas (enunciados e unidades tonais) de acordo com a percepção de falantes e tendo em vista a Teoria da Língua em Ato [2], que considera a prosódia a interface para a interpretação da fala. Transcrições de fala espontânea que compõem o corpus C-ORAL-BRASIL [6] foram segmentadas em enunciados e unidades tonais por diferentes membros do projeto, divididos em dois grupos de acordo com seu grau de expertise, e os resultados foram submetidos a testes Kappa [4]. Nosso objetivo é demonstrar que a segmentação da fala nessas unidades é estatisticamente consistente e, principalmente, apontar fatores que possam ter gerado o desacordo entre anotadores, em especial os fatores que dizem respeito à prosódia do português brasileiro (PB).

II. FUNDAMENTOS TEÓRICOS

A Teoria da Língua em Ato baseia-se na Teoria dos Atos de Fala, proposta por [1], e na identificação perceptual de fronteiras prosódicas proposta por [9]. A Teoria dos Atos de Fala defende que um ato de fala é composto por três partes: a locução, a ilocução e a perlocução. A locução é composta pelo conteúdo segmental da fala, enquanto a ilocução corresponde ao tipo de ato que se realiza com o material locutivo. A perlocução diz respeito ao efeito causado pela ilocução no interlocutor. Segundo [9], quebras prosódicas são variações prosódicas perceptíveis no contínuo da fala, suficientes para a segmentação desta em unidades discretas menores.

Partindo desse arcabouço teórico, a Teoria da Língua em Ato postula que a fala espontânea estrutura-se em enunciados, a menor unidade lingüística que pode ser pragmaticamente interpretada, porque veicula uma ilocução. Na entidade de enunciado, os domínios lingüístico e acional estão unidos, mediados pela prosódia, que permite reconhecer o tipo de ilocução veiculado. O enunciado é delimitado por quebras de valor perceptível conclusivo (ou terminal). No exemplo seguinte, a cadeia de fala foi realizada em dois enunciados, cada qual autônomo e separado do outro por um perfil terminal.

- *BAL: cê tá com um jarro d' água // que tem uma espessura assim //¹

O enunciado pode ser estruturado internamente por unidades tonais menores, na maioria das vezes, de valor informacional. Cada tipo de unidade prosódica tem parâmetros prosódicos e distribuição específicos (movimentos de F0, intensidade e duração) e os limites entre elas são marcados por quebras de valor não-terminal. Para a realização de um enunciado é imprescindível apenas a unidade informacional de comentário (COM), veículo da força ilocucionária, caso em que é chamado “simples”, como nos dois exemplos a seguir:

- *BEL: sabe que que aconteceu quando eu fui de intercâmbio //
- *BAL: uhn //

Nas duas ocorrências acima retiradas do corpus, há dois enunciados compostos apenas por uma unidade informacional de comentário. Nota-se que não há relação entre a entidade *proposição*, da escrita, e um enunciado, que seria equivalente a esta na fala, em termos de estrutura interna. O segundo enunciado consiste integralmente numa interjeição, dotada de valor ilocucionário (valor na comunicação). Muitas vezes, no entanto, o enunciado é formado por mais de uma unidade tonal, como a seguir:

- *BAL: as recarregáveis / tão aqui //

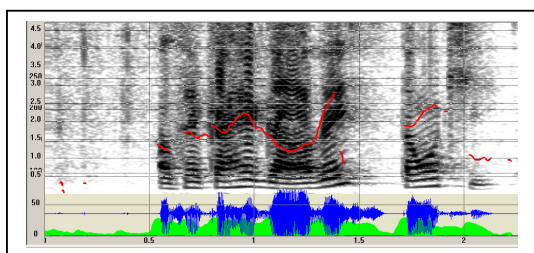


Figura 1. Espectrograma de um enunciado complexo de perfil TOP / COM.

O mesmo conteúdo locutivo “as recarregáveis tão aqui” poderia ser realizado em apenas uma unidade tonal:

- *BAL: as recarregáveis tão aqui //

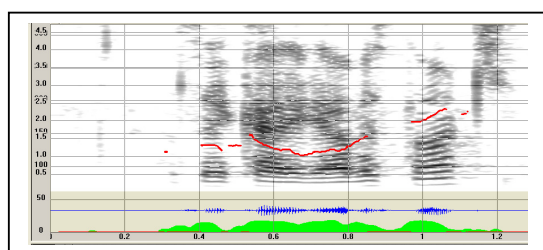


Figura 2. Espectrograma de um enunciado simples de perfil COM //.

¹ O símbolo // marca um perfil terminal, enquanto o símbolo / corresponde a um perfil não-terminal.

A diferença entre as duas possibilidades reside no fato de que há uma quebra não-conclusiva entre as duas unidades que formam o enunciado complexo, organizado em uma unidade de tópico (TOP), campo de aplicação da força ilocucionária, e em uma unidade de comentário. A Teoria da Língua em Ato prevê que falantes de uma língua são capazes de reconhecer essas unidades e as unidades maiores em que elas se organizam, os enunciados, baseando-se apenas na percepção das quebras que as limitam.

III. METODOLOGIA

Os dados utilizados neste trabalho pertencem a uma amostra do corpus C-ORAL-BRASIL, ramo brasileiro do corpus C-ORAL-ROM [3], representativo de quatro línguas românicas: espanhol, francês, italiano e português europeu. Para a composição do corpus, as gravações de variadas situações de fala espontânea, que permitem ampla amostragem de ilocuições, foram transcritas e segmentadas prosodicamente em enunciados e em unidades tonais e mais tarde tiveram o texto alinhado ao som com auxílio do software WinPitch [5]. A amostra analisada neste trabalho é constituída por textos utilizados no processo de treinamento de transcritores e anotadores, no segundo semestre de 2008, e por dois textos pertencentes ao minicorpus, descrito mais adiante.

Os alunos envolvidos no projeto C-ORAL-BRASIL passaram por um período de treinamento para a segmentação, que consistiu na segmentação de textos previamente transcritos por dois grupos de alunos de diferente formação e engajamento com o projeto. O grupo 1 foi formado por alunos de pós-graduação e em maior envolvimento com o projeto; o grupo 2, por alunos de graduação. O grupo 1 segmentou 6 textos de 4280 palavras, e o grupo 2, 7 textos de 7028 palavras. A tarefa consistia na segmentação dos textos, realizada individualmente por cada membro dos grupos, baseando-se apenas na percepção de quebras. As anotações individuais foram comparadas por meio do teste Kappa [4], que apontou os graus de acordo entre os anotadores. Em um segundo momento, um ano após o período de treinamento, o grupo 1 foi submetido a um novo teste, o follow-up, em que segmentou um monólogo (776 palavras) e um diálogo (573 palavras). O follow-up foi realizado somente com o grupo 1 para que se obtivesse uma noção do grau de acordo do minicorpus, amostra composta de cerca de um 1/5 mais representativo e de melhor qualidade acústica dos textos da metade informal do corpus (cerca de 20 textos e 30000 palavras), segmentado apenas por membros do grupo 1. Após a validação estatística, os casos de desacordo foram minuciosamente analisados à procura de fatores que pudessem explicá-los.

IV. ANÁLISE QUANTITATIVA

A validação estatística da segmentação do corpus [6] mostrou que o acordo relativo à anotação de quebras terminais é forte (Grupo 1=0,84; Grupo 2=0,80) e que casos de desacordo total, em que um anotador assinala uma quebra terminal e o outro, nenhuma quebra, são raros e devidos à falta de atenção. O acordo, no que diz respeito à anotação de quebras não-terminais, é mais baixo (Grupo1=0,66 e

Grupo2=0,63) e associado a aspectos específicos do PB. Na tabela a seguir estão os resultados obtidos a partir da validação estatística do acordo referente ao período de treinamento dos dois grupos:

Tabela I. Acordo relativo ao treinamento dos grupos 1 e 2

Treinamento Grupo 1			
Tipo de texto	Geral	Terminal	Não-terminal
Diálogos	0.81	0.87	0.61
Monólogos	0.79	0.78	0.69
Todos os textos	0.79	0.84	0.66
Treinamento Grupo 2			
Tipo de texto	Geral	Terminal	Não-terminal
Diálogos	0.76	0.81	0.60
Monólogos	0.76	0.79	0.65
Todos os textos	0.76	0.80	0.63

Na tabela observa-se um grau de acordo suficiente (a partir de 0.80) no que tange a percepção das quebras terminais tanto entre os membros do grupo 1 quanto do grupo 2.

Os testes Kappa relativos ao follow-up (Grupo 1) revelaram um aumento no acordo entre os anotadores, principalmente com relação à marcação de quebras não-terminais ($kappa$ terminal = 0.86 e $kappa$ não-terminal = 0.78). Nessa nova fase, o acordo geral entre os anotadores chegou a 0.86 [8], como demonstrado na tabela abaixo:

Tabela II. Acordo relativo ao treinamento e ao follow-up do grupo 1

Treinamento			
Tipo de texto	Geral	Terminal	Não-terminal
Diálogos	0.81	0.87	0.61
Monólogos	0.79	0.78	0.69
Todos os textos	0.79	0.84	0.66
Follow-up			
Tipo de texto	Geral	Terminal	Não-terminal
Diálogo	0.85	0.86	0.78
Monólogo	0.86	0.87	0.78
Todos os textos	0.86	0.86	0.78

V. ANÁLISE QUALITATIVA

Após a validação estatística do acordo entre anotadores, os textos envolvidos no processo de treinamento dos transcritores foram analisados em busca de fatores que pudessem ter gerado os casos de desacordo. A análise detalhada desses casos corrobora a hipótese estabelecida em [6] que associa a discrepância na segmentação a fatores específicos ao PB. O desacordo total é normalmente explicado pela existência de elementos nos textos transcritos que camuflam a percepção entoacional. O desacordo relativo à anotação de quebras não-terminais relaciona-se, por sua vez, a especificidades do PB. A tabela seguinte associa o tipo de discordância aos fatores que possam tê-la gerado:

Tabela III. Tipo de anotação prosódica x fatores que geram desacordo

Anotação	Fatores
----------	---------

Terminal x nada	Fatores relacionados à transcrição, sobreposição
Terminal x não-terminal	Fatores cognitivos, quebra sintática, sobreposição, insuficiência teórica
Não-terminal x nada	Fatores cognitivos, quebra sintática, sobreposição, fatores específicos ao PB

A seguir detalhamos os fatores relacionados à transcrição e aos seus critérios.

A. Problemas na transcrição

Os problemas nas transcrições são divididos entre termos transcritos incorretamente, segmentos incompreendidos (sinalizados por xxx, no caso de uma só palavra, e por yyyy, no caso de trechos maiores) e o caso especial dos fenômenos paralinguísticos, representados por *hhh* - não se sabia ao certo, ainda nesse período, se tosse, risos etc. deveriam ser segmentados -, exemplificados a seguir, em que a. e b. representam anotadores diferentes:

- Fragmentos transcritos incorretamente

- *EUG: essas Picadilly aqui é legal também // <rasteirinha> //²
*JAN: essa aqui que <cor cê tem> //
- *EUG: essas Picadilly aqui é legal também // <maciinha> //

A anotação a. foi feita por dois transcritores e a b. é a correta. A discrepância consiste na anotação de uma quebra terminal depois da palavra “rasteirinha”, que está no lugar de “maciinha”, e na não-anotação de quebra no mesmo contexto por parte de outro transcritor. Acreditamos que o fato de parte do conteúdo locutivo estar transcrito incorretamente possa ter influenciado a anotação prosódica.

Os fragmentos transcritos incorretamente geram também, por vezes, discrepâncias do tipo *não-terminal x nada*, como demonstrado a seguir:

Contexto: um vendedor descreve um tipo de sandália para a cliente.

- *EUG: uma sandália mais simples / uma sandália de borracha // mas assim / muito bonitinha //
- *EUG: uma sandália mais simples / uma sandália de borracha // mas assim muito bonitinha //

Nesse caso, a presença do “assim” pode ter levado os anotadores a tomar decisões diferentes, de modo que a.

² Os colchetes angulares indicam sobreposição de fala.

assinalou uma quebra não-terminal entre “assim” e “muito”, enquanto b. não assinalou quebra alguma.

- Segmentos incompreendidos:

Contexto: um colega explica a outro como usar o equipamento de gravação.

- *BAO: xxx porque o edit xxx //
- *BAO: xxx / porque o edit xxx //

Contexto: uma chefe de cozinha conversa com sua irmã.

- *HEL: um buffet / é especialista nisso / né // yyyy //
- *HEL: um buffet / é especialista nisso / né // yyyy

Em ambos os casos retratados acima, a presença de um trecho incompreendido pode ter induzido os anotadores a marcar quebras diferentes. Apesar da relativa independência existente entre o segmental e a anotação prosódica, o desconhecimento do primeiro pode dificultar a interpretação do segundo.

- Fenômenos paralinguísticos:

Contexto: uma mulher (REG) relata ao pesquisador BAO sua experiência de dar à luz.

- *REG: falei / é / vim ter filho pelo SUS // falei / no SUS não / minha filha / no susto //
- *BAO: hhh //
- *REG: falei / é / vim ter filho pelo SUS // falei / no SUS não / minha filha / no susto //
- *BAO: hhh

Nesse caso, não havia clareza sobre a decisão correta a ser tomada com relação aos fenômenos paralinguísticos. Ficou decidido posteriormente que deveriam ser segmentados apenas quando dotados de valor ilocucionário.

B. Fatores cognitivos, insuficiência teórica e sobreposição de fala

Somados aos fatores já expostos, há ainda fatores de ordem cognitiva, que respondem tanto por discordâncias totais quanto por discordâncias do tipo *terminal x não-terminal* ou *não-terminal x nenhuma quebra*, como distração e marcações erradas (caso em que há percepção da quebra, mas anotação em contexto claramente errado), o fator insuficiência teórica por parte dos transcritores nessa fase inicial de treinamento, quando ainda não tinham suficiente contato com a teoria e o fator sobreposição de fala (que naturalmente dificulta a percepção tanto do segmental quanto do suprasegmental).

A insuficiência teórica emerge principalmente na diferenciação de retrações, que contam como quebras não-terminais, e enunciados interrompidos, contabilizados como quebras terminais. A seguir observa-se um exemplo de retração e outro de enunciado interrompido:

- *HEL: porque cê &fa [1]³ montar um cardápio / dá um trabalho danado //
- *ECR: é égua // já tem o + mas tem egüinha também //⁴

A retração, demonstrada no primeiro exemplo, é comumente confundida com um enunciado interrompido. Ela caracteriza-se pela substituição de material semântico por outro de valor semelhante (um verbo por outro, por exemplo), havendo manutenção do programa. No segundo exemplo acima, diferentemente, é claro o abandono do enunciado iniciado por “já tem o”.

Contexto: LUC pergunta a sua avó sobre a sua experiência de ser sobrinha de um famoso poeta brasileiro.

- *LUC: &he / essa [1] &he / a [1] a [1] a história / desde o começo / assim / e tudo o mais //
- *LUC: &he / essa + &he / a [1] a [1] a história / desde o começo / assim / e tudo o mais //

Após hesitar um pouco (&he), a falante troca o demonstrativo “essa” pelo artigo “a”, mantendo o programa. Por isso o trecho não deve ser interpretado como um enunciado interrompido, mas sim como uma reformulação.

C. Quebra sintática

As quebras sintáticas são anotações correspondentes ao que seria uma segmentação sintagmática – claramente errôneas, dada a incompatibilidade de modelos. Na figura a seguir, percebe-se a anotação de quebras não-terminais entre os sintagmas verbais “reduzir”:

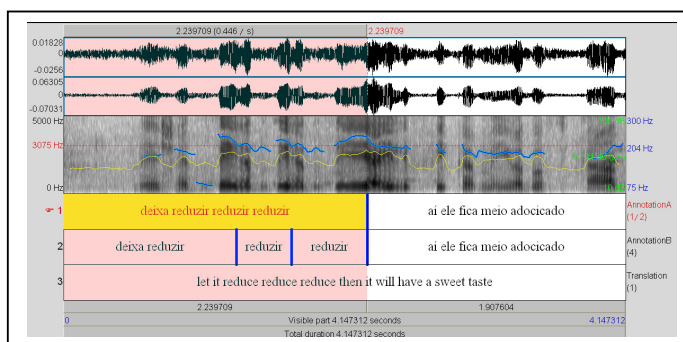


Figura 1. Quebra sintática

O fato de um anotador realizar uma segmentação em constituintes de um enunciado nos leva à hipótese de que possa haver transferência de parâmetros da sintaxe sentencial para a segmentação prosódica segundo a Teoria da Língua em Ato.

³ O número entre colchetes é referente ao número de palavras canceladas.

⁴ O símbolo + representa um perfil interrompido.

D. Fenômenos específicos ao PB

A discrepância na anotação de quebras *não-terminais x nenhuma quebra* pode ser explicada pela presença de fatores específicos ao PB, como ênfases, ritmo, co-articulação, quebras fracas e existência de marcadores discursivos. A discordância do tipo *nenhuma quebra x não-terminal* associada a eles emerge do processo de percepção e análise de um mesmo grupo tonal como diferentes unidades tonais. A análise desses fatores levou em conta alguns parâmetros acústicos, como intensidade e variação da frequência fundamental, diferentemente do processo de treinamento pelo qual passaram os anotadores, que deveriam basear-se apenas na percepção.

1) Ênfase

A ênfase representa um fator que gera desacordo na anotação prosódica quando é confundida com uma quebra. No espectrograma a seguir, vemos uma saliência no segmento /o/ da palavra “some”, representada por uma ascendência na curva que indica a frequência fundamental:

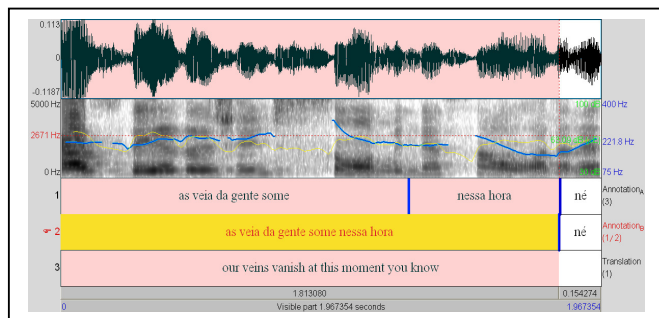


Figura 2. Ênfase

Provavelmente devido à presença da ênfase, o anotador A marcou uma quebra terminal após a saliência, que, nessa ocorrência, coincide com o foco ilocucionário.

2) Coarticulação

Por coarticulação entendemos que a produção de um segmento está articulada à produção de outro. Na figura seguinte, os limites entre segmentos /e/ das palavras “e” e “essa” são difíceis de serem distinguidos:

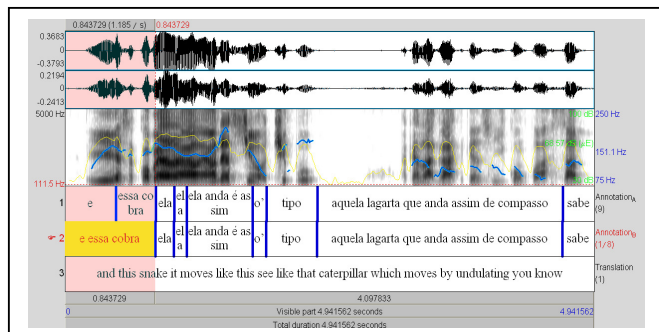


Figura 3. Coarticulação

Nesse caso, devido à coarticulação, o anotador B não assinalou uma quebra não-terminal entre “e” e “essa”. Estudos mostram, no entanto, que no PB há compatibilidade de quebra prosódica e coarticulação⁵.

3) Quebra fraca

Como o próprio nome indica, a quebra fraca é muito sutil e, por isso, muitas vezes imperceptível para alguns. Na figura seguinte, o anotador A assinalou uma quebra não-terminal entre os segmentos “brinca” e “que”:

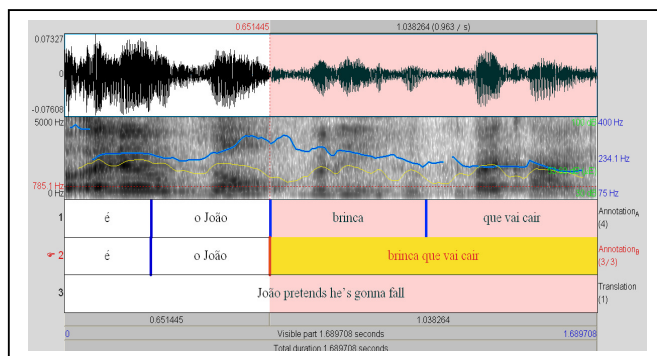


Figura 4. Quebra fraca

A presença da quebra não é clara no espectrograma e certamente não delimita a fronteira entre duas unidades informacionais.

4) Ritmo

A figura abaixo mostra um trecho em que houve bastante discrepância com relação à anotação de quebras simples. A discrepância diz respeito à anotação de quebras não-terminais entre os grupos acentuais, confundidos com unidades tonais.

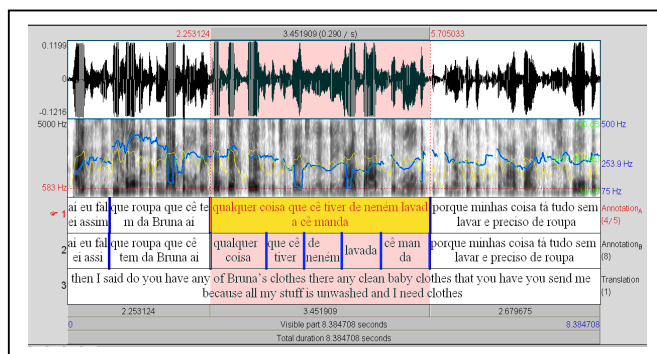


Figura 5. Ritmo

No trecho assinalado, percebe-se a presença dos grupos acentuais, cada um com uma sílaba proeminente. A hipótese é a de que a presença dos focos em cada grupo acentual tenha feito com que alguns anotadores os segmentassem como unidades tonais da Teoria da Língua em Ato. Outra evidência para essa hipótese é a ausência, nos grupos acentuais demarcados, das características que definem as unidades tonais de valor informacional segundo

⁵ Cf. [8]

a Teoria - os grupos acentuais assinalados acima não compartilham da distribuição no enunciado nem do perfil entoacional das unidades.

5) *Marcador discursivo*

Marcadores discursivos são entidades pragmáticas que ajudam a regular o discurso. Em línguas como o espanhol e o italiano⁶ o marcador é necessariamente realizado em uma unidade tonal separada do restante do enunciado. No português do Brasil, nota-se que a presença da quebra não-terminal é contingente. Na figura seguinte, observa-se a marcação de uma quebra não-terminal, separando o marcador do restante do enunciado:

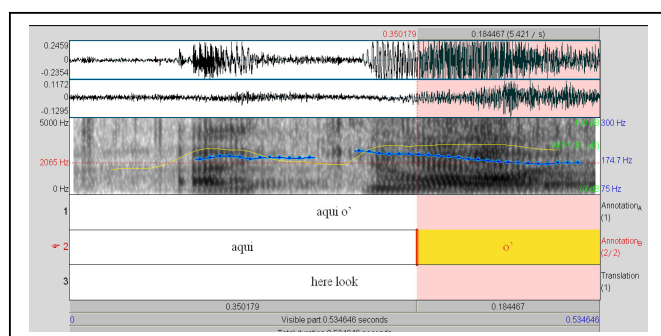


Figura 6. Marcador discursivo

A partir da análise do espectrograma e do áudio correspondente não se percebe nenhuma indicação de quebra. A sequência é realizada em uma só unidade tonal. Nota-se, também, uma possível transferência das regras de pontuação da escrita, em que muitas vezes se isola o marcador por vírgulas.

VI. CONCLUSÕES

A validação estatística da segmentação substancia a hipótese de que unidades pragmáticas como os enunciados e as unidades tonais (em vez de sentenças e sintagmas, na escrita) são o cerne da organização informacional na fala espontânea, o que embasa a segmentação do corpus nessas unidades. A forte percepção de quebras terminais está ligada à percepção da força ilocucionária, que emerge pela entoação, já que é a presença dessa força que permite a autonomia de *chunks* na comunicação.

A emergência de fatores específicos ao PB como possíveis causas da discordância entre anotadores e a sua descrição revela-se interessante. O fato de o acordo entre anotadores ser maior entre os membros do grupo 1 sugere que o conhecimento lingüístico possa influenciar a segmentação da fala nessas unidades.

VII. REFERÊNCIAS

[1] J. L. AUSTIN. How to do things with words. London: Oxford University Press, 1962.

[2] E. Cresti, Corpus di Italiano parlato, vol.1. Firenze: Accademia della Crusca, 2000, pp. 41-166.

[3] E. Cresti; M. Moneglia (eds.) C-Oral-Rom: Integrated Reference Corpora For Spoken Romance Languages. Amsterdam: John Benjamins, 2005, 303 p.

[4] J. L. Fleiss "Measuring nominal scale agreement among many raters". Psychological Bulletin, 76, 378-382, 1971.

[5] P. Martin. WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. Lisbon: LREC. May 2004. Disponível em: <http://lablita.dit.unifi.it/coralrom/papers/index.html>. Acesso em: 6 Set 2007.

[6] M. Moneglia; T. Raso; M. Mittmann; H. Mello. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-ORAL-ROM. Speech Prosody, 2010, Chicago. <http://aune.lpl.univ-aix.fr/~sprogisp2010/program.htm#W1>

[7] T. Raso; H. Mello. "The C-ORAL-BRASIL corpus". In: M. Moneglia; A. Panunzi (orgs.) Bootstrapping Information from Corpora in a Cross Linguistic Perspective. Firenze: Firenze University Press, 2010, pp. 193-213.

[8] T. Raso; M. Mittmann. "Validação estatística dos critérios de segmentação da fala espontânea no corpus C-ORAL-BRASIL". RELIN 17, 2009, pp. 73-91.

[9] J. t'Hart; R. Collier; A. Cohen. A perceptual study on intonation: an experimental approach to speech melody. Cambridge: Cambridge University Press, 1990.

⁶ Cf. [6]

