Proposta de delimitação de elementos discursivos baseada na prosódia

Luciana Lucente
Universidade Federal de Alagoas

This article aims to show the consistent alignment between an automatic segmentation of speech in stress groups and the delimitation of discourse segments according to Grosz and Sidner model (1986). The comparison between the manual segmentation of spontaneous speech in discourse segments that carry specific purposes and the automatic segmentation of stress groups shows a persistent concentration of a specific number of stress groups in each discourse segment. Based on the previous findings (Lucente 2012) this article compares a manual delimitation of discourse segments and an automatic delimitation of the same discourse segments according to the correlation between the number of stress groups in each discourse segment. The results point to a robust coincidence between the two segmentations.

Keywords: discourse segments, prosody, stress groups, phrase stress

1. Introdução

A análise da conversação explora diversas vertentes sob o mesmo título. Diversos trabalhos da área exploram teorias Bakhtinianas e levam em consideração elementos ideológicos e da anunciação, outros tendem para a linguística textual segundo uma perspectiva textual-interativa, na qual os trabalhos de Jubran (2006), Koch (2002, 2006) e Marcuschi (2003) se inserem.

A perspectiva Bakhtiniana de análise da conversação se concentra na relação entre o sujeito falante e os gêneros textuais através dos quais nos comunicamos, seja em textos orais ou escritos. Dentre os elementos que Bakhtin (2003) lista e define, o conceito de oração descrito pelo autor interessa a este trabalho. Para Bakhtin a oração em si não tem autoria e só a partir do momento em que se torna um enunciado, em uma situação discursiva, é que passa a representar a intenção do falante. Este sim, o enunciado, seria uma unidade real do discurso.

CHIMERA. Romance Corpora and Linguistic Studies 3.2 (2016), 375-384. ISSN 2386-2629

Seguindo perspectiva semelhante sobre o que compõem o enunciado nos textos orais, a perspectiva textual-interativa, adotada por Koch (2002, 2006), Jubran (2006) e Marcuschi (2003), se preocupa com o funcionamento da língua em contextos de uso e, portanto, com o emprego da atividade discursiva na formulação de textos orais. Seguindo os pressupostos desta teoria, a construção de um texto oral é feita *online* (Jubran 2006), ou seja, os enunciados do discurso são construídos durante a atividade do falante ou dos falantes, uma vez que essa construção por ser feita em conjunto quando mais de um falante estão engajados em uma conversa.

Adotando perspectiva similar, na qual o discurso se constrói conjuntamente pelos participantes deste, porém sob um arcabouço computacional, Grosz e Sidner (1986) propõem uma análise da conversação que considera aspectos formais e funcionais. Tal proposta, conhecida como modelo de Grosz&Sidner, ou modelo G&S, visa a segmentação de textos orais em unidades discursivas menores, seguindo uma hierarquia que se apoia em três princípios: intenção, atenção e estrutura linguística.

2. Modelo de segmentação discursiva

O Modelo G&S foi desenvolvido como um modelo computacional que, ao descrever a estrutura do discurso, oferece as bases necessárias para sua descrição e 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 nos 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: 175).

A hipótese defendida nesse modelo é de que qualquer enunciado do 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.

Sendo assim, podemos dizer que o discurso possui três componentes responsáveis por sua estruturação e interação: i) estrutura da sequência dos enunciados, ou estrutura linguística (*linguisticstructure*); 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 possam determinar como um enunciado se encaixa com outras partes do discurso, possibilitando que os participantes entendam por que algo foi dito e o que isso significa (Grosz & Sidner 1986: 177), sem que sejam mencionados aspectos sobre o significado do discurso como um todo.

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. 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 uma conversa existe mais do que um único objetivo que os leva a participar de tal 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 – *discoursepurpose*).

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 – *focusspace*), 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.

O modelo G&S se aplica diretamente à análise da estrutura do discurso por meio da segmentação discursiva e da construção de uma relação hierárquica entre esses segmentos, que se conectam por meio de elementos textuais que compreendem, além de conectivos como conjunções e preposições, elementos de construção textual.

3. Segmentação automática baseada em parâmetros prosódicos

No entanto a detecção sobre quais elementos textuais se basear para a segmentação se apresenta até o momento bastante arbitrária, pois alguns segmentadores podem realizar uma segmentação mais fina, considerando

quaisquer sintagmas como elemento de quebra na segmentação, enquanto outros podem optar por uma segmentação mais ampla, considerando apenas quebras em sentenças subordinadas ou coordenadas.

Uma forma de deixar mais clara a segmentação neste método é a uma segmentação baseada na análise prosódica dos enunciados. Trabalhos como os de Lucente (2012) e Hirschberg e Litman (1993) tem mostrado que os participantes de um diálogo utilizam de informação prosódica, associada à estrutura linguística, para detectar os propósitos de cada segmento discursivo.

3.1 Metodologia

Lucente (2012) apresenta a transcrição e segmentação de um corpus de fala espontânea – o corpus VoCE (Lucente 2012) – feito manualmente seguindo os pressupostos de segmentação estabelecidos no modelo G&S. Este estudo mostrou que existe uma coincidência entre as fronteiras destas unidades discursivas e acentos frasais detectados automaticamente. Uma amostra de aproximadamente 30 minutos de fala mostrou uma tendência dos dados na qual cada segmento discursivo tem máxima concentração em dois grupos acentuais - que é o intervalo entre dois acentos frasais consecutivos. A Figura 1 ilustra esses dados.

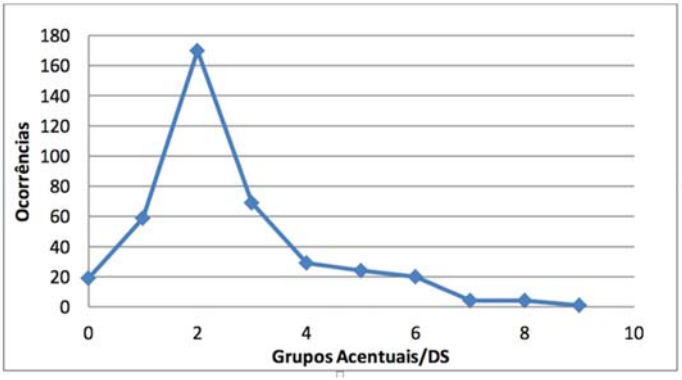


Figura 1. Concentração de ocorrências de grupos acentuais por segmentos discursivos (Lucente 2012)

Para o presente trabalho foram usados áudios extraídos do corpus VoCE (Lucente 2012). As gravações utilizadas tem duração total aproximada de 10 minutos (10 arquivos com aproximadamente 1 minuto de duração), esão compostas pela fala espontânea de 5 diferentes falantes, sendo 2 do sexo

feminino e 3 do sexo masculino, com idades aproximadas entre 30 e 50 anos. Assim como todas as gravações que compõem o Corpus VoCE, as utilizadas nesta análise são classificadas como sendo exemplos de fala espontânea por terem sido obtidas a partir de *podcasts* de programas de rádio. Os programas de rádio selecionados contaram com a participação de convidados que foram entrevistados em estúdio, o que garantiu a qualidade da gravação e a espontaneidade da fala. A fala dos entrevistadores não faz parte desse corpus.

A segmentação destas gravações é um fator imprescindível para uma proposta de delimitação automática dos segmentos da fala em unidades que carregam propósitos comunicativo, como pressupõe a teoria G&S. Sendo assim, as gravações utilizadas foram segmentadas primeiramente em unidades vogal-a-vogal, ou unidades V-V (Barbosa *et al.* 2005; Marcus 1981).

Tal técnica consiste na segmentação dos áudios em unidades que vão do início de uma vogal até o início da vogal seguinte, fazendo com que o segmento englobe uma vogal e a consoante seguida a esta. Dessa forma é possível marcar, por meio da segmentação, exatamente a transição consoante-vogal, ou seja, a transição C-V.

A marcação da transição C-V é importante para esta técnica de análise, pois possibilita a detecção do início de cada vogal, que funciona como ponto de ancoragem para a coordenação e o alinhamento entre os elementos envolvidos na produção da fala (Barbosa 2006). A transição C-V funciona como ponto de ancoragem, pois i) é o ponto de maior acúmulo de amplitude da atividade vocálica (*beat*); ii) é a transição que o ouvido humano está mais apto a perceber (Marcus 1981); iii) é o ponto de alinhamento entre o ritmo da fala e o ambiente externo, como mostram os experimentos com *p-centre*, conforme os trabalhos de Marcus (1981), Barbosa *et al.* (2005) e Pompino-Marschall (1989).

A segmentação dos arquivos em unidades V-V utilizando o *software* Praat[1] possibilita posteriormente a segmentação automática dos arquivos em grupos acentuais por meio do programa *SGDetector*, desenvolvido e descrito em Barbosa (2006), que funciona como um *script* de comandos quando aplicado ao Praat.

Um grupo acentual é composto pelo intervalo entre dois acentos frasais, que por sua vez são determinados por períodos de maior duração dentre as médias de duração das vogais. Dessa forma, o que o programa *SGDetector* faz é detectar esses pontos de maior duração das vogais de acordo com sua duração média para a língua em questão (o português brasileiro, neste caso). A detecção destas durações é possível devido à segmentação prévia do corpus em unidades V-V,

[1] Praat: doing phonetics by computer (version 5.1.05). http://www.praat.org/

pois o programa consegue medir a duração do segmento V-V e detectar se a soma das durações das unidades para um determinado segmento ultrapassa a duração prevista. Veja exemplo desta segmentação na Figura 2 abaixo.

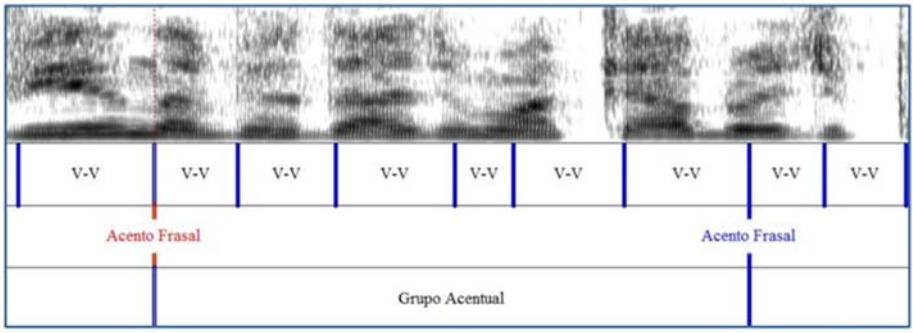


Figura 2. Exemplo de como se organizam na segmentação: i) as unidades V-V; ii) os acentos frasais; e iii) os grupos acentuais (Lucente 2012)

Esta técnica de segmentação ainda conta com tarefas manuais, pois a segmentação em unidades V-V, apesar de poder ser feita com auxílio do programa *BeatExtractor* (Barbosa 2006), que marca os inícios das vogais baseando-se no *beat*de cada vogal, necessita de alguns ajustes manuais. A transcrição fonética dos dados também é feita manualmente, utilizando caracteres específicos para esta técnica e não o alfabeto IPA.

Para maior produtividade na segmentação dos dados de fala seria desejável uma técnica que unisse a segmentação automática em unidades V-V e a marcação dos grupos acentuais em um único passo. No entanto, para o presente trabalho o processo ainda foi feito de forma parcialmente manual.

Após esta etapa da segmentação no Praat os arquivos analisados apresentam as seguintes camadas de notação: a primeira camada com a segmentação V-V, a segunda com segmentação por palavras e a terceira com a marcação dos acentos frasais, conforme a Figura 3.

4. Resultados

Seguindo esta técnica de segmentação, os resultados prévios encontrados coincidem em parte com os encontrados em Lucente (2012), que tinha como objetivo demonstrar que fronteiras discursivas, prosódias e entoacioanis estavam alinhadas, e tinham como elemento atrator os acentos frasais.

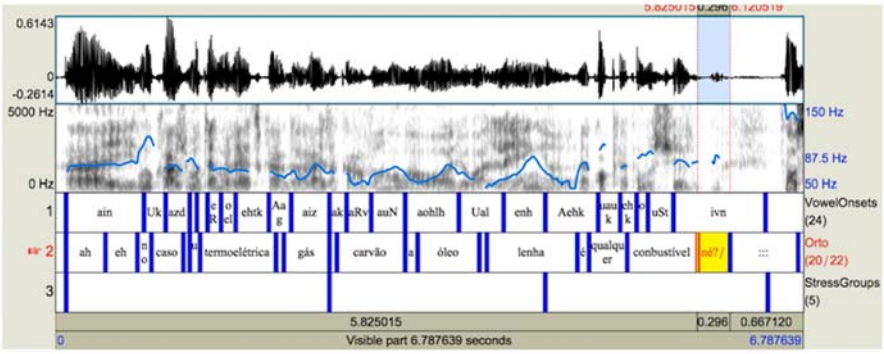


Figura 3. Exemplo de como se organizam na segmentação em três camadas na janela do Praat

Aqui o objetivo foi delimitar segmentos discursivos baseando-se na marcação dos grupos acentuais, diferente de Lucente (2012), que tinha como objetivo observar o alinhamento entre contornos entoacionais e fronteiras de grupos acentuais. Para tanto, após a segmentação dos áudio em unidades V-V, a partir dos quais se obteve a marcação dos grupos acentuais, foi verificada a coincidência entre a marcação automática dos grupos acentuais - correlato prosódico - e a delimitação dos segmentos discursivos - correlato discursivo - de acordo com a proposta de Grosz & Sidner (1986).

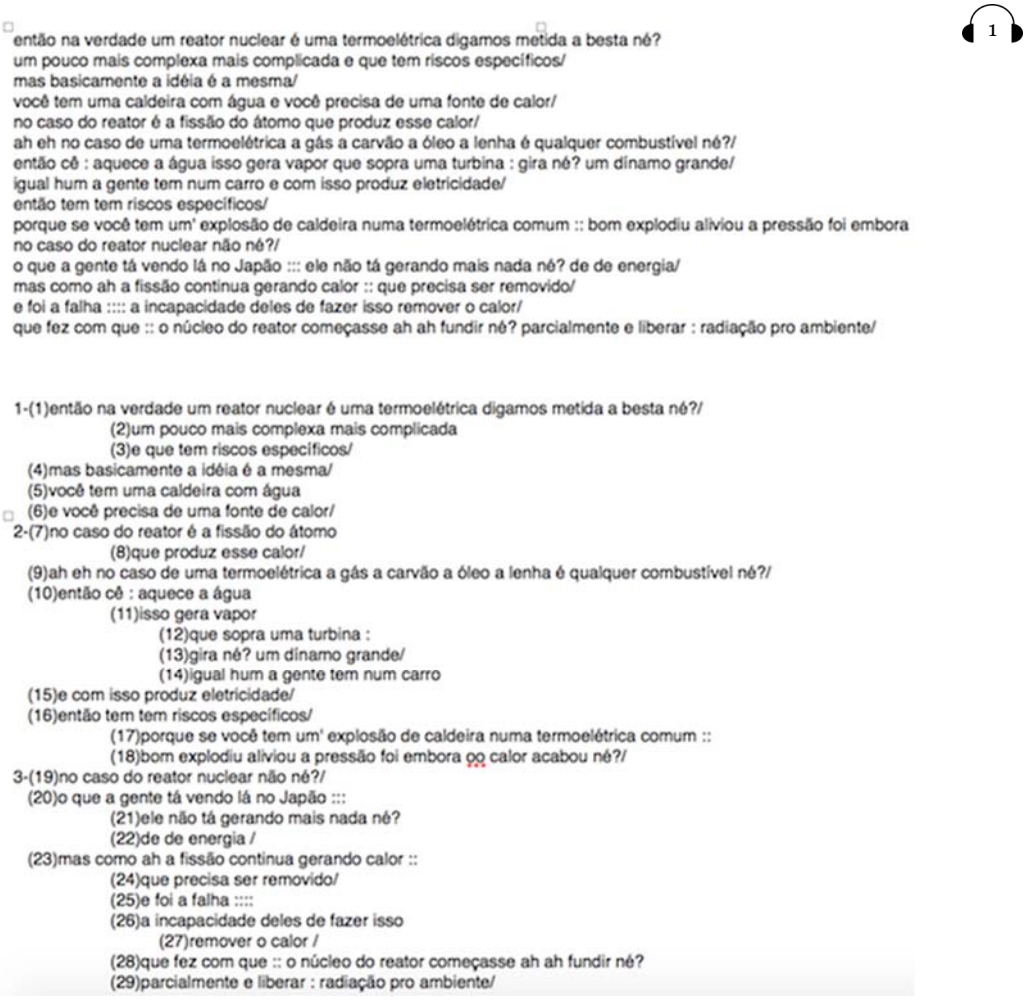
Os resultados encontrados mostraram uma coincidência alta entre segmentações feitas manualmente e as feitas com base nos grupos acentuais.

Em alguns casos a segmentação discursiva manual se apresenta mais estreita que a marcação dos grupos acentuais. Por exemplo, para o texto (1) *reator_nuclear* a segmentação manual marcou 29 segmentos discursivos, enquanto a segmentação em grupos acentuais encontrou 16 segmentos coincidentes, marcados por "/". Entretanto, para o texto (2) *pacce* a segmentação manual encontrou 16 segmentos discursivos, que coincidiram com 15 dos segmentos encontrados automaticamente. As segmentações podem ser vistas a seguir.

Os resultados das segmentações, tanto manuais, quanto automáticas, mostram que o texto oral se segmenta em momentos de hesitações, em parênteses, paráfrases, marcadores discursivos, como "tipo", "bom" e "sei lá", artigos, conjunções e elementos de negação. A detecção automática de tais elementos pode auxiliar no estabelecimento de critérios de segmentação

sintático-prosódicos para dados do português brasileiro mais claros para os usuários do modelo G&S do que os estabelecidos pelo próprio modelo.

1. Segmentação automática (acima) e manual (abaixo) do texto reator_nuclear, nas quais os segmentos coincidentes estão marcados por "/":



2. Segmentação automática (acima) e manual (abaixo) do texto pacce, nas quais os segmentos coincidentes estão marcados por "/":


```
primeiro acho assim/
ahhh a gente pode até listar/
algumas coisas que são clássicas/
tipo n'sei uma calça jeans/
uma camiseta branca/
um vestidinho preto/
tô falando p'ras meninas, ne?/
um trend coat ou uma jaqueta perfecto/
um terninho são coisas que /
não saem de moda/
mas as proporções /
mudam totalmente/
ntão não adianta você achar que um/
terninho de hoje/
é igual o terninho dos/
anos oitenta ou dos anos sessente que não é
```

```
1 - (1)primeiro acho assim
    (2)ahhh a gente pode até listar/algumas coisas que são clássicas
    (3)tipo n'sei/
        (4)uma calça jeans
        (5)uma camiseta branca
        (6)um vestidinho preto
                (7)tô falando p'ras meninas, ne?
        (8)um trend coat ou uma jaqueta perfecto
        (9)um terninho/
2- (10)são coisas que não saem de moda
        (11)mas as proporções/ mudam totalmente
3- (12)n'tão não adianta você achar/
        (13)que um terninho de hoje
        (14)é igual o terninho/dos anos oitenta ou dos anos sessenta/
                (15)que não é.
```

5. Discussão

A análise preliminar dos dados de segmentação, tanto manual, baseada no modelo de Grosz e Sidner, quanto a segmentação automática, realizada com o programa SG Detector (Barbosa 2006), mostra resultados relevantes para uma proposta de segmentação discursiva automática com base na prosódia do PB.

A observação dos dados mostra que existe coincidência consistente entre acentos frasais - que delimitam grupos acentuais - e unidades discursiva providas de sentido.

Esses resultados podem facilitar a segmentação de dados de fala com propósitos discursivos à medida que apontam para a construção de um programa que segmente dados de fala com base em grupos acentuais. Lucente (2012) mostra que essa segmentação também se alinha a aspectos relevantes da curva

entoacional do PB, pois a cada início de grupos acentual, ou segmento discursivo, estão alinhados contornos >LH.

Sendo assim, os próximos passos a seguir nesta pesquisa serão: i) um levantamento estatístico em um corpus grande sobre a relação entre grupos acentuais por segmentos discursivos; ii) desenvolver um programa que segmente automaticamente os segmentos discursivos tento como base os acentos frasais; iii) associar marcações entoacionais a estes segmentos.

Referências

Bakhtin, M. 2003. Os gêneros do discurso. In M. Bakhtin (ed.), *Estética da criação verbal*. São Paulo: Martins Fontes, 261-306.

Barbosa, P.A. 2006. *Incursões em torno do ritmo da fala.* Campinas: Pontes.

Barbosa, P.A., Arantes, P., Meireles, A.R. & Vieira, J.M. 2005. Abstractness in speech-metronome synchronisation: p-centres as cyclic attractors. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*. Lisbon: ISCA, 1441-1444.

Grosz, B.J., & Sidner, C.L. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics* 12(3): 175-204.

Hirschberg, J. & Litman, D. 1993. Empirical Studies on the Disambiguation of Cue Phrases, *Computation Linguistics* 19(3): 501-530.

Jubran, C.C.A.S. 2006. A Perspectiva Textual-Interativa. In C.C.A.S. Jubran & I.G.V. Koch (eds), *Gramática do Portugues Culto Falado no Brasil. Construção do Texto Falado*, Vol. 1. Campinas: Editora da Unicamp, 27-36.

Koch, I.G.V. 2006. Especificidade Do Texto Falado. In C.C.A.S. Jubran & I.G.V. Koch (eds), *Gramática do Portugues Culto Falado no Brasil. Construção do Texto Falado*, Vol. 1. Campinas: Editora da Unicamp, 39-46.

Koch, I.G.V. 2002. *Desvendando os Segredos do Texto.* São Paulo: Cortez.

Lucente, L. 2012. *Aspectos Dinâmicos da Fala e da Entoação no Portugues Brasileiro*. Campinas: Editora da Unicamp.

Marcus, S.M. 1981. Acoustic Determinants of Perceptual-Center (P-Center) Location. *Perception and Psychophysics* 30(3): 247-256.

Marcuschi, L.A. 2003. *Análise da Conversação*. São Paulo: Ática.

Pompino-Marschall, B. 1989. On the Psychoacoustic Nature of the P-Center Phenomenon. *Journal of Phonetics* 17: 175-192.