



OLIVEIRA JR., Miguel; CRUZ, Regina; SILVA, Ebson Wilkerson. **A relação entre a prosódia e a estrutura de narrativas espontâneas: um estudo perceptual.** *Revista Diadorim / Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro*. Volume 12, Dezembro 2012. [<http://www.revistadiadorim.lettras.ufrj.br>]

A RELAÇÃO ENTRE A PROSÓDIA E A ESTRUTURA DE NARRATIVAS ESPONTÂNEAS: UM ESTUDO PERCEPTUAL

Miguel Oliveira Jr. (UFAL)

Regina Cruz (UFPA)

Ebson Wilkerson Silva (UFAL)

RESUMO: O objetivo central do presente estudo foi examinar em que medida a estrutura de narrativas orais espontâneas é reconhecida por examinadores inexperientes e leigos, e, como corolário, demonstrar qual o papel que as pistas prosódicas têm nesse processo de reconhecimento. Para isso, um protocolo experimental foi desenvolvido, em que narrativas em diferentes condições (só texto, texto e áudio, só áudio e só áudio filtrado) foram apresentadas. Participaram do estudo 48 informantes. Os resultados indicam que, mesmo sem acesso a informações lexicais, sintáticas e semânticas, as pessoas identificam, de forma consistente, uma estrutura discursiva em narrativas espontâneas, o que sugere ter a prosódia um relevante papel na percepção da estrutura discursiva em narrativas orais espontâneas.

PALAVRAS-CHAVE: narrativas espontâneas, estrutura da narrativa, prosódia, percepção.

ABSTRACT

The present paper reports the results of a perceptual experiment that was conducted in order to find out to what extent naïve, untrained labelers can benefit from prosodic information present in spontaneous narratives so as to derive an underlying structure associated with them. Results from inter-rater agreements suggest that even when the lexical, syntactic, and semantic information of a narrative is blocked, labelers can still identify its underlying structure, which clearly indicates that listeners make substantial use of prosodic information in the task of identifying the structure of spontaneous narratives.

KEYWORDS: spontaneous narrative, structure, prosody, perception.

Introdução

O grau de coesão sintática, semântica e/ou pragmática entre palavras em um enunciado determina se elas pertencem ou não a um mesmo constituinte linguístico. De igual modo, as relações que se estabelecem entre enunciados vão definir se eles fazem parte de uma estrutura linguística ainda maior, geralmente denominada “discurso”. Nesta perspectiva, discurso é considerado uma estrutura linguística composta por unidades hierarquicamente dispostas que preservam uma orientação similar.

Vários estudos têm demonstrado que elementos prosódicos (pausa, tons de fronteira, diferença de *pitch*, por exemplo) são utilizados para delimitar macroestruturas discursivas em textos¹ orais de diferentes gêneros (Geluykens & Swerts, 1994; Grosz & Hirschberg, 1992; Passonneau & Litman, 1993). O uso de tais elementos prosódicos contribui para a identificação semântica das unidades discursivas, assim como explicita quais as intenções do falante para com sua audiência, facilitando, desse modo, o processo da comunicação.

Embora já exista um número considerável de trabalhos sugerindo que o falante utiliza-se da prosódia para estruturar o fluxo da informação discursiva, ainda há pouca evidência de que tal procedimento tenha relevância perceptual. É sabido que qualquer estudo envolvendo aspectos prosódicos da fala deve considerar não apenas a produção, mas também a percepção de tal fenômeno. Assim, a relevância das variáveis prosódicas na demarcação da estrutura discursiva, tal como observada em análises acústicas, só pode ser plenamente validada depois da consideração de seu papel sob a perspectiva da percepção.

Ao analisar, em um estudo de produção, uma série de variáveis prosódicas como possíveis elementos estruturadores da narrativa oral espontânea, Oliveira Jr.(2000) demonstrou que algumas delas exercem papel crucial na estruturação discursiva, segmentando o texto narrativo em seções semanticamente independentes. O propósito do presente estudo é ampliar a análise apresentada em Oliveira Jr.(2000), examinando até que ponto o ouvinte pode se beneficiar das marcas prosódicas presentes na narrativa espontânea para daí derivar a estrutura subjacente a esse tipo de discurso.

A hipótese central aqui defendida é que a prosódia facilita o ouvinte a perceber mais efetivamente a estrutura do texto narrativo, o que obviamente acarreta numa melhor compreensão do discurso como um todo. Confirmada essa hipótese, é validada a proposta, parcialmente testada em Oliveira Jr.(2000), de que a prosódia tem papel fundamental na elucidação da estrutura subjacente básica do texto narrativo.

1. Embora os termos “discurso” e “texto” sejam usados na literatura para designar diferentes fenômenos linguísticos, no presente artigo serão usados indistintamente.

Estado da Arte

Estudos experimentais testando o peso das variáveis prosódicas na percepção da estrutura do discurso falado têm aparecido ainda com certa infrequência. O maior número de publicações relacionadas a tal tipo de pesquisa é relacionado ao grupo do *Institute for Perception Research* (IPO) e seus associados, e tem as mais das vezes como foco de estudo as línguas holandesa e inglesa. Apoiados, na maioria dos casos, por uma análise acústica dos dados, tais estudos buscam compreender de que maneira os elementos prosódicos são percebidos na fala e com que proporção tais elementos estão associados a uma função delimitadora/estruturadora. Alguns desses estudos são brevemente exemplificados a seguir.

Collier (1993), tomando como material de análise um *corpus* constituído de sentenças lidas por três diferentes locutores, procurou observar de que maneira as variáveis prosódicas aí presentes estão associadas à noção de finalidade sob a perspectiva da percepção. Usando uma unidade empírica denominada *Perceptual Boundary Strength* (PBS), que é a medida usada pelos participantes do experimento para designar que partes da fala soam mais “finalizantes”, ele demonstrou que quanto maior os valores da PBS, tal como julgado pelos participantes em uma escala de 1 a 10 pontos, maior o número de elementos prosódicos presentes na parte da fala associada àquela PBS. A conclusão é de que as sentenças que ocupam uma posição final em um monólogo são geralmente acompanhadas de uma série de marcas prosódicas que, não apenas do ponto de vista acústico, mas sobretudo pelo prisma da percepção, constituem evidência de sua posição no discurso (ver também Collier, de Pijper, & Sanderman, 1993).

Utilizando um *corpus* mais adequado para esse tipo de pesquisa, fala produzida espontaneamente², Sweerts, Collier & Terken (1994) demonstraram que os sujeitos não apenas identificam com sistematicidade o final de uma unidade discursiva maior, mas também são capazes de prever quando “há muito mais [sentenças] por vir” ou quando “há apenas uma mais [sentença] por vir” em um monólogo. Estes resultados sugerem que a prosódia é usada principalmente para indicar continuidade ou finalidade no discurso (Brown, 1980; Yule, 1980), e é, talvez, demasiado simplista.

No intuito de testar se o julgamento dos sujeitos em experimentos de percepção de elementos prosódicos como dispositivos estruturadores é influenciado pela informação lexical, sintática e semântica do texto, de Pijper & Sanderman (1994) compararam resultados obtidos de experimento realizado com fala deslexicalizada (através da manipulação dos picos espectrais do material de análise) com os oriundos de experimento com fala normal e leitura em voz alta. A alta correlação entre os valores de

2. Para uma discussão acerca da importância do uso de material espontâneo no estudo da prosódia, conferir a coletânea de artigos editada por Sagisaka, Campbell & Higuchi (1997).

PBS nos dois tipos de experimentos evidenciou que a informação segmental não tem influência significativa na percepção da estrutura discursiva. A prosódia seria, desse modo, suficiente para que se derive a estrutura subjacente ao discurso oral.

Também fazendo uso de fala ininteligível, Swerts & Geluykens (1994) testaram a habilidade de ouvintes de reconhecer fronteiras discursivas em textos espontâneos baseados unicamente na informação suprasegmental dos mesmos. Ao contrário de Pijper & Sanderman (1994), aqui foi utilizada uma técnica relativamente mais simples: os participantes do experimento foram solicitados a dizer o mais rápido possível quando eles julgassem que uma dada unidade discursiva havia chegado ao fim. Os resultados também revelaram que ouvintes conseguem derivar a estrutura do discurso a partir da prosódia. Correlações estatisticamente significantes foram encontradas para a percepção de fronteiras discursivas e a presença de pausas longas e variação de tom.

A influência dos elementos prosódicos na percepção da estrutura discursiva foi também testada em Swerts (1996), que, ao invés de mascarar a fala para fins de comparação, utilizou resultados de experimentos feitos em condições diferenciadas. Participantes do experimento foram divididos em dois grupos: um que teve acesso unicamente à transcrição do texto discursivo em análise e outro que teve acesso não apenas à transcrição, como também ao texto falado *per se*. Os resultados da comparação entre as condições acima descritas mostraram que o acesso ao texto falado incorre numa maior concordância entre participantes acerca da estrutura textual. Isso implicaria que a informação prosódica contida no texto faz com que a estrutura do discurso se torne ainda mais clara sob o ponto de vista da percepção. Usando um modelo linear aditivo para testar de que modo os diferentes elementos prosódicos se correlacionam com os resultados obtidos no teste perceptual, Swerts (1996) notou que, entre os elementos investigados, a pausa teria um maior peso na decisão do ouvinte no que concerne à estrutura do discurso. A diferença de tom vem em segundo lugar, seguida do tom de fronteira.

Grosz & Hirschberg (1992), Hirschberg & Grosz (1992) e Grosz, Hirschberg & Nakatani (1994) estudaram a relação entre a variação acústico-prosódica e a estrutura do discurso sob o prisma da percepção. Usando um modelo de análise independente (Grosz & Sidner, 1986) que distingue dois modos de processamento do discurso - um no nível local e outro no nível global, estas autoras examinaram a relação entre elementos prosódicos e estrutura discursiva baseada no resultado de experimentos feitos com participantes que tiveram acesso apenas à transcrição do texto falado e de participantes que, além da transcrição, puderam ouvir a gravação dos textos transcritos. Elas encontraram uma associação estatisticamente significativa entre aspectos da variação de tom, amplitude e variáveis temporais e a estrutura global e local do discurso em análise. Como em Swerts (1996), foi observado que o julgamento dos participantes que tiveram acesso à gravação dos textos refletiam ainda mais significativamente a

estrutura textual derivada da informação prosódica, sugerindo que os ouvintes fazem uso de tal informação na tarefa de indicar a estrutura do discurso em teste (Litman & Passonneau, 1993; Litman & Passonneau, 1995; conferir também Passonneau & Litman, 1993; Passonneau & Litman, 1997).

Em um estudo piloto realizado para testar a reprodutibilidade de um modelo de segmentação do discurso baseado em intenções (Passonneau & Litman, 1997), Silva & Oliveira (2011) demonstraram que as pessoas concordam de forma bastante significativa acerca da estrutura da narrativa como um reflexo das intenções do falante. Para além disso, os autores observaram que a concordância tende a ser maior quando, além da transcrição, as pessoas têm acesso ao material sonoro das narrativas analisadas. Segundo eles, isso se dá como resultado do papel que a prosódia exerce na percepção da estrutura da narrativa. No presente estudo, uma extensão do experimento descrito em Silva & Oliveira (2011), exploramos as relações entre intenções, estrutura narrativa e prosódia, utilizando um protocolo que inclui não apenas o acesso às transcrições de narrativas acompanhadas ou não de áudio, mas também de narrativas desacompanhadas de descrição e de narrativas cujo conteúdo segmental é filtrado, tornando-o ininteligível. O objetivo é examinar em que medida a estrutura de narrativas orais espontâneas é reconhecida por examinadores inexperientes e leigos, e, como corolário, demonstrar qual o papel que as pistas prosódicas têm nesse processo de reconhecimento.

É importante, antes de passarmos para a próxima seção, em que são apresentados os procedimentos metodológicos do estudo, que se discuta um conceito central para este artigo: o da *espontaneidade da fala*, uma vez que trabalharemos com narrativas espontâneas. A definição do que caracteriza o discurso “espontâneo” e os métodos que são empregados para elicitá-lo são temas cercados por controvérsias. Em geral, nos estudos da linguagem, a fala é rotulada como sendo espontânea ou lida. Esta classificação binária é enganosa, porque não leva em conta o *continuum* em que a fala de fato se insere no que respeita à sua modalidade de uso (Fujisaki, 1997; Laan, 1997). Muito comumente, a literatura costuma designar o discurso como espontâneo quando ele apresenta uma série de aspectos linguísticos que estão frequentemente presentes em enunciados não roteirizados (ou minimamente preparados em termos de organização e expressão dos pontos de informação e de comunicação), tais como a ocorrência de rupturas e reparos, o uso frequente de contrações, uma sintaxe mais relaxada e a incidência de pausas preenchidas, como “uhm” e “eh”. Estudos têm demonstrado que as pessoas classificam de maneira sistemática a fala, quando solicitadas a utilizar uma classificação binária (espontânea *versus* lida) (Laan 1997; Levin, Schaffer & Snow, 1982), o que poderá indicar a validade de tal classificação.

Esta controvérsia não é recente: remonta às discussões de Labov (1971) sobre o paradoxo do observador. De acordo com Labov (1971), há algumas técnicas que podem ser utilizadas para contornar o paradoxo do observador em trabalho de campo. Uma das mais bem-sucedidas e, por essa razão,

amplamente utilizada na pesquisa (socio)linguística, é a elicitación de narrativas. As narrativas são consideradas excelentes exemplares de fala natural porque ao contar uma história, as pessoas transportam-se para o universo da narrativa e, ao fazê-lo, não dão tanta importância ao contexto da gravação. O contexto em que a narrativa aparece durante uma gravação em trabalho de campo pode, no entanto, interferir na qualidade da mesma.

Quando alguém decide contar uma história durante uma conversa, sem que seja solicitado a fazê-lo, deixa transparecer o julgamento que faz acerca do que acredita ser suficientemente interessante para justificar o longo turno conversacional que ocupará. Isto, evidentemente, impõe ao falante uma maior responsabilidade, uma vez que ele terá, necessariamente, de fazer um esforço para evidenciar que a narrativa que está contando é não apenas relevante para a conversa em curso, mas também digna de ser ouvida. Narrativas que são contadas em resposta a uma solicitação explícita podem, por outro lado, ter características completamente diferentes do que aquelas que aparecem naturalmente na conversa.

Ao contrário das narrativas espontâneas, o objetivo central de histórias elicitadas é responder a uma pergunta. A princípio, não se poderia esperar de um falante nesse contexto mais que isso. Desse modo, narrativas elicitadas podem ser qualitativamente diferentes de narrativas espontâneas, ou não elicitadas. De acordo com Wolfson (1979), as pessoas sabem que regras são apropriadas em entrevistas como eventos de fala. Isso explicaria o fato de algumas narrativas, quando elicitadas, assumirem a forma de resumos: são geralmente mais curtas, pontuais e exibem pouquíssimos detalhes, assim como respostas às perguntas em uma entrevista devem ser.

O presente estudo assume a mesma definição de Labov e Waletzky (1967) para a classificação de narrativas: um método de recapitulação de experiências passadas que se caracteriza por uma sequência de proposições refletindo a sequência de situações que de fato ocorreu. Narrativas espontâneas aqui são consideradas narrativas não elicitadas, que aparecem naturalmente no discurso dos falantes.

Métodos

O material utilizado no presente estudo faz parte de um corpus de *entrevistas espontâneas* (Wolfson, 1979), em que os participantes foram solicitados a falar livremente sobre qualquer tema a partir de uma lista de 28 tópicos possíveis.²

As gravações foram feitas em uma sala acusticamente tratada utilizando um gravador profissional (Marantz PMD201) e um microfone unidirecional dinâmico (GenexxaIntertan 33-984 DCA),

2. Conferir Oliveira Jr. (2000) para uma descrição detalhada do *corpus*.

posicionado a cerca de 15-30 cm da boca dos participantes. A duração total das entrevistas variou de 45 a 62 minutos. Os participantes já conheciam o pesquisador responsável pela gravação há algum tempo, o que contribuiu consideravelmente para o alto grau de espontaneidade nas gravações.

Desse *corpus*, quatro narrativas foram selecionadas para o presente estudo. A seleção foi feita levando-se em conta um critério comum em estudos cujo objetivo é garantir um certo grau de espontaneidade nos dados de fala (Wolfson, 1979): o descarte do material inicial das entrevistas; nesse caso, os dez primeiros minutos a partir do início da gravação. Todas as narrativas apareceram naturalmente nas entrevistas, na maioria das vezes como uma ilustração de um dado argumento ou tópico. Os participantes não foram solicitados a contar histórias. No entanto, a maioria deles produziu naturalmente pelo menos uma narrativa ao longo de suas falas.

As narrativas selecionadas derivam de entrevistas de quatro diferentes sujeitos: três mulheres e um homem, todos estudantes de pós-graduação no momento da coleta de dados, variando de 25 a 37 anos de idade. Os informantes eram naturais de diferentes regiões do Brasil: um do Nordeste, um do Centro-Oeste e dois do Sul. Nenhum deles sabia os detalhes do estudo a que se destinavam as gravações.

A duração média das narrativas selecionadas para o presente estudo é de 69 segundos, sendo a mais longa delas de 114 segundos e a mais breve, de 46 segundos.³ Cada narrativa ganhou quatro diferentes condições de apresentação (C). Na condição de apresentação 1 (C1), apenas a transcrição ortográfica da narrativa, feita de uma maneira linear, sem marca de pontuação e sem indicação de paragrafeamento, é apresentada. Na condição de apresentação 2 (C2), são apresentados simultaneamente a transcrição da narrativa e o áudio a que a transcrição se refere. Na condição de apresentação 3 (C3), apenas o áudio da narrativa é apresentado. Na condição de apresentação 4 (C4), uma versão deslexicalizada da narrativa é apresentada.⁴

Os participantes do presente estudo foram recrutados de maneira voluntária, mediante cartazes afixados na Faculdade de Letras da Universidade Federal de Alagoas. As únicas restrições feitas para a participação no estudo foram: (i) a língua nativa (português); (ii) o grau de escolaridade (terceiro grau, completo ou incompleto) e (iii) a saúde auditiva (não reportar existência de qualquer tipo de distúrbio auditivo). Um total de 48 pessoas participou do estudo, sendo 32 mulheres (idades entre 20 e 27 anos) e 16 homens (idades entre 18 e 25 anos).

3. As narrativas podem ser ouvidas no seguinte repositório: http://www.archive.org/details/Pesquisa_273

4. O arquivo de áudio original foi filtrado com filtro passa-banda, resultando em um fala ininteligível, que preserva, no entanto, a informação prosódica. A filtragem foi feita no Praat. Toda a informação de frequência superior a 400Hz foi descartada.

Todos os participantes leram e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE), de acordo com a exigência do Ministério da Saúde e do Comitê de Ética em Pesquisa⁵. Após a assinatura do TCLE, receberam instruções gerais impressas acerca do experimento, seguidas de exemplos: duas narrativas adicionais, extraídas do *corpus* supramencionado.

A tarefa solicitada aos participantes desse experimento foi indicar pontos nas narrativas em que julgavam existir uma intenção, por parte do falante, de finalizar uma unidade comunicativa. Nenhuma definição para a noção de unidade comunicativa foi apresentada; os participantes foram instruídos a indicar as fronteiras dessas unidades em caráter puramente subjetivo.

Os exemplos traziam narrativas segmentadas, seguidas de uma análise de suas partes constitutivas, considerando-se as possíveis intenções do falante. As segmentações e análises foram apresentados à guisa de ilustração, e não como modelos a serem seguidos. Também, a título de ilustração, uma das narrativas foi apresentada em sua versão filtrada, para que os informantes se familiarizassem com o que ouviriam na condição experimental 4 (C4).

Os participantes foram então informados de que receberiam quatro diferentes narrativas em quatro diferentes condições (descritas acima). Foram solicitados a ouvir, ler, ou ler e ouvir cada uma das narrativas com atenção e, em seguida, segmentá-las de acordo com o julgamento que fizessem acerca da intenção do falante de finalizar uma unidade comunicativa.⁶

Nas condições C1 e C2, em que receberam uma transcrição da narrativa, os participantes foram solicitados a fazer a segmentação na transcrição, por meio de barras transversais. Nas condições C3 e C4, em que foram expostos apenas à versão em áudio da narrativa, os participantes foram solicitados a fazer a segmentação simultaneamente à execução do áudio, pressionando a tecla *Enter* de um teclado de computador. As respostas aos estímulos apresentados nessas duas condições foram registradas no aplicativo computacional ELAN (Brugman & Russel, 2004)⁷.

Os estímulos foram apresentados de uma maneira semialeatória: cada uma das quatro narrativas que compõem o *corpus* do estudo, em diferentes condições experimentais, apareceu pelo menos três vezes em cada ordem de apresentação (4 narrativas x 4 condições x 3 ordens = 48 grupos de apresentação únicos). A apresentação semialeatória dos estímulos assegura a generalidade dos resultados.

5. O presente estudo foi avaliado pelo Comitê de Ética em Pesquisa da Universidade Federal de Alagoas, tendo sido aprovado para realização em 24/11/2010 (Processo n. 013526/2010-01).

6. Os participantes poderiam levar o tempo que julgassem necessário para fazer a segmentação de cada uma das narrativas, podendo também ouvir as versões em áudio quantas vezes julgassem necessário para completar a tarefa. No geral, todavia, os participantes ouviam as narrativas apenas duas vezes.

7. O ELAN é um aplicativo computacional desenvolvido pelo Max Planck Institute for Psycholinguistics: <http://tla.mpi.nl/tools/tla-tools/elan/>.

Estudos em discurso e processamento de discurso tradicionalmente computavam concordância interexaminadores em termos de porcentagem (Di Eugenio, 2000). Carletta (1996) foi a primeira a apontar problemas nesse tipo de cômputo para a pesquisa em processamento de discurso, por não descartar a concordância devida ao fator chance, sugerindo, em seu lugar, o emprego do teste estatístico Kappa (K), um coeficiente de concordância interexaminadores (Cohen, 1960). O coeficiente Kappa pode variar de 1 a -1, indicando concordância ou discordância completa; o valor 0 indica o acaso. Landis & Koch (1977) consideram valores maiores que 0,75 como indicativo de excelente concordância.

A fim de calcular os valores de Kappa para o presente estudo, as narrativas que compõem o *corpus* foram segmentadas em palavras, sendo o fim de cada palavra considerado uma potencial fronteira discursiva. O objetivo do estudo foi então analisar de que maneira os participantes/examinadores concordam no julgamento que fazem sobre a localização de fronteiras discursivas em narrativas.⁸

Resultados

A figura 1 abaixo traz os resultados da concordância entre os participantes do experimento quando expostos às narrativas na condição 1 (C1), em que tiveram acesso apenas à transcrição da narrativa:

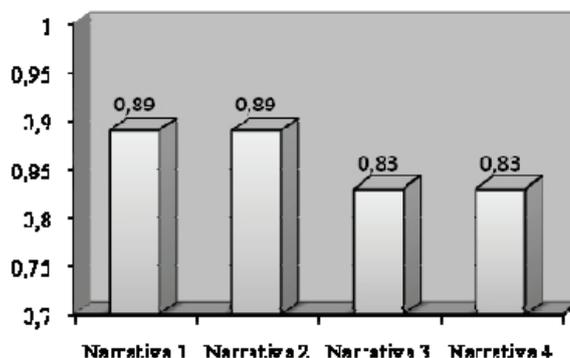


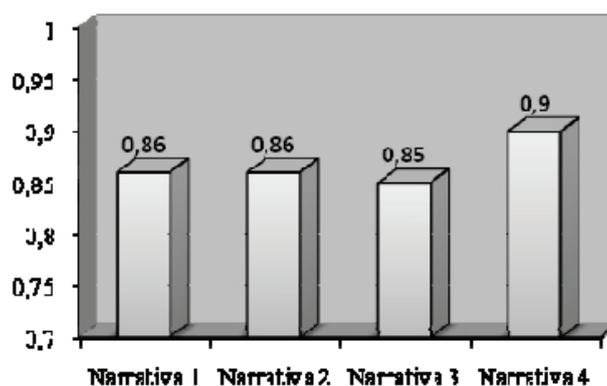
Figura 1

Valores de kappa (K) para cada uma das narrativas na Condição 1 (C1)

Os resultados na Figura 1 indicam claramente que os examinadores concordam entre si, de maneira estatisticamente significativa, acerca de como as narrativas são segmentadas em termos de intenções do falante. Esses resultados corroboram o que reporta a literatura na área (Collier, 1993; Collier, de Piyper, & Sanderman, 1993; Sweerts, Collier & Terken, 1994).

8. A localização exata das fronteiras nas condições C3 e C4 foi estabelecida a partir da consideração dos resultados gerais obtidos nos julgamentos de fronteiras nas condições C1 e C2, levando-se em conta uma latência média padrão em experimentos com estímulos auditivos de 200ms. (Campbell, Winkler & Kujala, 2007).

A figura 2 a seguir mostra os resultados de concordância entre examinadores para cada uma das narrativas na condição 2 (C2), em que tiveram acesso tanto à transcrição quanto ao áudio da narrativa:



Valores de kappa (K) para cada uma das narrativas na Condição 2 (C2)

Os resultados reportados acima, com uma população bem maior que a utilizada em Silva & Oliveira (2011), não reiteram em sua totalidade a tendência observada naquele estudo de que a concordância entre examinadores tende a ser maior quando estes têm acesso ao material sonoro das narrativas analisadas. Em alguns casos, na verdade, os resultados são opostos àquilo que foi verificado no estudo piloto: o grau de concordância entre examinadores em duas narrativas na C2 é menor que na C1 no presente estudo. Vale salientar, todavia, que nesses casos a diferença é de apenas 0,03. Se, por outro lado, considerarmos os valores de *kappa* para a Narrativa 4, caso que se assemelha aos observados no estudo piloto, notaremos uma diferença de valores um pouco maior: na C2, o nível de concordância entre examinadores aumenta em 0,07 pontos, em se comparando com o valor obtido na C1.

A figura 3 traz os resultados de concordância entre examinadores para cada narrativa na Condição 3 (C3), quando tiveram acesso apenas ao áudio das narrativas.

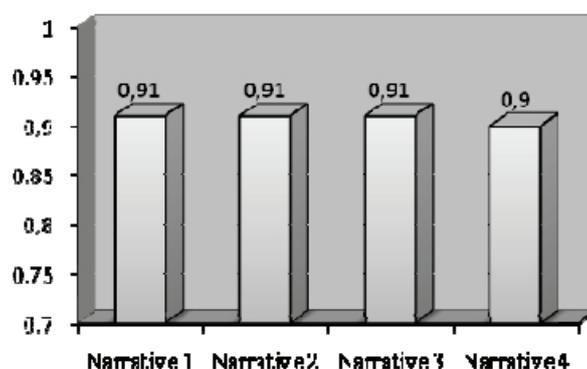


Figura 3

Valores de kappa (K) para cada uma das narrativas na Condição 3 (C3)

Os resultados na figura 3 são muito homogêneos e mostram que os examinadores concordam de maneira bastante significativa acerca da segmentação das narrativas quando expostos apenas à versão em áudio das mesmas.

A figura 4 traz os resultados de concordância entre examinadores para cada uma das narrativas na Condição 4 (C4), quando tiveram acesso apenas à versão deslexicalizada do áudio original das narrativas.

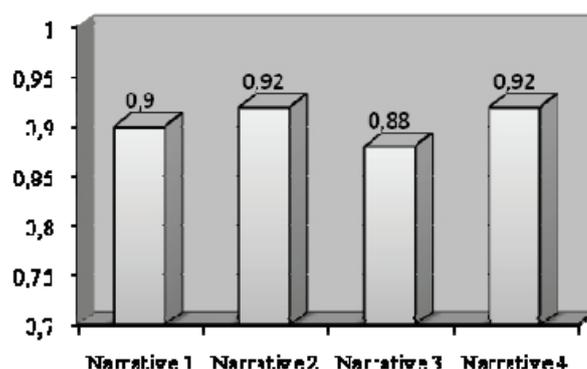


Figura 4

Valores de kappa (K) para cada uma das narrativas na Condição 4 (C4)

Quando expostos a uma versão deslexicalizada do áudio de uma narrativa, os examinadores concordam de maneira bastante significativa acerca de sua segmentação. É importante observar que o maior valor de *kappa* neste estudo (0,93), foi obtido na C4.

Por fim, a Figura 5 abaixo traz valores médios de *kappa* nas diferentes condições experimentais em que as narrativas foram apresentadas aos examinadores:

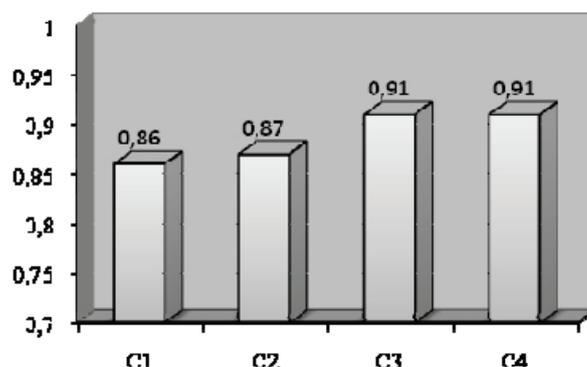


Figura 5

Valores médios de kappa (K) em cada uma das quatro condições que os estímulos foram apresentados aos examinadores

Observe-se que os maiores valores de *kappa* estão associados a C3 e C4: as análises que foram feitas baseadas exclusivamente nas versões em áudio das narrativas.

Discussão

O que os resultados apresentados acima indicam é que pessoas conseguem identificar de forma bastante homogênea fronteiras discursivas em narrativas espontâneas, mesmo não sendo especialistas na área, nem tido recebido qualquer treinamento prévio para fazê-lo. Silva & Oliveira (2011) já haviam demonstrado isso com um *corpus* em que o conteúdo informacional das narrativas foi disponibilizado aos participantes do experimento. O presente estudo dá um passo além, mostrando que mesmo quando as informações lexicais, sintáticas e semânticas da narrativa são bloqueadas, as pessoas ainda conseguem identificar a sua estrutura subjacente. Isso indica claramente que as pessoas fazem uso substancial da informação prosódica na tarefa de identificar a estrutura de narrativas espontâneas.

Parece ser evidente, no entanto, que a prosódia não é o único recurso utilizado na identificação da estrutura subjacente de narrativas. Os resultados de concordância entre examinadores para narrativas na C1, que foram apresentadas apenas em sua versão transcrita, demonstram claramente que o acesso ao conteúdo lexical e semântico das narrativas é suficiente para que examinadores concordem na organização estrutural de narrativas espontâneas. A prosódia, nesse caso, é utilizada como uma informação adicional, que tornaria a concordância entre examinadores acerca da estrutura discursiva ainda mais robusta⁹. Vários estudos (Collier, 1993; de Pijper & Sanderman, 1994; Swerts, 1996; Swerts, Collier & Terken, 1994), incluindo um estudo piloto que utilizou o mesmo *corpus* do presente estudo (Silva & Oliveira, 2011), já demonstraram isso antes. Embora os resultados em dois casos particulares neste estudo (concordância das narrativas 1 e 2 na C2, em comparação com a C1) aparentemente contradigam esse achado, os resultados gerais apresentados aqui, quando considerados em bloco (cf. Figura 5), ainda mostram uma tendência em direção às conclusões reportadas pela literatura.

Os resultados reportados aqui evidenciam ainda que a informação prosódica é, por si só, suficiente para que se identifique a estrutura de narrativas espontâneas – e este é talvez o achado mais importante do presente estudo. Quando têm acesso apenas à versão deslexicalizada das narrativas do *corpus* (C4), os participantes ainda assim concordam em um grau estatisticamente significativo acerca da segmentação desses textos, o que sugere ter a prosódia papel relevante na percepção da estrutura

9. Bader (1998) e Fodor (2002) argumentam que na leitura silenciosa, um contorno entoacional padrão é projetado no estímulo, o que pode influenciar a resolução da ambiguidade sintática. Denominam esse fenômeno *Prosódia Implícita*. Embora a hipótese de prosódia implícita tenha sido estudada apenas no nível da frase, não será inapropriado especular que o fenômeno também se verifique em unidades linguísticas maiores que a frase.

discursiva. Trabalhos anteriores já demonstraram que a percepção da estrutura discursiva não depende exclusivamente de informações lexicais, sintáticas e semânticas (de Pijper & Sanderman, 1994; Hirschberg & Grosz, 1992; Kreiman, 1982; Swerts, 1996; Swerts, Collier, & Terken, 1994). A maior parte desses estudos, no entanto, fez uso de material lido, elicitado, não espontâneo. Além disso, a maioria trata de línguas como o inglês ou o holandês. A relevância do presente estudo está também no fato de apresentar resultados a partir da utilização de um *corpus* de dados de fala espontânea do português brasileiro. Embora já exista uma quantidade razoável de trabalhos que abordem diferentes aspectos prosódicos do Português Brasileiro, incluindo as funções da prosódia no discurso, estudos sobre prosódia como um dispositivo de estruturação da fala espontânea no português falado no Brasil ainda são muito escassos. O presente trabalho constitui, desse modo, uma contribuição para esta área de pesquisa.

Considerações Finais

O presente estudo teve por objetivo avaliar em que medida a estrutura de narrativas espontâneas orais é reconhecida por pessoas não treinadas. Além disso, procurou investigar qual o peso que a prosódia tem nesse processo de reconhecimento. Para isso, um protocolo experimental foi utilizado, com o propósito de testar se há concordância significativa entre examinadores não treinados acerca da percepção de localização de fronteiras discursivas, a partir de um modelo baseado em intenções (Passonneau & Litman, 1997).

Os resultados apresentados aqui corroboram o que já se tinha observado na literatura para outras línguas e com corpora diferenciados: a estrutura subjacente a um dado gênero discursivo oral pode ser identificada com um alto grau de confiabilidade, mesmo quando, preservadas as características prosódicas do texto, informações lexicais, sintáticas e semânticas não são disponibilizadas. Isso sugere que a prosódia, além de atuar significativamente na estruturação do discurso, sob a perspectiva da produção (Geluykens & Swerts, 1994; Grosz & Hirschberg, 1992; Passonneau & Litman, 1993), tem um relevante papel também para a percepção da estrutura discursiva.

O próximo passo do presente estudo é analisar que elementos prosódicos especificamente estão associados a pontos fortes de fronteiras perceptuais (PBS), nas diferentes condições experimentais discutidas aqui. Oliveira Jr. (2000), em um estudo de produção, observou que as fronteiras entre unidades maiores das narrativas (ou fronteiras narrativas) são prosodicamente diferentes. Especificamente, verificou-se que (i) pausas ocorrem mais sistematicamente no final das seções narrativas e são frequentemente mais longas que em outras posições; (ii) a diferença de tom entre duas unidades entoacionais é maior quando essa unidade coincide com uma fronteira de seção narrativa; e (iii) tons baixos ocorrem em fronteiras de seções narrativas; tons altos ocorrem dentro de seções narrativas.

É sabido que qualquer estudo envolvendo aspectos prosódicos da fala deve considerar não apenas a parte da produção, mas sobretudo a percepção de tal fenômeno. A relevância das variáveis prosódicas na demarcação da estrutura discursiva só pode ser plenamente validada depois da consideração de sua eficácia sob a perspectiva da percepção. O presente estudo, portanto, consiste em uma contribuição nesse sentido.

Referências Bibliográficas

- BADER, M. Prosodic influences on reading syntactically ambiguous sentences. In: FODOR, J. & FERREIRA, F. (eds). *Reanalyses in sentence processing*. Dordrech: Kluwer Academic, p.1-46, 1998.
- BROWN, G., CURRIE, K. & KENWORTHY, J. *Questions of Intonation*. London: Croom Helm, 1980.
- BRUGMAN, H.; RUSSEL, A. Annotating Multimedia/ Multi-modal resources with ELAN. In: PROCEEDINGS OF THE FOURTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2004, Lisboa, Portugal. 2004. p. 2065-2068.
- CAMPBELL, T.; WINKLER, I.; KUJALA, T. N1 and the mismatch negativity are spatiotemporally distinct ERP components: disruption of immediate memory by auditory distraction can be related to N1. *Psychophysiology*, v. 44, n. 4, p. 530-540, 2007.
- CARLETTA, J. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, v. 22, n. 2, p. 249-254, 1996.
- COHEN, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37-46, 1960.
- COLLIER, R. On the communicative function of prosody: some experiments. *IPO Annual Progress Report*, v. 28, p. 67-75, 1993.
- COLLIER, R.; DE PIJPER, J. R.; SANDERMAN, A. Perceived prosodic boundaries and their phonetic correlates. In: PROCEEDINGS OF THE ARPA WORKSHOP ON HUMAN LANGUAGE TECHNOLOGY, 1993, Plainsboro, New Jersey, USA. Plainsboro, New Jersey, USA: Morgan Kaufman Publishers, 1993. p. 341-345.
- DE PIJPER, J. R.; SANDERMAN, A. A. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *J. Acoust. Soc. Am*, v. 96, n. 4, p. 2037-2047, 1994.
- DI EUGENIO, B. On the usage of Kappa to evaluate agreement on coding tasks. In: PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2000, Atenas, 2000, p. 441-444.

- FODOR, J. Prosodic disambiguation in silent reading. In: HIROTANI, M. (ed.). *Proceedings of Nels*, 32. Amhert: GLSA, 2002.
- FUJISAKI, H. *Prosody, models, and spontaneous speech*. In Y. Sagisaka, N. Campbell & N. Higuchi. *Computing Prosody: Computational Models for Processing Spontaneous Speech*. New York: Springer, 1997. p.27-42.
- GELUYKENS, R.; SWERTS, M. Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, v. 15, p. 69–77, 1994
- GROSZ, B.; HIRSCHBERG, J. Some intonational characteristics of discourse structure. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 1992, Banff. Banff, 1992. p. 429–432.
- GROSZ, B.; HIRSCHBERG, J.; NAKATANI, C. A Study of intonation and discourse structure in directions. In: PROCEEDINGS OF THE AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE WORKSHOP ON THE INTEGRATION OF NATURAL LANGUAGE AND SPEECH PROCESSING, 1994, Seattle. Seattle, 1994.
- GROSZ, B.; SIDNER, C. L. Attentions, intentions, and the structure of discourse. *Computational Linguistic*, v. 85, p. 363–394, 1986.
- HIRSCHBERG, J.; GROSZ, B. Intonation features of local and global discourse structure. In: PROCEEDINGS OF THE DARPA WORKSHOP ON SPOKEN LANGUAGE SYSTEMS, 1992, Arden House. Arden House, 1992.
- HIRSCHBERG, J.; NAKATANI, C. H.; GROSZ, B. J. Conveying discourse structure through intonation variation. In: PROCEEDINGS OF THE ESCA WORKSHOP ON SPOKEN DIALOGUE SYSTEMS: THEORIES AND APPLICATIONS, 1995, Visgo, Denmark. Visgo, Denmark: ESCA, 1995.
- KREIMAN, J. Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, v. 10, p. 163–175, 1982.
- LAAN, G. P. M. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a reading speaking style. *Speech Communication*, v. 22, p. 43–65, 1997.
- LABOV, W. The transformation of experience in narrative syntax. In: _____. *Language in the inner City*. Philadelphia: University of Pennsylvania Press, 1972. p.354-98.
- LABOV, W; WALETZKY, J. Narrative analysis: oral versions of personal experience. In: HELMS, J. *Essays on the verbal and visual arts*. Seattle: University of Washington Press, 1967.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159–174, 1977.
- LEVIN, H.; SCHAFFER, C.; SNOW, C. The Prosodic and Paralinguistic Features of Reading and Tell-

ing Stories. *Language and Speech*, v. 25, n. 1, p. 43–54, 1982.

LITMAN, D. J.; PASSONNEAU, R. J. Combining Multiple Knowledge Sources for Discourse Segmentation. In: PROCEEDINGS OF THE 33RD ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL-95), 1995, Cambridge, MA. Cambridge, MA, 1995. p. 108–115.

LITMAN, D. J.; PASSONNEAU, R. J. Empirical evidence for intention-based discourse segmentation. In: PROCEEDINGS OF THE ACL WORKSHOP ON INTENTIONALITY AND STRUCTURE IN DISCOURSE RELATIONS, 1993, Columbus, Ohio. Columbus, Ohio, 1993. p. 60–63

OLIVEIRA Jr, M., *Prosodic Features in Spontaneous Narratives*. 2000. 291 fl. Tese (Doutorado em Linguística) - Simon Fraser University, Vancouver, 2000.

PASSONNEAU, R. J.; LITMAN, D. J. Discourse Segmentation by Human and Automated Means. *Computational Linguistics*, v. 23, n. 1, p. 103–139, 1997.

PASSONNEAU, R. J.; LITMAN, D. J. Intention-based segmentation: Human reliability and correlation with linguistic cues. In: PROCEEDINGS OF THE 31ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL-93), 1993, Columbus, Ohio. Columbus, Ohio, 1993. p. 148–155.

SAGISAKA, Y.; CAMPBELL, N.; HIGUCHI, N. *Computing prosody*. New York: Springer Verlag, 1997.

SILVA, E. W.; OLIVEIRA, M., Jr. A percepção dos elementos prosódicos como marca de estruturação de narrativas espontâneas. In: RESUMOS DO III COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA, 2011, Belo Horizonte. Belo Horizonte, 2011. p. 15–17.

SWERTS, M. Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, v. 101, n. 1, p. 514–521, 1996.

SWERTS, M.; BOUWHUIS, D. G.; COLLIER, R. Melodic cues to the perceived “finality” of utterances. *J. Acoust. Soc. Am*, v. 96, n. 4, p. 2064–2075, 1994.

SWERTS, M.; COLLIER, R.; TERKEN, J. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, v. 15, n. 1-2, p. 79–90, 1994.

SWERTS, M.; GELUYKENS, R. Prosody as a marker of information flow in spoken discourse. *Language and Speech*, v. 37, n. 1, p. 21–43, 1994.

WOLFSON, N. Speech events and natural speech. *Language in Society*, Cambridge, n.5, p.189-209, 1979.

YULE, G. Speakers’ topics and major paratones. *Lingua*, v. 52, p. 33–47, 1980.