

Análise Acústica Experimental em Diálogos

Thalita Siqueira do Valle Barbosa
Departamento de Linguística
Universidade Estadual de Campinas
Campinas, Brasil
thalitasvb@gmail.com

Abstract—O objetivo principal deste trabalho é avaliar o grau de simultaneidade de fala entre os participantes dos diálogos selecionados previamente, além de mostrar o funcionamento dinâmico dessas falas simultâneas. Os diálogos também são analisados a partir de parâmetros prosódico-acústicos como a frequência fundamental.

Fonética Acústica; Prosódia; Diálogo; Acústica do diálogo

I. INTRODUÇÃO

Segundo a literatura que envolve a Análise da Conversação, que trata do estudo da interação verbal e não verbal em situações cotidianas, e procura descrever a forma de interações formais e informais. A conversação, segundo Levinson (1983, p.284) é o gênero básico da interação humana, e é um produto organizado e que pode ser estudado. Segundo Marcuschi, esta organização se deve a “um reflexo de um processo subjacente desenvolvido, percebido e utilizado pelos participantes da atividade comunicativa, ou seja, as decisões interpretativas dos interlocutores decorrem de informações contextuais e semânticas mutuamente construídas ou inferidas de pressupostos cognitivos, étnicos e culturais, entre outros.”

Ainda segundo Marcuschi (1986) é justamente pelo caráter dialógico da linguagem que buscamos entender estas interações e suas organizações. É importante notar que definimos como conversação tudo o que é caracterizado por pelo menos dois falantes e uma troca de turno.

Notamos que nessa literatura há um pressuposto de que a conversação é feita em turnos: enquanto uma pessoa fala a outra ficaria em silêncio esperando por sua vez, assim como Levinson define a fala conversacional como “um tipo de fala em que dois, ou mais participantes, livremente se alternam na fala”. Teríamos um senso de ritmo conversacional, mas não é isso que é visto nas análises recentes, como no trabalho de Campbell, que, a partir de “um amplo número de conversas por telefone em que pessoas eram pagas ‘apenas para falar’ uma com a outra” conseguiu abrir caminho para uma nova visão. Ao demonstrar que, ao contrário do que havia sido postulado, a fala conversacional não é feita como questão e resposta ou tópico e comentário, na verdade há um padrão muito mais complexo de fala simultânea em que os participantes tentam dominar a interação. Não há barreiras entre um turno e outro. Os participantes normalmente não esperam em silêncio enquanto o outro está num momento de fala, na verdade há alguma reação e quando o tópico já é dominado pelo ouvinte do momento, podendo ocorrer falas

simultâneas. Há uma constante troca de falas (‘speech’) e gestos. Portanto, é possível afirmar que nas conversações é muito comum haver turnos em que haja mais de um falante, por mais breve que isso aconteça, pois não há nenhuma ordem pré-definida dos turnos.

Um modelo muito importante e respeitado de tomada de turno é o de Sacks *et al.* (1974), que diz que as falas simultâneas são descritas como ‘comuns, mas breves’, seriam elas breves por se posicionarem essencialmente em finais de turno, ou em torno dos chamados *transition relevance place* (TRP), que seria onde o falante em voga pode terminar o seu turno. De acordo com este modelo as sobreposições seriam um resultado de uma ‘auto-escolha’ e de uma projeção do fim do turno, ou seja, a ‘auto-escolha’ seria o caso em que o falante atual não escolhe o próximo participante, o que abre a possibilidade de algum dos demais participantes de escolher e continuar o diálogo, ou por outro lado um dos outros participantes antes mesmo do término do turno do falante corrente sobrepe seu turno e continua o diálogo exatamente no ponto em que aconteceria a conclusão do turno do outro falante, ou seja a projeção deste fim de turno.

Em trabalhos mais recentes como o de Schegloff (2000) que trata os *overlaps* de turnos competitivos como sendo lugares em que os participantes qualificam essa sobreposição na fala como problemática e com necessidade de resolução. Também é importante ressaltar a diferença entre o chamado *backchannel* e o *overlap*. Em nosso trabalho foram avaliadas apenas as falas simultâneas e não os backchannels, que são caracterizados na literatura como mensagens verbais e não-verbais mandadas por um interlocutor no papel de ouvinte na conversação com o intuito de informar que está seguindo a fala do outro. As mensagens verbais consistem de, por exemplo, ‘uh-huh’, ‘mm-hm’, e ‘sim’, enquanto que as mensagens não-verbais são caracterizadas por sorrisos ou um balançar de cabeça. Na literatura, por expressar que o ouvinte está compreendendo, ou aceitando, ou não a conversa, essas mensagens receberam diversos nomes como ‘signals of attention’ (Fries, 1952), ‘accompaniment signals’ (Kendon, 1967), ‘listeners responses’ (Dittman and Llewellyn, 1968) e ‘backchannel’ (Yngve, 1970). Essas mensagens são muito utilizadas em trabalhos dentro da Linguística que envolvam pesquisa no campo de língua e cultura, o que não é o nosso caso, portanto iremos excluir este tipo de fala dos nossos dados, mesmo porque não têm a finalidade de interromper o interlocutor.

II. MATERIAL E MÉTODOS

O material utilizado foram gravações de diálogos do programa “Diálogos Impertinentes” produzido pela TV PUC, com apoio do jornal Folha de S. Paulo e Sesc São Paulo, exibido pelo SESCTV, que foram retirados de um projeto desenvolvido no LIIAC (Laboratório Integrado de Análise Acústica e Cognição) da Pontifícia Universidade Católica de São Paulo (PUC-SP) disponível no site <http://www.pucsp.br/liaac/dialogos.html>, e fazem parte de um projeto do Grupo de Pesquisa de Estudos sobre a Fala, que tem como líder a Profª. Dra. Sandra Madureira. O corpus dessa análise é composto por cinco diálogos, totalizando 295 minutos, tendo como nomes os temas principais dos diálogos, sendo: ‘A Moda’, ‘A Arte’, ‘A Felicidade’, ‘A Sabedoria’ e ‘A Alimentação’.

Os métodos incluem o software livre chamado Praat (<http://www.fon.hum.uva.nl/praat/>), em que foi possível fazer as marcações de sobreposição de falas. A intenção foi mostrar onde se localizam as falas de cada um dos participantes do diálogo, normalmente quatro participantes faziam parte de cada um dos diálogos, e mostrar onde se encontravam as falas simultâneas.

Foi utilizado ainda neste trabalho um script criado por Barbosa, chamado de *Conversation Analysis*, que roda no programa Praat, a fim de contar o tempo de turno de cada participante em relação aos demais, Turn Share (TS), cf. Kousidis, Dorrán, McDonnell e Coyle, 2009, bem como os tempos de silêncio e de fala sobreposta.

Com esse programa, é possível marcar o tempo ativo de cada participante i no diálogo, $AT_i = L_i/L$, em que L_i é a duração total da fala do participante i e L o tempo do diálogo. O tempo total percentual de turno de cada participante é dado por: $TS_i = AT_i / \text{Soma}(AT_i)$, sendo TS_i a proporção de turno do participante i (*turn share*) e AT_i o tempo ativo do falante i . Essas equações provêm do artigo de Kousidis, Dorrán, McDonnell e Coyle (2009).

Portanto, a partir da separação das falas dos participantes feita no programa Praat e do script, foram feitas contagens do tempo de turno de cada participante em relação aos demais, Turn Share (TS), bem como os tempos de silêncio e de fala sobreposta, qual a interferência delas no diálogo, além de observarmos se há marcas acústicas que possam mostrar a natureza dessas interações.

III. RESULTADOS

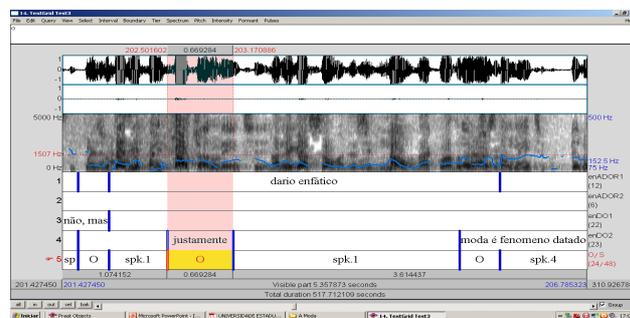
A partir dos diálogos selecionados, conseguiu-se contar o tempo de fala ativa e o Turn Share de cada diálogo. Foram, também, feitas tabelas para cada um dos diálogos referidos, contendo, na primeira coluna, qual parte do diálogo estamos tratando. Cada uma das entrevistas são separadas em várias partes para melhor manuseamento no Praat. A segunda coluna se refere às sobreposições que aconteceram neste diálogo, com a etiqueta ‘spk.X1/O/spk.X2’, sendo que ‘spk.X1’ é o falante corrente em dada parte do discurso, ‘O’ significando sobreposição, e ‘spk.X2’ sendo o falante que causou a sobreposição. A terceira coluna é a duração da

sobreposição em milissegundos. E na quarta, e última coluna, foi assinalado se, além de ocorrer a sobreposição, o falante roubou o turno. Para exemplificar, segue a etiquetagem e contagem do diálogo “A Felicidade”.

Diálogo	Etiqueta	Duração da sobreposição(ms)	Roubo de Turno
A Felicidade 1	spk.1/O/spk.3	486	x
A Felicidade 2	spk.3/O/spk.1	693	
A Felicidade 2	spk.4/O/spk.1	358	x
A Felicidade 3	spk.1/O/spk.4	483	x
A Felicidade 3	spk.1/O/spk.4	565	x
A Felicidade 4	spk.3/O/spk.4	275	x
A Felicidade 4	spk.3/O/spk.1	293	x
A Felicidade 5	spk.1/O/spk.4	1561	x

A partir desta tabela, observa-se que, dentre o total de overlaps neste diálogo, temos uma porcentagem de 40% de overlaps com roubo de turno. Além disso, o maior número de falas simultâneas com roubo de turno é feita pelo falante 4.

Também foram separadas algumas imagens e explicações acerca de alguns casos interessantes encontrados durante a pesquisa, como no exemplo abaixo:



Neste caso é interessante notar que mesmo com as sobreposições de fala ocorridas, o diálogo é entendido por eles e continuado.

Como exemplo dos resultados do script *Conversation Analysis* para contar o *Turn Share*, temos o do diálogo “A Sabedoria”:

“Parte 1:
 $AT_1 = 0.57$, $AT_2 = 0$, $AT_3 = 0.31$, $AT_4 = 0.11$
 $TS_1 = 0.57$, $TS_2 = 0$, $TS_3 = 0.31$, $TS_4 = 0.11$
 $TO = 0$, $TSIL = 0.02$ ”

Parte 2:

AT1 = 0.11, AT2 = 0.67, AT3 = 0, AT4 = 0.23

TS1 = 0.11, TS2 = 0.66, TS3 = 0, TS4 = 0.23

TO = 0, TSIL = 0

Parte 3:

AT1 = 0, AT2 = 0.14, AT3 = 0.39, AT4 = 0.48

TS1 = 0, TS2 = 0.14, TS3 = 0.39, TS4 = 0.47

TO = 0.03, TSIL = 0.004

Parte 4:

AT1 = 0.009, AT2 = 0.17, AT3 = 0.25, AT4 = 0.58

TS1 = 0.009, TS2 = 0.16, TS3 = 0.25, TS4 = 0.58

TO = 0.003, TSIL = 0

Parte 5:

AT1 = 0.02, AT2 = 0.29, AT3 = 0.28, AT4 = 0.39

9

TS1 = 0.02, TS2 = 0.30, TS3 = 0.29, TS4 = 0.40

TO = 0.02, TSIL = 0

Parte 6:

AT1 = 0.03, AT2 = 0.02, AT3 = 0.21, AT4 = 0.78

TS1 = 0.03, TS2 = 0.02, TS3 = 0.21, TS4 = 0.75

TO = 0.04, TSIL = 0.003

Parte 7:

AT1 = 0.03, AT2 = 0.11, AT3 = 0.66, AT4 = 0.21

TS1 = 0.02, TS2 = 0.11, TS3 = 0.66, TS4 = 0.21

TO = 0.006, TSIL = 0

Parte 8:

AT1 = 0.02, AT2 = 0.02, AT3 = 0.47, AT4 = 0.50

TS1 = 0.02, TS2 = 0.02, TS3 = 0.47, TS4 = 0.50

TO = 0.02, TSIL = 0.007

Parte 9:

AT1 = 0.06, AT2 = 0.07, AT3 = 0.26, AT4 = 0.62

TS1 = 0.06, TS2 = 0.07, TS3 = 0.26, TS4 = 0.61

TO = 0.02, TSIL = 0.02

Parte 10:

AT1 = 0.05, AT2 = 0.07, AT3 = 0.47, AT4 = 0.43

TS1 = 0.05, TS2 = 0.07, TS3 = 0.46, TS4 = 0.42

TO = 0.01, TSIL = 0.01

Através destes dados podemos perceber que as taxas de maior sobreposição ocorrem nas partes 3, 4 e 6, e coincidentemente o falante 4 apresenta o maior tempo ativo, e que a partir da comparação com o 10 *Turn Share* é um pouco menor, o que significa que está o seu turno 'roubado', ou seja, sobrepe-se a sua fala. É muito importante reparar que nem sempre o tempo ativo é igual ao *turn share*, isto apenas reafirma que os falantes não falam em turno, que "normalmente é definido como parte da fala que pertence ao falante e dura até o outro falante dar a vez"⁶, mas as vezes 'atropelam' a fala uns dos outros. Interessante notar que o tempo de sobreposição é um bom indicador de 'alta ativação'. Portanto, um aplicação direta desses dois quantificadores (TS e TO) pode indicar uma ativação de contexto emocional na fala neste diálogo, por ser espontâneo.

IV. CONCLUSÕES

A partir dos dados analisados durante esta pesquisa foi possível concluir, primeiramente, que em um diálogo

espontâneo as falas de seus participantes não são feitas em turnos, justamente por haverem várias interrupções no meio das falas com a finalidade de manter o canal em andamento (backchannel) ou por haverem roubos de turno, com trechos de fala simultânea.

Pensando nas sobreposições de fala, podemos dizer que alguns dos nossos dados demonstram que, ao contrário do que se pensava no início do projeto, eles não servem apenas para roubar o turno do falante corrente, mas também para demonstrar opiniões no meio da fala do outro.

Uma importante consideração é a de que no momento dessas falas simultâneas ocorrem exemplos em que dois participantes estão falando ao mesmo tempo e se entendem, não ocorrendo o roubo do turno, como se o falante que havia se sobreposto apenas ajudasse o falante corrente. Também temos que o sujeito que vai cometer a sobreposição tem duas saídas no processo dialógico, a primeira seria elevar o pitch a fim de que o outro pare de falar e o deixe tomar o turno, nem que seja momentaneamente. Ou ele continuar no mesmo pitch do falante corrente e ir se sobrepondo, o que é interessante perceber é que ele continua a conversa e os participantes se entendem, sem nenhum problema para continuar seu diálogo, por isso podemos dizer que os falantes se usam da sobreposição de falas como uma saída comum no diálogo, e não falam em turnos como se acreditava. Com o script criado a partir das equações de Kousidis (2009), temos que para sobreposição o intervalo de tempo antes da fala simultânea é considerado como um indicador de quem irá cometer a sobreposição. Por exemplo, se é um intervalo do falante.1, então a sobreposição é atribuído ao falante 2. O valor da sobreposição de cada falante é definido como o número de ocorrências de sobreposição na organização do diálogo, pelo número de segmentos de falas do falante corrente. Essencialmente, isto descreve quantas vezes o falante começa falando antes do interlocutor terminar. No entanto, é importante ressaltar que esta pesquisa ainda pode ser aprofundada com mais sujeitos (dados). Foram estudadas algumas questões teóricas de base, e analisado e discutido um corpus que trouxesse indícios de padrões e que pudesse possibilitar a elaboração de hipóteses pertinentes ao contexto teórico apresentado. Devo ainda informar que pretendo fazer uma análise mais aprofundada das questões que envolvem este trabalho, e ainda ir além dele, em minha próxima pesquisa acadêmica.

REFERÊNCIAS

- [1] MARCUSCHI, L. A., "Análise da conversação". São Paulo: Ática, 1986.
- [2] LEVINSON, S.C., "Pragmatics", Cambridge University Press, 1983.
- [3] CAMPBELL, N., "Approaches to conversational speech rhythm: speech activity in two-person telephone dialogues", 2007.
- [4] SACKS, H; SCHEGLOFF, E.E.; JEFFERSON, G. "A simplest systematics for the organization of turn taking for conversation." Language, 50, 696-735, 1974.
- [5] Kousidis, S.; Dorrán, D.; McDonnell, C. & Coyle, E., Towards Flexible "Representations for Analysis of Accommodation of Temporal Features in Spontaneous Dialogue Speech", 2009.

