

As principais medidas da fala

Tommaso Raso

Maryualê M. Mittmann

1. Objetivos do capítulo

Neste capítulo apresentamos uma primeira análise de alguns indicadores da estruturação da fala informal do português brasileiro (PB) representado no *corpus* C-ORAL-BRASIL. A partir das unidades naturais e de referência da fala espontânea adotadas para a compilação do C-ORAL-BRASIL¹ (a saber, os turnos dialógicos e os enunciados), mostramos como estas variam em tamanho e complexidade de acordo com a ramificação *corpus*. Além dos dados do próprio C-ORAL-BRASIL, incluímos ainda dados de outros *corpora* para uma análise comparativa.

Esta análise enfoca duas dimensões principais: a primeira concerne à variação linguística resultante do meio físico-ambiental e do canal através do qual a língua é usada (variação diamésica), esfera que opõe o meio oral ao meio escrito (Berruto, 1993), enquanto a segunda dimensão comparada trata da variação interlinguística da fala espontânea informal.

No que se refere à variação diamésica, nosso objetivo é mostrar as especificidades da fala informal em relação à escrita, tomando como ponto de comparação principal as estratégias lexicais características de uma e de outra. Para esta comparação utilizamos principalmente os dados do *corpus*

¹ Moneglia (1999, 2011) apresenta de forma clara os argumentos em favor do enunciado como unidade de referência da fala e sua relação com compilação de *corpora* orais.

NILC/São Carlos (Pinheiro; Aluísio, 2003). Trata-se de um *corpus* de escrita de acesso livre, disponibilizado pelo projeto AC/DC² (Santos; Sarmiento, 2002) da Linguateca³ (Santos, 2009).

A comparação interlinguística foi realizada com base nas seções informais dos *corpora* de português europeu (PE), italiano, espanhol e francês do projeto C-ORAL-ROM (Cresti; Moneglia, 2005). Com isso, pretendemos incluir o PB na comparação das línguas românicas representadas no C-ORAL-ROM realizada por Cresti (2005b). Isso só é possível porque os *corpora* dos projetos brasileiro e europeu têm a mesma estrutura e foram compilados segundo os mesmos critérios fundamentais (veja-se o Capítulo 2).

Apresentamos ainda algumas medidas que permitem observar certos aspectos da organização da fala, são eles:

1. a organização através da segmentação prosódica, que pode estruturar o enunciado em uma (enunciado simples) ou mais unidades (enunciado complexo) (veja-se o Capítulo 3 para uma explicação mais aprofundada e o Capítulo 4 para os critérios de implementação e validação);
2. a ocorrência de estruturas sintáticas complexas, aferidas pela presença de um verbo finito no enunciado, e pelo controle de quando este verbo constitui um núcleo verbal no enunciado;
3. a combinação das duas estratégias anteriores, visto que há a possibilidade de termos enunciados simples com e sem núcleo verbal e enunciados complexos com e sem núcleo verbal;
4. a ocorrência de conjunções coordenativas (*e* e *mas*) e subordinativas (*que* e *porque*), e sua posição dentro do enunciado.

2. Uso de verbos e nomes na fala espontânea e na escrita

Na língua falada espontânea, a unidade de referência é o enunciado, identificado com base em critérios prosódicos e pragmáticos.⁴ Por outro

² O AC/DC é um projeto que objetiva melhorar o acesso à *corpora* em língua portuguesa, tornando esses recursos acessíveis na rede e disponibilizando uma interface de consulta <<http://www.linguateca.pt/ACDC/>>.

³ A Linguateca é um centro de recursos linguísticos para o processamento da língua portuguesa, financiado pelo governo português que pode ser acessado em <www.linguateca.pt>.

⁴ Um enunciado é a menor unidade linguística autônoma do ponto de vista pragmático e é delimitado no fluxo da fala por fronteiras prosódicas de tipo terminal (veja-se o Capítulo 3 deste volume).

lado, a unidade de referência da escrita é a sentença.⁵ Percebe-se aqui uma distinção fundamental no que se refere ao objeto considerado em cada diamesia. Enquanto na fala a língua estrutura-se em unidades pragmáticas, delimitadas pela entonação, na escrita a língua organiza-se em unidades sintáticas, delimitadas pela pontuação.

Derivam daí estratégias muito distintas de estruturação linguística, originando diferenças que apenas começamos a investigar em maior detalhe. Conforme indica Cresti (2005b), a organização no nível lexical da língua falada é diferente daquela da língua escrita. Na fala de modo geral, a proporção de verbos é maior do que a de nomes, uma tendência contrária à da escrita, na qual nomes são mais frequentes do que verbos.

Apresentamos aqui uma primeira análise, na qual observamos as particularidades quanto ao uso de verbos e nomes, na fala e na escrita. A título de ilustração, vejam-se os resultados obtidos a partir do *Corpus* do Português⁶ (Davies; Ferreira, 2006), no que se refere à distribuição de nomes e verbos na fala e na escrita (Tabela 1). Este é um *corpus* do português que conta com cerca de 45 milhões de palavras. Trata-se de um *corpus* balanceado, porém nele estão incluídos os dados do PE em conjunto com o PB, e dados diacrônicos (a partir de 1.300) juntamente com dados sincrônicos de ambos os países.

Tabela 1
Frequência de verbos e nomes nas seções escrita e falada do *Corpus* do Português

Seção do <i>corpus</i>	Escrita/ Acadêmico		Escrita/ Jornalístico		Escrita/ Ficção		Fala	
Tamanho (palavras)	5.754.506		6.488.917		5.937.484		2.083.296	
Frequência de verbos	694.469	12%	875.016	13%	1.107.567	19%	409.896	20%
Frequência de nomes	1.510.299	26%	1.555.535	24%	1.323.628	22%	406.059	19%
Total de verbos e nomes	2.204.768	38%	2.430.551	37%	2.431.195	41%	815.955	39%

Fonte: Davies; Ferreira, 2006.

⁵ Para uma discussão sobre os conceitos de enunciado e sentença, veja-se Cresti (2001).

⁶ O *Corpus* do Português pode ser acessado através do site <<http://www.corpusdoportugues.org/>>.

Os dados extraídos do *Corpus* do Português mostram parcialmente a tendência indicada por Cresti (2005b) e já observada por outros pesquisadores para a língua inglesa (Biber *et al.*, 1999). Na escrita dos gêneros acadêmico e jornalístico, percebe-se uma concentração muito maior de nomes do que de verbos; contudo, na seção do gênero ficção nota-se uma distribuição mais equitativa entre nomes e verbos.⁷ De modo geral, os dados de escrita apresentam cerca de 15% de verbos e 24% de nomes (somando-se todas as seções escritas), enquanto a fala conta com um certo equilíbrio entre verbos (20%) e nomes (19%).

Os dados acima também corroboram os resultados levantados por Bacelar do Nascimento, Gonçalves e Veloso (2005, p. 205) para o PE. A língua falada espontânea (*corpus* C-ORAL-ROM do PE) apresenta uma proporção de 15% de nomes e 18% de verbos, enquanto o PE escrito do gênero jornalístico (*corpus* RL-News) apresenta 24% de nomes e 12% de verbos.

2.1 Ocorrência de nomes e verbos na fala e na escrita do PB

Para comparação com a escrita, utilizamos aqui o *corpus* NILC/São Carlos. Este *corpus* apresenta a desvantagem de não ser balanceado quanto aos gêneros textuais que o compõem. O gênero jornalístico apenas representa mais de 79% das palavras, e, além disso, a totalidade dos textos deste gênero é proveniente de um único veículo de comunicação (ver Pinheiro; Aluísio, 2003). Ainda assim, o *corpus* NILC/São Carlos é o único *corpus* exclusivo do PB escrito, cujos dados são inteiramente disponibilizados para consultas.⁸ Veja-se na Tabela 2 a distribuição de palavras por seção do *corpus*.

⁷ Biber *et al.* (1999) apontam que tanto a prosa acadêmica quanto a ficcional em língua inglesa apresentam cerca de três nomes para cada verbo, uma tendência muito diferente daquela observada em português, o que pode ser o reflexo de tradições literárias e acadêmicas distintas.

⁸ Consultas ao *corpus* NILC/São Carlos, bem como acesso às listas de palavras e tabelas contendo as frequências de unidades do *corpus*, encontram-se na página do projeto AC/DC: <<http://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>>.

Tabela 2
Número de palavras por seção do *corpus* NILC/São Carlos

Seção	Sigla	Total de palavras	% total de palavras	% verbos	% nomes
Jornalístico (apenas CETENFolha)	JOCF	27.742.555	79,45%	12%	20%
Ensaio	ENS	2.193.635	6,28%	12%	21%
Jornalístico (sem CETENFolha)	JO	2.079.134	5,95%	13%	20%
Texto legal	LE	1.111.859	3,18%	10%	19%
Literatura	LI	921.365	2,64%	14%	18%
Texto didático	DI	426.765	1,22%	12%	22%
Enciclopédia	ENC	286.558	0,82%	10%	22%
Revistas	RE	153.786	0,44%	13%	22%
Texto epistolar	EP	3.350	0,01%	11%	20%
Total		34.919.007		12%	20%

Fonte: NILC/São Carlos v. 9.2 (Linguatca, 2011).

Naturalmente, a proporção total reflete muito o peso da seção jornalística, que chega a quase 80% do *corpus*. Contudo, os detalhes mostram as diferentes proporções nos outros gêneros de escrita representados no *corpus*. De modo geral, confirma-se na amostra de escrita brasileira extraída do NILC/São Carlos a mesma tendência observada no *Corpus* do Português, ressaltando-se algumas diferenças de pequena proporção. Esta similaridade de dados com relação ao *corpus* balanceado nos permite realizar uma comparação com o *corpus* de língua falada assumindo o NILC/São Carlos como representativo da variedade escrita do PB.

Em relação à variedade falada do PB representada pelo C-ORAL-BRASIL (Gráfico 1), nota-se que a proporção de nomes (14%) é semelhante àquela do PE falado (C-ORAL-ROM), que apresenta 15% de nomes. Já em relação aos verbos, o PB apresenta uma frequência bem maior do que aquela observada no *corpus* de fala do PE (18%), e mesmo da seção falada do *Corpus* do Português (20%), apresentando um total de 24% de verbos.

É interessante notar que a soma do total de verbos e nomes nas amostras consideradas é relativamente semelhante. A menor proporção encontrada é a do NILC/São Carlos (32%), e a maior proporção de nomes

mais verbos em relação ao total de palavras é a do *Corpus* do Português (39%). Esta última é bastante semelhante na proporção total encontrada no C-ORAL-BRASIL (38%), ainda que a proporção de nomes e verbos separadamente seja bem diferente, pois no C-ORAL-BRASIL encontramos uma proporção muito maior de verbos do que de nomes.

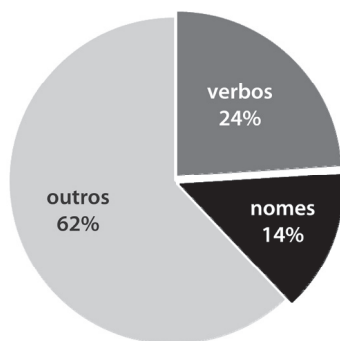


Gráfico 1 - Proporção de nomes e verbos no C-ORAL-BRASIL

Acreditamos que a distância entre nomes e verbos observada no PB em relação ao PE falado seja, pelo menos parcialmente, devido às diferenças na estrutura do *corpus*. É importante lembrar que os valores obtidos para o PE são relativos à totalidade deste *corpus*, que conta com uma seção informal e uma seção formal. No C-ORAL-BRASIL, até o momento, apenas a seção informal foi concluída.

De todo modo, não parece confirmado para o português, tanto o brasileiro quanto o europeu, a observação de Biber *et al.* (1999) acerca da frequência de verbos e nomes na fala e na escrita. Segundo documentado pelos pesquisadores em *corpora* de língua inglesa, na escrita teríamos uma proporção de nomes cerca de três vezes maior do que de verbos. Isso parece não acontecer na escrita do PB (ou mesmo do PE, até onde podemos avaliar) nem mesmo nos gêneros em que a distância entre nomes e verbos é maior, como a escrita legal e a enciclopédico-acadêmica.

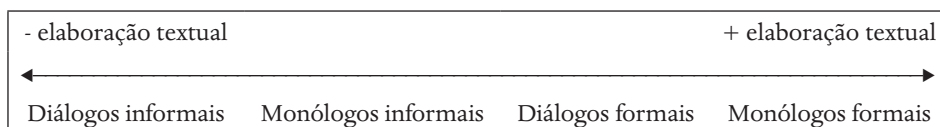
2.2 Análise interlinguística da ocorrência de nomes e verbos na fala espontânea

Constatamos que no português falado os verbos ocorrem com mais frequência do que os nomes, uma tendência contrária àquela do português escrito. Passaremos agora para uma análise mais detalhada da fala,

considerando não somente os dados do PB e do PE, mas comparando o C-ORAL-BRASIL com todos os *corpora* do C-ORAL-ROM, quais sejam, o italiano, o espanhol (europeu) e o francês.

Os *corpora* do C-ORAL-ROM apresentam uma seção informal, composta por textos dialógicos e monológicos em situações públicas e privadas; e uma seção formal, composta por interações dialógicas e monológicas, além de textos midiáticos e interações telefônicas. No caso dos textos oriundos de interações face a face, foi possível observar que a formalidade da situação acarreta um maior grau de elaboração textual, e com isso emerge na fala uma maior quantidade de estruturas complexas do que ocorre em textos produzidos em situações informais. Cresti (2005b) propõe que a complexidade estrutural dos textos segue a mesma escala da formalidade. Quanto mais informal a situação, menor é a elaboração textual; quanto mais formal a situação, maior será a elaboração textual.

Os textos falados se distribuiriam segundo a seguinte escala (adaptada de Cresti, 2005b, p. 218):



Conforme o que foi observado no C-ORAL-ROM, a frequência relativa de verbos e nomes parece estar relacionada com a formalidade da situação na qual o texto foi produzido. Maior formalidade parece estar correlacionada ao maior uso de formas nominais na fala. No contexto atual do C-ORAL-BRASIL, não é possível fazer a comparação em relação às seções formais, mas ainda assim se pode verificar se há diferença na frequência de verbos e nomes em textos de caráter dialógico *versus* monológico.

A Tabela 3 apresenta uma comparação entre as seções informais dos *corpora* do C-ORAL-ROM (PE – Português Europeu; IT – Italiano; ES – Espanhol; FR – Francês) com o C-ORAL-BRASIL no que se refere à frequência relativa de nomes e verbos. Para cada *corpus* apresentam-se os valores de verbos e nomes em textos dialógicos (dl, que abrangem diálogos e conversações) e monólogos (mn).

Tabela 3
 Frequência percentual de nomes e verbos
 nas seções informais do C-ORAL-BRASIL e C-ORAL-ROM

Parte do discurso	PB		PE*		IT*		ES*		FR*	
	dl	mn	dl	mn	dl	mn	dl	mn	dl	mn
verbos	23,92	20,66	18,67	18,60	20,62	19,84	15,16	15,22	19,80	18,16
nomes	12,43	15,12	13,45	15,60	16,27	18,99	11,30	12,86	12,70	13,82
Total	36,35	35,78	32,12	34,20	36,89	38,83	26,46	28,08	32,50	31,98

*Fonte: Cresti, 2005b, p. 218-219.

Podemos notar que, efetivamente, a frequência relativa de nomes e verbos muda de acordo com as ramificações dos *corpora*. Em PB, temos no dialógico 23,92% de verbos e 12,43% de nomes. No monológico, os verbos diminuem para 20,66%, e os nomes aumentam para 15,12%. A tendência para um aumento dos verbos no dialógico e dos nomes no monológico é verificada em todas as línguas românicas analisadas acima. A língua que apresenta a menor variação é o PE, na qual a frequência de verbos é praticamente igual, seja nos textos monológicos seja nos dialógicos. Talvez isso seja consequência de algumas características do *corpus* do PE, no qual a distinção entre a tipologia monológica e a dialógica é menos marcada.

Olhando para as particularidades do PB em relação às outras línguas apresentadas na Tabela 3, notamos que o PB é a língua com a maior porcentagem de verbos, tanto no dialógico (quase 24%) quanto no monológico (20,66%). A segunda língua que possui mais verbos é o italiano, com 20,62% no dialógico e 19,84% no monológico. Singular é a situação do espanhol, com apenas 15,16% de verbos no dialógico e praticamente o mesmo número no monológico.

Quanto aos nomes, o PB apresenta uma baixa frequência no dialógico (12,43%, superior unicamente ao espanhol, que tem 11,30%), mas uma proporção maior no monológico (15,12%, um aumento absoluto de 2,69% e percentual de mais de 20%), ocorrência inferior somente em comparação com o italiano, que apresenta 19% de nomes em sua seção informal monológica. A soma de nomes mais verbos é relativamente semelhante em todos os *corpora*, como já observado anteriormente também em relação à escrita. Contudo, impressiona a baixíssima porcentagem tanto de nomes quanto de verbos em ambas as tipologias para o espanhol, fator que mereceria ser melhor aprofundado.

Em síntese, corrobora-se a ideia de que a situação comunicativa determina fatores estruturais dos textos falados, os quais estão relacionados ao tipo de preenchimento lexical e também ao grau de elaboração textual.

3. Unidades estruturais naturais da fala espontânea

Uma das características constitutivas da língua falada é o fato de ser transmitida através do som. Este fato tem implicações importantes no modo como a informação é estruturada na fala. As unidades estruturais naturais da fala espontânea são aquelas delimitadas pela própria dinâmica da comunicação, os turnos dialógicos; e aquelas marcadas linguisticamente, através da entonação, como unidades concluídas e comunicativamente autônomas, os enunciados.

Nesta seção apresentamos algumas medidas relacionadas a essas unidades naturais que nos permitem, através de dados quantitativos, compreender como se organiza a fala espontânea.

3.1 O turno dialógico

A primeira unidade de referência natural de um texto falado é o turno dialógico, definido como um trecho de fala contínuo de um mesmo falante, delimitado pela fala de outro. Apesar de ser uma unidade de organização natural da fala espontânea face a face, não é adequado considerar o turno dialógico como a unidade de referência para a análise linguística. Isso porque a constituição de um turno é dada muito mais pela dinâmica comunicativa do que por fatores linguísticos. Efetivamente, não parece ser possível estabelecer uma propriedade de caráter estritamente linguístico que englobe qualquer turno dialógico possível. Isso torna-se claro à medida que observamos dados e exemplos da fala espontânea.

Os dados da Tabela 4, extraídos do C-ORAL-BRASIL, nos mostram os valores médios de número de enunciado por turno e de número de palavras por turno. O valor mínimo refere-se ao texto do *corpus* C-ORAL-BRASIL com a menor média (de enunciados ou palavras por turno), e o valor máximo refere-se ao texto com a maior média (de enunciados ou palavras por turno). Essas medidas merecem algumas considerações.

Tabela 4
Valores médios de enunciados e palavras por turno

Tipologia interacional	Enunciados por turno			Palavras por turno		
	Valor mín.	Valor máx.	Média geral	Valor mín.	Valor máx.	Média geral
Conversações	1,19	2,12	1,46	4,38	14,01	7,45
Diálogos	1,46	3,53	1,83	6,36	25,25	9,64
Monólogos	1,89	90,00	3,01	12,85	44,94	28,56

Na primeira parte da tabela acima, que mostra a média de enunciados por turno, observamos que, quanto às conversações e aos diálogos, os textos com o número mínimo e o número máximo de enunciados por turnos não se distanciam muito da média geral. Nos monólogos, ao contrário, este não é o caso. São duas as razões principais para isso, detalhadas a seguir.

Primeiramente, em interações informais, não existe um monólogo perfeito. No C-ORAL-BRASIL temos um caso excepcional, representado pelo monólogo familiar *bfammn30*, que apresenta o valor máximo de enunciados por turno. Trata-se de um único turno que soma 90 enunciados, em um trecho no qual não há intervenção por parte do interlocutor com o monologante. Em todos os demais textos do *corpus*, em maior ou menor medida, o interlocutor sempre interage com o monologante.

Em segundo lugar, nos textos monológicos, quando o interlocutor interage com o monologante, frequentemente ele o faz manifestando o seu acordo, expressando que está acompanhando o discurso do monologante, ou então suas impressões sobre a fala do outro através de enunciados bastante curtos e estruturalmente simples. Esses enunciados têm uma função interativa, mais do que uma função informativa.

Um exemplo paradigmático está no trecho abaixo de *bfammn02* (exemplo 1), em que o monologante (DFL) segue na construção da própria narrativa, independentemente das intervenções do interlocutor (LUC). O texto do monólogo continua, portanto, sendo fruto do planejamento inicial do falante, e não é consequência da interação do interlocutor. Essa é a característica que nos leva a considerar esse texto como um monólogo, ainda que, em média, o texto *bfammn02* seja composto por turnos curtos e, muitas vezes, mais simples do ponto de vista estrutural.

Exemplo 1 (bfammn02)

*DFL: *que o meu avô / era de uma família abastada / porque o professor ia em casa / nã ia po grupo não // ia em casa / o professor / ensinava / a ler / escrever / noções de geografia / história / ciência /*

*LUC: *hum hum //*

*DFL: */ aquilo que o / professor achava mais importante // e ele então nã teve / uma escola //*

*LUC: *<hum hum> //*

*DFL: *<ele> teve / &he / um professor // que foi em casa / que o instruiu // tinha uma &la [/1] caligrafia maravilhosa / <que mamãe> tinha até o cartão dele / mamãe falava que ele tinha um português correto //*

*LUC: *<hum hum> //*

*DFL: *mas ele quis que todos os filhos estudassem // formassem //*

*LUC: *hum hum //*

*DFL: *o mais velho / &ti [/1] &Fla [/1] chamava Flaviano + porque a minha bisavó chamava Flávia Augusta //*

*LUC: *hum hum //*

Os dados da Tabela 4 nos fazem notar também que há diferenças entre conversações e diálogos, mesmo que esses dois tipos de interação sejam tendencialmente muito semelhantes entre si. Os dados mostram uma média de enunciados por turno superior no diálogo. Isso se deve ao fato de uma maior competição pelo turno nas conversações. Outra razão importante é o fato de, algumas vezes, certas interações dialógicas permitirem trocas de turnos relativamente longos entre os interlocutores. Isso ocorre principalmente no caso da gravação de sessões, em que os participantes simplesmente se engajam em discussões em volta de algum assunto, ao invés de estarem interagindo um com o outro na execução de alguma tarefa⁹ (situação mais acional, o que, tendencialmente, leva à maior interatividade).

De modo geral, o número de enunciados por turno é uma medida muito significativa do nível de interatividade dos textos. Como regra geral, quanto maior o número de enunciados por turno, menor será o grau de interatividade e maior o grau de elaboração textual. Contudo,

⁹ No C-ORAL-BRASIL, há o registro da situação em que foi realizada a gravação nos metadados de cada transcrição (veja-se o DVD neste livro).

a medida do turno, assim como pode ser capturada na fala alinhada do *corpus*, não é suficiente para compreender a dinâmica da fala espontânea e a sua interatividade. Em algumas situações, a interatividade não é expressa linguisticamente, é antes dada através da sintonia entre a ação verbalizada de um falante e a ação não verbalizada de outros. Infelizmente, tais casos não são captados e representados na transcrição da fala adotada no *corpus*, e as medidas que temos refletem este limite.

Um exemplo evidente disso é a interação entre um professor de ginástica e o aluno. A situação é fortemente interativa, porque cada enunciado do professor depende da ação executada pelo interlocutor, mas essa dinâmica não é refletida na transcrição (exemplo 2).

Exemplo 2 (*bpubdl03*)

*GUI: volta aqui / faz força // mais // beleza // contrai o abdômen / joga o tronco só um pouquinho pra frente // aí // beleza // descansou // vou baixar um pouquinho mais // vai // pera aí // deixa eu passar a faixa // aí // vai // força // aqui / pra frente // isso // pesado //

Casos como esse mostram que, na verdade, a média de enunciados por turno (assim como o turno é capturado no sistema de alinhamento) representa um valor aproximado para o alto. Essa medida, nas situações dialógicas, é certamente mais baixa do que o mostrado pelos dados.

3.2 Estruturação da fala em enunciados simples e complexos

As transcrições do C-ORAL-BRASIL baseiam-se nos parâmetros estabelecidos pela Teoria da Língua em Ato (apresentada no Capítulo 3 deste livro). De acordo com esse panorama teórico, a unidade de referência da fala espontânea deve ser considerada o enunciado. O enunciado corresponde a uma sequência linguística que é suficiente e autônoma do ponto de vista pragmático (Cresti, 2000a; Cresti; Moneglia, 2010). O enunciado é delimitado entonacionalmente por uma fronteira prosódica de tipo terminal. Fronteiras prosódicas não terminais delimitam unidades internas ao enunciado (unidades tonais), que estão relacionadas a diferentes funções informacionais. Assim, os enunciados podem ter estrutura simples (sem segmentação interna), ou complexa (com segmentação interna).

De modo geral, a maior ou menor complexidade estrutural dos enunciados está relacionada a uma maior ou menor elaboração textual.

Assim, uma primeira maneira de atestar a diferença entre a complexidade dos enunciados comparando as tipologias interacionais é observar quão complexos, em termos entonacionais, são os enunciados. A Tabela 5 mostra a média de unidade tonal por enunciado em cada seção do *corpus*. As medidas foram calculadas com base apenas nas unidades concluídas (enunciados interrompidos e *retractings*¹⁰ não foram considerados, mas são apresentados mais adiante).

Observando a complexidade relativa ao número de unidades tonais por enunciado, que é diretamente relacionada à complexidade informacional,¹¹ notamos que a complexidade dos textos dialógicos (diálogos e conversações) é muito menor do que a complexidade dos monológicos. De fato, a média de unidade tonal por enunciado é bastante semelhante em conversações (1,58) e em diálogos (1,61). Nos monólogos essa média sobe para 2,42 unidades tonais por enunciado.

A diferença entre o contexto familiar e o público é marcante somente nos monólogos. Enquanto nos dialógicos a diferença na média de unidades tonais em contexto público e privado é muito pequena (de 0,10 nas conversações e de 0,04 nos diálogos), nos monólogos, a média é de 2,42 unidades tonais por enunciado em contexto familiar e de 3,56 no contexto público, uma diferença de 1,14 unidades tonais a mais em contexto público.

Tabela 5
Complexidade dos enunciados do C-ORAL-BRASIL

Contexto	Tipologia	Unidades tonais por enunciado	Palavras por enunciado	Palavras por unidade tonal ^(Continua)
	Conversações	1,58	5,05	3,19
	Diálogos	1,61	5,33	3,31
Familiar	Conversações + diálogos	1,60	5,19	3,25
	Monólogos	2,42	8,74	3,61
	Total	1,78	5,99	3,36

¹⁰ Para explicações detalhadas e exemplos dos diferentes tipos de unidades e anotação realizada na transcrição, veja-se o Capítulo 2, que contém a descrição do *corpus*.

¹¹ Veja-se o Capítulo 3.

Contexto	Tipologia	Unidades tonais por enunciado	Palavras por enunciado	Palavras por unidade tonal
Público	Conversações	1,68	5,37	3,21
	Diálogos	1,57	5,02	3,20
	Conversações + diálogos	1,61	5,17	3,20
	Monólogos	3,56	13,09	3,68
	Total	2,08	7,09	3,41
Familiar + público	Conversações	1,60	5,12	3,19
	Diálogos	1,60	5,25	3,29
	Conversações + diálogos	1,60	5,19	3,24
	Monólogos	2,61	9,48	3,63
	Total	1,83	6,16	3,37

Se olharmos a complexidade dos enunciados através da medida da palavra, e não da unidade tonal, chegamos às mesmas conclusões, como mostram os dados da tabela acima. Essa medida mostra que a complexidade dos monólogos é quase o dobro da complexidade dos dialógicos.

Já o número de palavras por unidade tonal é praticamente uma constante e independe tanto do contexto quanto do tipo de interação. Esse valor representa o limite médio do conteúdo lexical, que é possível ser envelopado em uma mesma unidade prosódica em PB. Naturalmente, uma medida mais precisa a ser considerada seria a do número de sílabas fonéticas por unidade tonal.

Na Tabela 6 apresentamos o percentual de enunciados simples e complexos do C-ORAL-BRASIL em comparação com a seção informal do C-ORAL-ROM, de acordo com a tipologia da interação. Os textos dialógicos (conversações e diálogos) representam 2/3 do número de enunciados de cada amostra, enquanto a tipologia monológica conta com uma porcentagem de enunciados de 1/3 do total. Dado que conversações e diálogos se comportam da mesma maneira,¹² podem ser considerados juntos na categoria “dialógico”.

¹² Os dados detalhados para cada texto do *corpus* estão disponíveis no DVD, neste livro.

Tabela 6
 Proporção de enunciados simples e complexos
 no C-ORAL-BRASIL e C-ORAL-ROM

Tipologia	Estrutura	PB	PE*	IT*	ES*	FR*
Dialógico	Simple	58,7%	50,2%	52,0%	57,8%	69,2%
	Complexo	41,3%	49,8%	48,5%	42,2%	30,8%
Monológico	Simple	43,2%	32,4%	30,5%	32,4%	44,1%
	Complexo	56,8%	67,6%	69,5%	67,6%	55,9%

*Fonte: Cresti, 2005b, p. 222.

De acordo com a Tabela 6, o tipo da interação (monológica ou dialógica) representa uma variável significativa na estruturação dos enunciados. Os monólogos apresentam uma complexidade de enunciados claramente superior à dos dialógicos em todas as línguas comparadas.

No PB, enquanto nas conversações e diálogos (dialógico) a porcentagem de enunciados simples é de 58,7%, nos monólogos essa proporção cai para 41,3%. Nos monólogos, os enunciados complexos representam a maioria (56,8%). Já o PE apresenta uma distribuição praticamente equivalente de enunciados simples (50,2%) e complexos (49,8%) no dialógico, o que, mais uma vez, atesta uma diferença estrutural quanto aos textos dialógicos do *corpus* do PE, como já observado na distribuição de verbos e nomes. O italiano (IT) é a língua que apresenta a maior proporção de enunciados complexos nos monólogos (69,5%), enquanto o francês é a língua com a maior proporção de enunciados simples no dialógico (69,2%).

Segundo Cresti (2005b), a opção por uma estrutura simples ou complexa não está ligada à formalidade dos textos, mas sim à estrutura do evento comunicativo (tipologia interacional), que opõe dialógicos (informais e formais) a monólogos (informais e formais). Os dados apresentados até aqui fornecem alguma evidência disso, indicando que textos monológicos apresentam maior elaboração textual do que os dialógicos; entretanto, especialmente em PB, a diferença na proporção de enunciados simples e complexos em situação dialógica e monológica ainda parece pequena. Consideramos importante explorar com mais detalhe a questão da complexidade dos enunciados no *corpus* brasileiro. Esse detalhamento é apresentado na seção seguinte.

3.3 Complexidade informacional na fala espontânea

É possível aprofundar o raciocínio sobre a complexidade dos enunciados, pois é importante levar em consideração diferentes tipos de enunciados complexos. Enunciados complexos são aqueles que apresentam mais de uma unidade tonal, e, portanto, mais de uma unidade informacional – a unidade de Comentário e uma ou mais unidades textuais e/ou dialógicas. Por sua vez, os enunciados simples são aqueles que têm apenas uma unidade tonal/informacional, a unidade de Comentário.¹³ Assim, um enunciado complexo pode ser composto por:

- a) comentário e uma ou mais unidades dialógicas;
- b) comentário e uma ou mais unidades textuais;
- c) comentário e uma ou mais unidades dialógicas em conjunto com uma ou mais unidades textuais.

Torna-se então interessante desagregar os dados dos enunciados complexos e não tratá-los como uma única categoria, pois a natureza informacional dessa complexidade pode ter motivações diferentes nos enunciados de textos dialógicos e monológicos. Devemos tratar de maneira diferente os enunciados que são complexos porque há unidades que acrescentam informação ao conteúdo locutivo do enunciado (tipos b e c) daqueles que são complexos unicamente do ponto de vista dialógico, mas não informacional (tipo a).

Para desagregar esses dados, precisamos de um *corpus* que tenha recebido etiquetagem das funções informacionais de cada unidade tonal. No momento, dispomos de um *subcorpus* do C-ORAL-BRASIL que recebeu etiquetagem informacional. Este *subcorpus* compreende 20 textos e corresponde a 15% do *corpus*. A estrutura do *subcorpus* é a mesma do *corpus*. Essa amostra conta com 31.318 palavras em 20 textos: seis conversações,

¹³ As unidades do padrão informacional são definidas conforme a Teoria da Língua em Ato, apresentada no Capítulo 3. Em síntese, a TLA postula a hipótese de que cada unidade de um padrão prosódico de um enunciado corresponde a uma unidade de um padrão informacional. Todo padrão é composto de pelo menos uma unidade nuclear, que tem como função veicular a força ilocucionária, concedendo autonomia pragmática ao enunciado. Essa unidade é chamada de Comentário. Há ainda outras unidades, que podem servir para a elaboração textual (Tópico, Apêndice, Introdutor Locutivo e Parentético) ou cumprir função dialógica, pois são direcionadas ao interlocutor e regulam o funcionamento da interação (Fático, Conativo, Expressivo, Conector Discursivo, Alocutivo e Incipitário).

sete diálogos e sete monólogos. As conversações abrangem um total de 9.774 palavras; os diálogos têm 11.331 palavras, e os monólogos, 10.213 palavras. Ao final deste capítulo encontra-se um anexo contendo a descrição detalhada de cada texto do *subcorpus*. A proporção de enunciados simples e complexos do *subcorpus* é mostrada na Tabela 7.

Tabela 7
Enunciados simples e complexos no total do C-ORAL-BRASIL
e no *subcorpus* etiquetado informacionalmente

Amostra	Contexto	Tipologia	Enunciados simples	Enunciados complexos
C-ORAL-BRASIL	Familiar + público	dialógico	58,7%	41,3%
		monológico	43,2%	56,8%
<i>Subcorpus</i>	Familiar + público	dialógico	61,8%	38,2%
		monológico	37,9%	62,1%

Vale ressaltar que, nas interações informais, o monólogo perfeito não existe, como é o caso das interações formais. Portanto, na seção informal temos textos fundamentalmente monológicos (veja-se a discussão no Capítulo 2). O *subcorpus* foi construído para refletir de modo mais preciso a distinção entre textos monológicos e dialógicos. No *corpus* como um todo, temos diversos textos dialógicos com turnos longos e vários textos monológicos que contêm pequenas inserções de trechos dialógicos. Com isso, a proporção de enunciados simples e complexos no *corpus* inteiro tende a ser menor do que a observada no *subcorpus*, conforme pode ser verificado na Tabela 7.

Como o que nos interessa aqui é uma análise da relação entre diversos tipos de enunciados (descritos em a, b e c acima) e as tipologias (dialógica e monológica), devemos fazer essa análise para o *subcorpus* e projetar os resultados para o *corpus*, tendo em mente essa diferença. O C-ORAL-BRASIL como um todo possui 3% a mais de enunciados complexos nos textos dialógicos do que o *subcorpus*. Ao mesmo tempo, este último apresenta 5% a mais de enunciados complexos nos textos monológicos do que o *corpus* como um todo.

Os dados da Tabela 8 mostram os valores absolutos e a proporção dos enunciados complexos que apresentam somente unidades dialógicas,

dos complexos que apresentam apenas unidades textuais e dos mistos, que contêm tanto unidades textuais quanto dialógicas. A proporção foi calculada em relação ao total de enunciados complexos por tipologia.

Tabela 8
Diferentes tipos de enunciados complexos no *subcorpus* do C-ORAL-BRASIL

Tipologias interacionais	Enunciados complexos somente unidades dialógicas		Enunciados complexos somente unidades textuais		Enunciados complexos mistos	
Conversações	196/348	56,3%	108/348	31,0%	44/348	12,6%
Diálogos	232/399	58,1%	125/399	31,3%	42/399	10,5%
Dialógico*	428/747	57,3%	233/747	31,2%	86/747	11,5%
Monológico	63/219	28,8%	100/219	45,7%	56/219	25,6%
Total	491/966	50,8%	333/966	34,5%	142/966	14,7%

*Somatório das conversações com os diálogos.

Novamente esses dados mostram uma diferença praticamente nula entre as tipologias dialógicas (diálogos e conversações). Além disso, ao tratar de maneira diferenciada os enunciados complexos, percebemos que nas tipologias dialógicas há forte predominância de enunciados que são complexos unicamente pela presença de uma ou mais unidades dialógicas, mas sem unidades textuais, visto que 57,3% dos enunciados em textos dialógicos apresentam apenas unidades de auxílios dialógicos. Já quanto aos monólogos, não somente possuem maior ocorrência de enunciados complexos de modo geral, mas ainda possuem domínio de enunciados textualmente mais complexos, visto que 45,7% dos enunciados nessa tipologia apresentam apenas unidades textuais, e mais 25,6% são mistos, ou seja, apresentam tanto unidades textuais quanto dialógicas.

No Gráfico 2 somamos os casos de enunciados mistos com os complexos com somente unidades textuais nas tipologias dialógicas e monológicas. Desse modo, comparamos a proporção de enunciados complexos com somente unidades dialógicas com a proporção de enunciados complexos que apresentam pelo menos uma unidade textual.

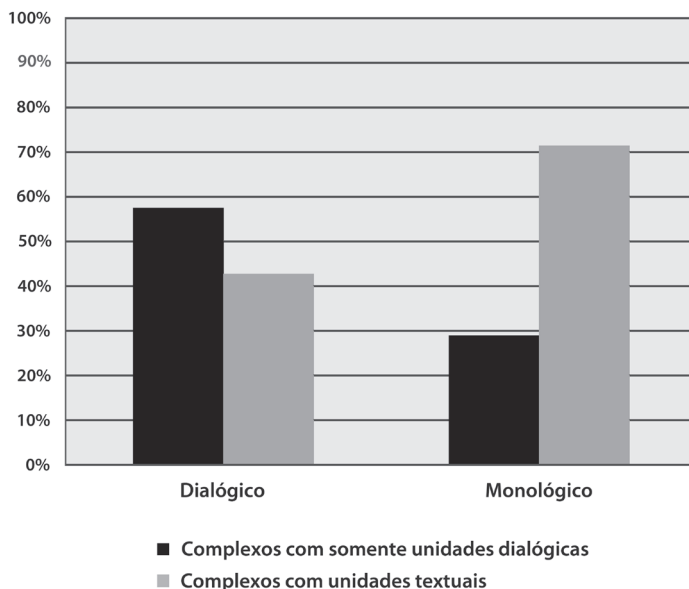


Gráfico 2 - Proporção de enunciados complexos com somente unidades dialógicas e com unidades textuais

Essas considerações nos permitem dizer que a diferença da complexidade dos enunciados nas duas tipologias (dialógica e monológica) não é somente de natureza quantitativa, mas também qualitativa. Se considerarmos a complexidade unicamente informacional, a distância entre textos dialógicos e monológicos aumenta muito. Ao mesmo tempo, como pode ser verificado nos dados da Tabela 8, a diferença na proporção de enunciados complexos de cada tipo entre dialógico e monológico é sempre bem superior à diferença observada entre o *subcorpus* e o *corpus* inteiro. Em vista disso, podemos assumir que no C-ORAL-BRASIL como um todo os dados se comportam da mesma maneira como a descrita para o *subcorpus*.

3.4 Fenômenos de fragmentação da fala

Na fala espontânea, é natural que ocorra uma eventual fragmentação, seja no nível da palavra como no nível do enunciado. Apresentamos nesta seção os dados relacionados à fragmentação da fala. Entendemos por fenômenos de fragmentação:

1. palavras incompletas produzidas pelo falante, que são marcadas na transcrição com o símbolo “&”;

2. *retracting*, fenômeno que ocorre quando um falante imediatamente se retrata do que foi dito, sempre no mesmo enunciado, podendo ou não repetir uma ou mais palavras pronunciadas. É marcado na transcrição com a barra simples seguida do número de palavras envolvidas no *retracting*, dentro de colchetes (ex. [/3]);
3. interrupção de enunciado, que é assinalada na transcrição através do símbolo “+”.

Os dois primeiros tipos de fenômenos estão presentes no exemplo 3.

Exemplo 3 (*bfamd124*)

*DUD: <ela é do [/2] é [/1] é> [/1] é do &dese [/1] desenvolvimentismo do Juscelino //

No enunciado do exemplo 3 aparece uma palavra interrompida (&dese) e quatro fenômenos de *retracting*. Note-se que os números dentro dos colchetes indicam a quantidade de palavras canceladas pelo falante. Sem os *retractings* e a palavra interrompida, o enunciado seria como mostrado no exemplo 3a.

Exemplo 3a

*DUD: ela é do desenvolvimentismo do Juscelino //

Em relação ao terceiro tipo de fragmentação indicado acima, há duas possibilidades: quando o próprio falante se interrompe e muda o plano do discurso iniciando um enunciado novo, ou quando uma circunstância externa interrompe o falante, normalmente um outro falante presente na situação. O exemplo 4 mostra um enunciado interrompido devido à mudança de plano da falante (VER).

Exemplo 4 (*bfamcv12*)

*VER: <então / a> + <não> // não // a gente vai negociar / e vai ter um acesso / que nã atrapalha ninguém //

O exemplo 5 mostra dois casos de enunciado interrompido que são de naturezas diferentes. No primeiro caso, o falante se interrompe porque o interlocutor rouba o turno. No segundo caso, a interrupção é interna ao turno do falante, mas parece diferente do que acontece no exemplo anterior. De fato, se no exemplo anterior não há dúvida de que o falante resolve

interromper o próprio enunciado e começar algo diferente, no exemplo seguinte é duvidoso se este tipo de fragmentação deve ser considerado um enunciado interrompido ou um *retracting*. Este exemplo é ilustrativo do fato de que nem sempre é fácil distinguir entre uma mudança de plano e uma correção.

Exemplo 5 (*bfamcv12*)

*GIL: *mas isso / &he / complicado no sentido assim +*

*CAR: *é &total + é muito mais leve do que um <telhado> //*

A Tabela 9 mostra a porcentagem dos três fenômenos de fragmentação de acordo com o contexto, a tipologia interacional e no total do *corpus*.

Tabela 9
Fenômenos de fragmentação no C-ORAL-BRASIL

Contexto	Tipologia	Palavras interrompidas	<i>Retracting</i> por unidade tonal	<i>Retracting</i> por enunciado	Enunciados interrompidos
Familiar	Conversações	3,9%	7,2%	11,4%	8,2%
	Diálogos	4,33%	9,5%	15,3%	7,9%
	Monólogos	4,43%	13,1%	31,7%	6,9%
Público	Conversações	4,76%	10,0%	16,8%	8,8%
	Diálogos	4,10%	8,3%	13,1%	7,5%
	Monólogos	5,30%	16,8%	60,2%	9,2%
Total		4,34%	10,2%	18,5%	7,9%

Observamos que, de modo geral, a quantidade de palavras interrompidas é baixa (4,34%) e não parece mudar muito conforme o contexto e a tipologia. Isso indica que esse tipo de fragmentação é um fenômeno provavelmente mais ligado ao falante do que ao tipo de texto produzido. Também o número de enunciados interrompidos se mantém abaixo de 10%, e a variação não parece se dar em função do contexto ou da tipologia interacional.

Já os fenômenos de *retracting* têm maior incidência na fala. Além disso, notamos que os monólogos são mais sensíveis a esse tipo de fenômeno de fragmentação. Isso pode ser explicado pela própria natureza processual da tipologia. Os monólogos são como um “pensamento falado”; são textos

maiores e mais elaborados, cujo planejamento é realizado durante a própria produção da fala e que, além disso, não são resultado de interatividade imediata com o interlocutor. Sendo assim, naturalmente os monólogos apresentam um grau de monitoramento maior, o que induz a um maior número de correções.

Se calculamos o percentual de *retractings* sobre o número total de enunciados, a porcentagem relativa do fenômeno sobe. Essa medida é especialmente interessante porque mostra que a maioria dos enunciados em tipologias monológicas contém pelo menos um *retracting*. Nessa medida, a distância entre as tipologias monológicas e dialógicas é bastante grande, devido justamente ao fato de os enunciados dos monólogos terem uma quantidade de unidades tonais em geral superior à dos diálogos e conversações.

Do ponto de vista da fala como um todo, é interessante observar que 26,4% dos enunciados possuem algum fenômeno de fragmentação, que são a interrupção (7,9%) e o *retracting* (18,5%). Ainda que mais de 1/4 dos enunciados na fala espontânea apresente fenômenos de fragmentação, essa quantidade relativamente significativa de enunciados envolve apenas 4,3% das palavras do *corpus*. Isso indica que a perda em termos de informação é muito menor do que as dificuldades na hora de veicular a informação.

3.4 Tamanho e duração das unidades de referência da fala

Apresentamos algumas medidas relacionadas ao tamanho das unidades de referência da fala (turnos, enunciados e unidades tonais) medidas para o C-ORAL-BRASIL em comparação com o C-ORAL-ROM, quais sejam: número médio de palavras por turno, por enunciado e por unidade tonal; número médio de enunciados por turno e por minuto; número médio de unidades tonais por enunciado e número de palavras por segundo. Essas medidas fornecem a possibilidade de comparar a densidade informacional e ilocucionária das línguas representadas nesses *corpora*. A Tabela 10 apresenta os valores comparativos dessas medidas.

Tabela 10
Valores médios do tamanho das unidades de referência
no C-ORAL-BRASIL e no C-ORAL-ROM

Medidas	PB	PE*	IT*	ES*	FR*
Palavras por turno	11,37	22,92	13,22	16,93	26,16
Palavras por enunciado	6,16	7,54	6,51	7,81	14,49
Palavras por unidade tonal	3,37	2,86	2,71	3,10	4,97
Enunciados por turno	1,85	3,04	2,03	2,17	1,81
Enunciados por minuto	29,93	24,48	23,55	24,50	14,41
Unidades tonais por enunciado	1,83	2,63	2,40	2,52	2,75
Palavras por segundo	2,76	3,08	2,56	3,19	3,48

*Fonte: DVD 9 do C-ORAL-ROM (Cresti; Moneglia, 2005).

Avaliando os dados da Tabela 10 comparativamente, vemos que o PB é a língua que tem o segundo maior número de palavras por unidade tonal (depois do francês). Observamos que o PB apresenta grande número de unidades tonais bastante longas, conforme pode ser verificado nos enunciados dos exemplos 6, 7 e 8. Essa possibilidade de produzir unidades tonais mais longas, considerando o número de palavras, é provavelmente devido ao fato de a fala representada no C-ORAL-BRASIL (variedade mineira) ser tendencialmente mais acentual, em oposição ao italiano, que é uma língua silábica e apresenta o menor número de palavras por unidade tonal.

Exemplo 6 (bfamdl10)

*HEL: *como se aquela pessoa que tá atendendo ali é que tinha que resolver o problema / porque ela que fez o negócio estragar / entendeu //*

Exemplo 7 (bfammn06)

*JOR: *um belo dia durante o almoço / o gerente de recursos humanos de uma multinacional / me informou que havia uma vaga na área comercial da empresa / e / se eu tinha interesse //*

Exemplo 8 (bfammn06)

*CAR: *a única coisa que eu fiquei muito triste que eu não falo perto dela / Maira / é que / quando fez oito dia que ela tava com a gente / eu / não falo perto dela / porque / isso ela não sabe / é que / &he / a mãe / mandou buscar porque tinha vendido ela por seiscentos reais //*

Observe-se que a primeira unidade do exemplo 6, um comentário, tem entre 25 e 28 sílabas gráficas (dependendo do fato de se considerar os fenômenos de sinérese ou não), que a segunda unidade do exemplo 7, um tópico, possui entre 18 e 19 sílabas gráficas, e que a primeira unidade do exemplo 8, também um tópico, possui 23 sílabas gráficas. Mas o interessante é que unidades desse tamanho não são excepcionais no *corpus*, o que diz algo sobre as características da fala do PB, ou pelo menos da fala mineira, que mereceriam ser estudadas.

Contudo, apesar de as unidades tonais no PB serem longas em termos de número de palavras e de sílabas, quando se passa à medida por enunciado, nota-se que os enunciados do *corpus* de PB são mais curtos (com 6,16 palavras em média por enunciado). O PB apresenta ainda um número alto de enunciados por minuto, justamente por ter enunciados em média mais curtos (mas seria interessante investigar a velocidade de fala também).

O número de palavras por turno e de enunciados por turno também são os mais baixos entre os *corpora* comparados. Isso indica que há uma grande interatividade nas sessões gravadas do C-ORAL-BRASIL. O PB se aproxima mais do *corpus* italiano do que dos demais neste quesito. Em conclusão, o PB e o IT são os mais interativos, seguidos pelo ES e pelo PE, enquanto o FR é o mais distante de todos.

Quando temos poucos enunciados por turno, esperamos também menos palavras por turno. Em todas as línguas, o número de enunciados por turno é diretamente proporcional ao número de palavras por turno, com exceção do francês, que apresenta número alto de palavras por turno e um número baixo de enunciados por turno. Visto que, em diversas medidas avaliadas, os dados do francês se mostram desviantes, resta saber se essa diferença significa que a fala do francês é diferente em termos da organização prosódica ou se as diferenças observadas são fruto de decisões operacionais diferentes quando da compilação, transcrição e segmentação do *corpus*.

A medida de palavras por segundo indica, em princípio, que PB e PE têm uma velocidade de elocução maior. Contudo, essa medida pode ser facilmente enviesada, por um lado, pela presença de silêncios, que diminuem o número de palavras por segundo; e por outro, pelas sobreposições de fala, que aumentam o número de palavras por segundo. Situações que requerem atividades e ações não verbais tendem a ter mais silêncios (a título de exemplo

no C-ORAL-BRASIL, podem-se ver os textos *ppubdl01* e *bpubdl02*). É possível, então, que o número de palavras por segundo do PE seja devido à falta de silêncios, que acontecem naturalmente em situações de fala espontânea mais acionais e interativas. Um estudo sobre a real velocidade de fala média do PB deveria portanto editar os arquivos, de maneira a evitar os longos silêncios devidos a momentos da ação em que não há fala, coisa que não acontece nos bate-papos, quando o silêncio gera o fim da comunicação.

4. Enunciados verbais e não verbais

Até agora averiguamos o grau de estruturação da fala espontânea em termos lexicais (presença de verbos e nomes) e também quanto à organização dos enunciados em simples e complexos. Nesta seção, fazemos o cruzamento dessas duas variáveis, verificando a ocorrência de enunciados simples com e sem verbo, e enunciados complexos com e sem verbo.

A presença de uma forma verbal no enunciado funciona como um indicador estrutural. Quando uma forma verbal finita é utilizada em um enunciado, isso dá origem à estruturação sintática complexa. Enunciados que não apresentam verbo têm configuração sintática extremamente reduzida. Usualmente apresentam uma única expressão, acompanhada, no máximo, por seus determinantes, modificadores e atributos (Cresti, 2005b).

Os exemplos a seguir apresentam, respectivamente, enunciado simples sem verbo (exemplo 9), enunciado simples com verbo (exemplo 10), enunciado complexo sem verbo (exemplo 11) e enunciado complexo com verbo (exemplo 12).

Exemplo 9 (*bfamdl29*)

*ALV: *como assim //*

Exemplo 10 (*bpubmn11*)

*SAN: *o forte sempre foi educação //*

Exemplo 11 (*bfamdl02*)

*BEL: *uhn / talvez na parte maior / não //*

Exemplo 12 (*bpubcv08*)

*IAR: *avança essa vinhetinha / pega na parte que a Jú montou / por favor //*

Com base no que foi observado para as línguas do C-ORAL-ROM, Cresti (2005b) relaciona a estruturação simples com a tipologia dialógica, e o mesmo vale para a estratégia não verbal. Textos dialógicos são mais propícios ao aparecimento de enunciados não verbais. Já a tipologia monológica apresenta uma prevalência de enunciados complexos com verbo.

As porcentagens dos enunciados simples e complexos com e sem verbo calculadas para o C-ORAL-BRASIL são baseadas no *corpus* que passou pela etiquetagem informacional através do *parser* PALAVRAS (ver Bick, 2000; e Capítulo 6 deste volume). Esse processo ignora as palavras em *retracting* e também enunciados lexicalmente vazios (preenchidos por palavras incompletas ou por yyyy). Assim, o total de enunciados com etiquetagem morfossintática é menor do que o total geral de enunciados, e as proporções apresentadas aqui não são exatamente iguais às já mostradas anteriormente.

Em uma primeira análise comparamos o C-ORAL-BRASIL com o C-ORAL-ROM. Os dados relativos aos textos dialógicos são mostrados na Tabela 11, e os dados dos monólogos estão na Tabela 12.

Tabela 11
Enunciados simples e complexos com e sem verbo nos textos dialógicos do C-ORAL-BRASIL e do C-ORAL-ROM

Dialógico	PB	PE*	IT*	ES*	FR*
Simple sem verbo	23,8%	30,1%	33,0%	33,3%	26,9%
Simple com verbo	36,3%	20,1%	18,6%	24,5%	42,3%
Complexo sem verbo	5,7%	11,9%	11,9%	11,9%	11,9%
Complexo com verbo	34,1%	40,4%	36,6%	34,1%	28,2%

*Fonte: Cresti, 2005b, p. 229.

Tabela 12
Enunciados simples e complexos com e sem verbo nos textos monológicos do C-ORAL-BRASIL e do C-ORAL-ROM

Monológico	PB	PE*	IT*	ES*	FR*
Simple sem verbo	18,29%	23,25%	17,11%	15,25%	7,94%
Simple com verbo	26,51%	17,32%	13,44%	17,18%	36,16%
Complexo sem verbo	3,83%	9,87%	10,58%	7,00%	0,36%
Complexo com verbo	51,38%	49,53%	58,87%	60,57%	55,54%

*Fonte: Cresti, 2005b, p. 229.

É interessante notar que o PB possui uma proporção alta de enunciados simples verbais. Na tipologia dialógica, podemos notar, por exemplo, que o espanhol e o PB têm uma frequência de enunciados simples bastante semelhante (57% e 60%). Porém, o número de enunciados simples verbais desses dois *corpora* se distancia muito (24,5% e 36,3%). O PB também apresenta um baixo número de enunciados complexos sem verbo, tanto na tipologia dialógica quanto monológica.

Os dois indicadores mais consistentes em termos de variação de acordo com a estrutura do *corpus* são as estratégias opostas: “complexo com verbo” e “simples sem verbo”. Esses dois tipos de estruturação dos enunciados se mantêm em uma distribuição semelhante nas línguas do C-ORAL-ROM (à exceção do francês). Em PB, contudo, a proporção de enunciados simples com verbo é extremamente elevada. Acreditamos que isso, em parte, seja decorrência do grande número de enunciados simples compostos por verbos utilizados para afirmar ou negar em PB, mas em outras circunstâncias foi notada a forte presença verbal nos enunciados do PB.

Para verificar essa hipótese, realizamos uma análise mais detalhada nos 20 textos do *subcorpus* do C-ORAL-BRASIL. Tal análise tem um caráter mais qualitativo: todos os enunciados do *subcorpus* foram classificados não somente quanto à presença ou ausência de uma forma verbal, mas também classificamos se o verbo presente era uma forma finita, e se constituía o núcleo sintático do enunciado. Os enunciados com verbo foram classificados como:

- 1) enunciado verbal cujo verbo não apresenta função verbal, mas sim pragmática. Incluímos as formas reduzidas do imperativo de *olhar* (*o'* e *a'*), as formas *é*, *tá* e outras formas verbais quando utilizadas para afirmar ou negar (o enunciado tem valor de *sim* ou *não*), e a expressão *n' é não*;
- 2) enunciado verbal cujo verbo apresenta função verbal e é o núcleo sintático do enunciado. Neste caso incluímos também orações relativas e clivadas que não possuem verbo nuclear. Foram incluídas também nesta categoria as formas *olha*, *pera aí*, *entendi* (confirmação) e outras similares, mesmo sendo duvidosa a função verbal dessas palavras nestes casos;
- 3) enunciado verbal cujo verbo apresenta função verbal, mas não constitui núcleo sintático do enunciado. Trata-se de enunciados compostos de estruturas sintáticas dependentes ou enunciados cujo verbo é uma forma não finita.

O *subcorpus* tem um total de 5.498 enunciados; destes, 4.960 foram analisados. Os 538 enunciados não analisados são enunciados interrompidos ou aqueles que continham muitas palavras não transcritas (não compreendidas pelo transcritor) e que por isso não permitiram uma avaliação da estrutura do enunciado. Os resultados estão sumarizados na Tabela 13.

Tabela 13
Enunciados simples e complexos verbais e não verbais do *subcorpus*

Tipologia	Complexidade do enunciado	Verbal núcleo	Verbal não núcleo	Verbal sem função verbal	Não verbal	Total
Conversações	Simple	9,11%	0,79%	3,91%	9,01%	22,82%
	Complexos	9,68%	1,05%	0,87%	1,94%	13,53%
Diálogos	Simple	12,46%	1,63%	5,38%	10,44%	29,92%
	Complexos	11,13%	1,03%	0,71%	2,10%	14,96%
Monológico	Simple	3,33%	0,71%	0,58%	3,23%	7,84%
	Complexos	8,89%	1,33%	0,06%	0,65%	10,93%
Total		55%	7%	12%	27%	100%

Do total de enunciados analisados, 27% são não verbais (1.357 enunciados não verbais), ou seja, não contêm nenhuma forma verbal. Dentre os enunciados não verbais, 17% (224) são constituídos por simples manifestações de afirmação, negação ou sinal de que o interlocutor está acompanhando a interação como: *hum hum*, *hum*, *ham ham*, *ahn*, *uhn uhn* e *uhn*. Mesmo excluindo esses casos, continuamos tendo uma grande quantidade de ilocuições organizadas em enunciados não verbais.

A maioria dos enunciados (3.603 casos, correspondendo a 73% das ocorrências), entretanto, apresenta pelo menos um verbo. Entre estes estão os enunciados que apresentam mais de uma estrutura sintática independente, das quais pelo menos uma constitui núcleo. Quinhentos e setenta e um (12%) enunciados apresentam verbo, mas estes não têm função verbal. Vejam-se os exemplos a seguir.

Exemplo 13 (*bfamdl03*)

*LAU: *mas nã é bonitim demais / chegar aqui //*

*LUZ: *é //*

Exemplo 14 (bfamdl04)

*KAT: *antes de passar aquele creme ele nũ tava assim / espetado pos lados nãõ //*

*SIL: **tava nãõ //**

Exemplo 15 (bfamdl04)

*SIL: *tomou sorvete / Heliana //*

*HRM: **tomei //**

Exemplo 16 (bfamcv02)

*RUT: *<vou ser obrigada> / né //*

*TER: **pois é / uai //**

Exemplo 17 (bfamcv03)

*CEL: *aqui o catorze bem morto lá agora // o' p' cê ver //*

O verbo não é o núcleo do sintagma em 324 ocorrências (7%).

Exemplo 18 (bfamdl01)

*FLA: *cê nũ quer comprar um trenzinho que espirra pro seu banheiro nãõ //*

*REN: **trenzim que espirra //**

Exemplo 19 (bfamdl04)

*SIL: *colocar esse aparelho do Tommaso / mesma coisa que tar num Big Brogher //*

Exemplo 20 (bpubdl02)

*EUG: *de que tipo de sandália cê quer //*

*JAN: **uma sandália que dê pra eu usar de dia e de noite //**

Exemplo 21 (bfamcv04)

*HEL: *pra / nenhum <objeto> / que tenha //*

Uma visualização mais clara dos dados relativos apenas aos enunciados com verbo é apresentada no Gráfico 3. Na maioria dos enunciados com verbo, este efetivamente representa o núcleo do enunciado. Contudo, a proporção de enunciados com verbo, mas nos quais este não apresenta função verbal, é muito maior em conversações e diálogos, e muito mais representativa em enunciados simples do que em enunciados complexos.

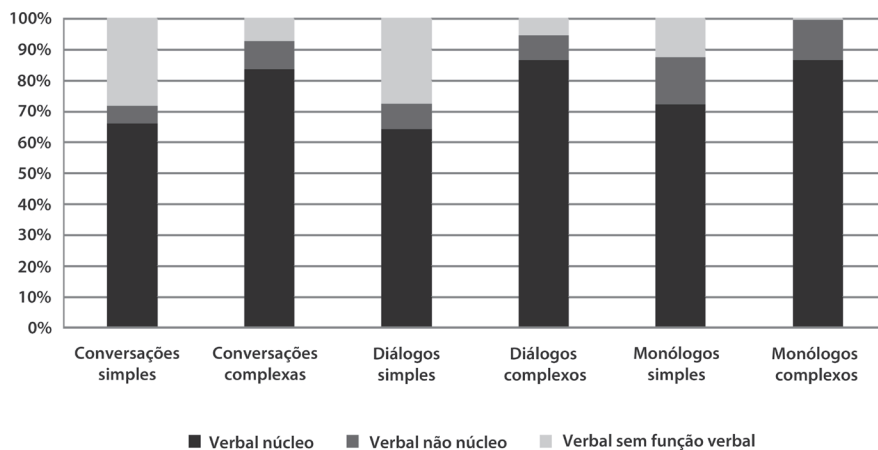


Gráfico 3 - Enunciados simples e complexos com verbo no *subcorpus*

Novamente, percebe-se que conversações e diálogos apresentam uma estruturação de enunciados muito semelhante, tanto na composição dos enunciados simples quanto dos complexos. Nos textos monológicos, a proporção de enunciados verbais em que o verbo não é núcleo é mais significativa do que a proporção de enunciados em que o verbo não cumpre, efetivamente, uma função de verbo.

5. Conjunções coordenativas e subordinativas na fala e na escrita

Como mostramos nas seções anteriores, a presença de um núcleo verbal no enunciado indica a formação de estruturas oracionais, revelando diferentes graus de complexidade estrutural da fala. Outra forma de acessar tais características é observar outros itens lexicais que indicam a formação de estruturas oracionais complexas: as conjunções coordenativas e subordinativas. Analisamos aqui a ocorrência das conjunções mais frequentes da fala, quais sejam, os itens coordenantes *e* e *mas*, e os subordinantes *que* e *porque*.

Primeiramente, vejamos a frequência de ocorrência desses itens nas seções de fala e escrita do *Corpus* do Português, comparando a ocorrência das conjunções *e*, *mas*, *que* e *porque* na fala e na escrita em língua portuguesa nas variedades europeia e brasileira (Tabela 13), passando em seguida para a análise comparativa da fala espontânea (C-ORAL-BRASIL) e na escrita do PB (*corpus* NILC/São Carlos, sem as ocorrências da seção CETEMFolha – JOCF) (Tabela 14).

Tabela 13
 Frequência das conjunções *e*, *mas*, *que* e *porque* nas seções escrita e falada do *Corpus* do Português, proporção do total de palavras da amostra

Seção do <i>corpus</i>	Escrita/Acadêmico		Escrita/Jornalístico		Escrita/Ficção		Fala	
Tamanho (palavras)	5.754.506		6.488.917		5.937.484		2.083.296	
Coordenantes	175.147	3,04%	169.895	2,62%	215.821	3,63%	68.581	3,3%
<i>e</i>	166.965	2,90%	150.913	2,33%	185.929	3,13%	53.024	2,5%
<i>mas</i>	8.182	0,14%	18.982	0,29%	29.892	0,50%	15.557	0,7%
Subordinantes	99.412	1,73%	159.546	2,46%	189.640	3,19%	88.853	4,3%
<i>que</i>	97.940	1,70%	154.159	2,38%	181.901	3,06%	80.845	3,9%
<i>porque</i>	1.472	0,03%	5.387	0,08%	7.739	0,13%	8.008	0,4%

Fonte: Davies; Ferreira (2006).

De modo geral, pode-se verificar a recorrência de estratégias distintas na fala e na escrita, e mesmo diferenças entre gêneros de escrita diversos. O texto acadêmico mostra a importância da coordenação (3,04%) na escrita, enquanto a subordinação é bem menos utilizada (1,73%). Nos demais gêneros, a coordenação é sempre mais utilizada, embora com uma diferença que parece pouco significativa. Na fala, a estratégia que aparece como mais importante é a subordinação (4,3%).

Na amostra de PB, confirmamos a mesma tendência observada no *Corpus* do Português para a diamesia falada. A principal estratégia na fala espontânea representada no C-ORAL-BRASIL é a subordinação (4,33%). Já a escrita (*corpus* NILC/São Carlos) mostra a mesma tendência observada nos gêneros Jornalístico e Ficção do *Corpus* do Português, com uma pequena prevalência da coordenação sobre a subordinação, mas que não chega a ser significativa.

Tabela 14
 Frequência das conjunções *e*, *mas*, *que* e *porque* no C-ORAL-BRASIL (fala) e no NILC/São Carlos (escrita), proporção do total de palavras

<i>Corpus e</i>	C-ORAL-BRASIL		NILC/São Carlos*	
Tamanho (palavras)	208.130		7.176.452	
Coordenantes	5.457	2,62%	154.810	2,16%
<i>e</i>	3.771	1,81%	142.132	1,98%
<i>mas</i>	1.686	0,81%	12.678	0,18%
Subordinantes	9.017	4,33%	146.359	2,04%
<i>que</i>	7.889	3,79%	142.130	1,98%
<i>porque</i>	1.128	0,54%	4.229	0,06%

*Sem dados da seção CETEMFolha.
 Fonte: Linguateca, 2011.

Realizamos ainda um levantamento da distribuição dessas conjunções no C-ORAL-BRASIL em relação ao número de enunciados, para realizar uma comparação com os dados do C-ORAL-ROM. Os resultados são apresentados na Tabela 15.

Tabela 15
 Percentual de ocorrência das conjunções *e*, *mas*, *que* e *porque* por enunciado no C-ORAL-BRASIL e no C-ORAL-ROM

Conjunção	PB	PE*	IT*	ES*	FR*
e (<i>e</i> , <i>y</i> , <i>et</i>)	10%	27%	17%	29%	20%
mas (<i>ma</i> , <i>pero</i> , <i>mais</i>)	5%	7%	6%	8%	8%
que (<i>che</i> , <i>que</i> , <i>que</i>)	22%	34%	20%	22%	18%
porque (apenas PB)	3%	-	-	-	-

*Fonte: Cresti, 2005b, p. 239; valores aproximados.

Notamos que o PE tem uma proporção bem maior de todas as conjunções analisadas do que o PB. É muito provável que isso seja devido às características do *corpus* do PE, que tem maior predominância de enunciados com alta elaboração textual, com maior proporção de enunciados complexos (ver Tabela 6), que são também mais longos e contêm mais unidades tonais (ver Tabela 10). Esses fatores devem gerar uma maior frequência de conjunções por enunciado.

A frequência da conjunção *que* em PB não é diretamente comparável à frequência da mesma conjunção em italiano, francês e espanhol. Primeiramente, é preciso considerar que uma porcentagem significativa de *que* é devida a locuções das quais essa conjunção é um dos componentes (*quando que*, *que que* etc.). Com isso, a frequência do *que* em PB tende a ser bem maior se comparada com as outras línguas. Em segundo lugar, em PB (e também em PE e ES) há a locução modal *ter que* (*tener que*), que não existe em IT e FR.

No *corpus* do PB, as conjunções investigadas são menos frequentes do que nos *corpora* das línguas que compõem o C-ORAL-ROM. É bem provável que uma das causas para tal seja a alta frequência de enunciados simples com relação ao C-ORAL-ROM (ver Tabela 6), além de o PB ter também uma média baixa de palavras por enunciado (ver Tabela 10).

Esse raciocínio, aparentemente, não se aplica se compararmos o PB e o francês, já que esta última língua apresenta mais enunciados simples do que o PB. Porém, se olharmos para o tamanho médio das unidades tonais e enunciados em número de palavras (Tabela 10), observamos que a média do francês é de 4,97 palavras por unidade tonal e 14,49 palavras por enunciado, enquanto a média de palavras por unidade tonal do PB é de 3,37, e por enunciado é de 6,16. Uma unidade tonal com alto número de palavras tem muito mais chance de originar estruturas complexas (como coordenação e subordinação), para as quais se faz necessário o uso de conjunções.

Os dados mostrados na Tabela 15 indicam a presença de coordenação em pelo menos 15% dos enunciados do PB, e de subordinação em pelo menos 25% dos enunciados. Entretanto, é preciso ressaltar que a simples ocorrência desses itens na fala não é suficiente para compreender de que modo esses operadores textuais atuam na composição dos enunciados e que funções executam. Analisando dados de fala espontânea, Cresti (2005b) atesta que, na fala, as conjunções muitas vezes não funcionam como operadores lógicos como acontece na escrita, especialmente as coordenativas *e* e *mas*. Para ilustrar o uso comum dessas conjunções na escrita, vejam-se os exemplos 22 e 23 extraídos do NILC/São Carlos:

Exemplo 22 (NILC/São Carlos)

par = 5415: As células “jogadas” para fora diferenciam-se e acabam constituindo um tecido de revestimento adulto denominado súber.

Exemplo 23 (NILC/São Carlos)

*par = 44588: Pode-se dizer que a ciência cognitiva tem um passado muito extenso **mas** uma história relativamente curta.*

Diferentemente do que ocorre na escrita, na fala espontânea, é muito comum o uso dessas conjunções com funções pragmáticas, como a abertura do turno ou a ligação entre enunciados, e, portanto, atos de fala diferentes. O valor linguístico desses operadores sintáticos irá variar de acordo com a posição que ocupam no enunciado. Um item lexical pode ocupar qualquer uma das três seguintes posições dentro do enunciado (Cresti, 2005b):

1. *posição inicial*: o item ocorre logo depois de uma quebra prosódica terminal ou logo no início de turno dialógico. Essa posição está relacionada a funções pragmáticas típicas da fala, como a abertura do

turno ou a ligação entre atos de fala (expressos pelos enunciados); a função, portanto, é a de marcar o começo de um turno (típica função pragmática) ou de ligar unidades autônomas (os enunciados), o que é uma função muito diferentes da coordenação sintática através de operadores lógicos;

2. *posição articulada*: o item ocorre imediatamente após uma fronteira prosódica de tipo não terminal; também as conjunções nessa posição não devem ser confundidas com os operadores lógicos; em caso contrário, se perderia a diferença funcional entre conjunções nessa posição e na posição interna à unidade tonal / informacional: conectar unidades informacionais ou conectar sintagmas. Nesta posição as conjunções *e* e *mas* normalmente mantêm funções sintáticas semelhantes às da escrita. Já as conjunções *que* e *porque* parecem exercer principalmente a função de conector de informação não essencial (conectar adjuntos ou parentéticos ao Comentário);
3. *posição linearizada*: o item faz fronteira com outras palavras, no meio de uma unidade tonal. Neste caso as conjunções coordenativas normalmente são parte de construções formulaicas, enquanto as conjunções *que* e *porque* cumprem função realmente de subordinação sintática.

Vejam-se exemplos das conjunções em cada uma das posições. No exemplo 24 temos a conjunção *porque* em posição inicial. Em 25 a conjunção *mas* está em posição articulada, e em 26 a conjunção *e* está em posição linearizada.

Exemplo 24 (bfamdl31)

*MAR: só *que* tem *que* ser uns vaso pesado // tipo aquele quadradinho lá / né //

*LIQ: é / uai //

*MAR: **porque** vaso pequenininho cai //

Exemplo 25 (bfammn12)

*JUN: a gente toca letra própria / **mas** também faz muito cover / né //

Exemplo 26 (bfammn03)

*ALO: recebi o dinheiro **e** pronto //

Percebe-se, através desses exemplos, que o uso das conjunções na fala é diferente daquele encontrado na escrita. No exemplo 24, a conjunção *porque* está em início de turno; neste caso, não está indicando relação de

subordinação sintática, mas está operando como um conector pragmático, introduzindo o turno ao mesmo tempo que o liga ao enunciado anterior.

De modo a verificar com maior detalhe como se dá a distribuição das conjunções *e*, *mas*, *que* e *porque*, controlamos a posição em que esses itens ocorrem nos enunciados do C-ORAL-BRASIL. Foram controladas as seguintes possibilidades quanto à posição da conjunção: (a) início de turno; (b) início de enunciado; (c) início de unidade tonal; e (d) meio de unidade tonal. Também verificamos as ocorrências de conjunção em unidade tonal dedicada, ou seja, quando o conteúdo locutivo da unidade tonal corresponde somente à conjunção, como mostrado no exemplo 27.

Exemplo 27 (*bpubdl10*)

*ADA: *eu / realmente eu acredito / que / o professor / ele tem que ter noções de [/ 1] desses outros fatores que vão influenciar a sala de aula //*

Para as conjunções, a posição em unidade tonal dedicada corresponde à função informacional de unidade dialógica (veja-se o Capítulo 3), ou seja, o que em outras tradições é chamado de marcador discursivo. Mais uma vez, a prosódia determina o valor funcional de uma unidade.

A análise excluiu os itens que estivessem envolvidos em fenômenos de *retracting* e os que estavam em enunciados interrompidos. Os resultados gerais estão sumarizados na Tabela 16. Na Tabela 17 está a frequência de ocorrência de cada conjunção em unidade tonal dedicada, conforme a posição no enunciado.

Tabela 16
Ocorrência das conjunções *e*, *mas*, *que* e *porque*,
conforme a posição no enunciado no C-ORAL-BRASIL

Posição da conjunção	<i>e</i>		<i>mas</i>		<i>que</i>		<i>porque</i>	
início de turno (posição inicial)	417	11,6%	442	28,4%	353	4,7%	164	15,7%
início de enunciado (posição inicial)	692	19,3%	476	30,6%	438	5,9%	345	33,0%
início de unidade tonal (posição articulada)	1.619	45,1%	612	39,4%	1.274	17,0%	352	33,7%
meio de unidade tonal (posição linearizada)	859	23,9%	24	1,5%	5.414	72,4%	184	17,6%
Total	3.587	100%	1.554	100%	7.479	100%	1.045	100%

Tabela 17
Ocorrência das conjunções *e*, *mas*, *que* e *porque* em unidade tonal dedicada conforme a posição no enunciado no C-ORAL-BRASIL

Posição da unidade	<i>e</i>		<i>mas</i>		<i>que</i>		<i>porque</i>	
em início de turno	39	10,7%	44	18,2%	11	10,6%	33	14,7%
em início de enunciado	119	32,5%	86	35,5%	17	16,3%	103	46,0%
no meio do enunciado	208	56,8%	112	46,3%	77	74,0%	88	39,3%
Total	366/3.587	10,2%	242/1.554	15,6%	105/7.479	1,4%	224/1.045	21,4%

Nas seções seguintes, discutimos os resultados para cada uma das conjunções separadamente.

5.1 A conjunção *e*

A conjunção *e* apresenta uma importante função pragmática de conectar turnos e atos de fala (quando em posição inicial de turno ou enunciado) em cerca de 30% das ocorrências: 11,6% em início de enunciado e 19,3% em início de turno. Neste último caso, é muitas vezes utilizada para introduzir ou reforçar um argumento.

Exemplo 28 (*bfamdl27*)

*LUA: *ô Gigi / serviço de pedreiro tá pagando tão bem / quem sabe cê nã muda de ramo hhh //*

*LIA: *oh hhh // não / mas caleja muito a mão de mulher // se nã fosse isso / minha filha / tava dentro //*

*LUA: ***e se nã fosse pela beleza / né / Gigi hhh //***

*LIA: *é //*

*LUA: *se nã fosse +*

*LIA: *ter que ficar &a [/1] cum [/1] ca mão / toda / cortada / Deus me livre //*

*LUA: *se nã fosse a questão estética //*

*LIA: ***e mãos que fazem maquiagem nã pode pegar serviço muito pesado não //***

A posição articulada é aquela que parece ser mais relevante para o uso do *e*. Esta conjunção ocorre em posição articulada (no início de unidade tonal, mas não de enunciado) em 45% dos enunciados. Estes resultados

corroboram os encontrados para a língua italiana, na qual o *e* está em posição articulada em cerca de 43% dos casos (Cresti, 2005b), mantendo uma função semelhante à que desempenha na escrita (coordenação copulativa).

Exemplo 29 (bfamcv20)

*CAD: *deu ele pra adoção / aí ele foi levado por esses jesuítas / foi / criado nos Estados Unidos / formou em medicina / e virou cirurgião plástico um dos mais famosos de Hollywood //*

A posição linearizada é a menos frequente (23,9%) e normalmente está associada a expressões formulaicas e construções fixas, apresentando então função composicional específica (lexical).

Exemplo 30 (bfamcv02)

*RUT: *portanto / eu nã vou / oferecer mundos e fundos não //*

Exemplo 31 (bfamcv04)

*HEL: *esse é novo // eu [/1] eu conheço o velho Imagem e Ação //*

No PB, também parece ser muito comum o uso da conjunção *e* em posição linearizada para compor construções verbais com valor aspectual.

Exemplo 32 (bfammn01)

*MAI: *quando chegou lá na frente / e' viu que ela ia pegar ele mesmo / ele [/1] ele tirou / uma capa // como é que chama o nomeda capa / cê falou aquele dia comigo //*

*DUD: *pelerine //*

*MAI: *é // tipo assim / o nome da capa // e' pegou e jogou //*

Exemplo 33 (bfamcv27)

*RAL: *quando eu descartei achei que nunca ia sair pra mim // aí foi e saiu // uhn //*

Em 10,2% das ocorrências da conjunção *e*, esta encontra-se em uma unidade tonal dedicada. Essa posição é típica, quanto às quatro conjunções em exame, da função de unidade dialógica. Neste caso, normalmente o *e* tem a função informacional de Conector Discursivo e não apresenta função de coordenação copulativa oracional, mas sim a de conectar unidades de Comentários Ligados em uma Estrofe, ou ligar subpadrões informacionais dentro de uma Estrofe, ou, naturalmente, de ligar enunciados: são todos casos em que a função de base acaba sendo aquela de conectar atos de fala.

Exemplo 34 (bfammn09)

*MAU: *ai / eu fui / me internaram / e / na época / eu tinha um médico / lá em Caxambu / que chamava doutor Luís Carlos / uma pessoa até muito boa / um médico bom da época / que ele cuidava de criança //*

Quando ocorre com essa função informacional, percebemos que o *e* normalmente é utilizado em posição medial (56,8%), o que indica que essa conjunção muito provavelmente conecta Comentários Ligados dentro de uma Estrofe.¹⁴

5.2 A conjunção *mas*

A conjunção *mas* aparece predominantemente em posição inicial, de turno (28,4%) e de enunciado (30,6%), totalizando 58% das ocorrências nessa posição. Também neste caso os dados do PB confirmam as observações de Cresti (2005b) para o italiano. Quando usada na abertura do enunciado, essa conjunção tem uma função argumentativa, marcando um contraste total ou parcial com a argumentação do interlocutor.

Exemplo 35 (bfamcv08)

*REN: *é // eu sabia que na hora que ela começou a pegar os trem / ela não lembrava / tudo que ela tinha passado desde o Ice //*

*BRU: *inda falei assim // fecha o caixa / e começa de novo //*

*REN: *é //*

*FLA: ***mas ela nã ia passar tudo de novo / também //*** *nã ia não //*

Em posição articulada (39,4%), *mas* está ligado à estratégia de estruturação textual e mantém a função adversativa (*mas* muitas vezes parcial).

Exemplo 36 (bfamcv08)

*DEB: *o ano passado eu não trabalhei com esse projeto / não trabalhei porque / exatamente porque o devedê estragou / na época / eu fiquei / muito brava / porque eu gostaria que a escola / providenciasse o devedê / não providenciou / não foi possível / né / **mas esse ano a minha postura foi diferente** / falei / não / eu acho que / meu papel de professora / social / né / cidadã / e eu tenho que fazer isso //*

¹⁴ Estrofes são entidades linguísticas concluídas, mas que, diferentemente dos enunciados, apresentam mais de uma unidade nuclear, chamadas de Comentários Ligados. Veja-se o Capítulo 3.

A conjunção *mas* é a menos recorrente em posição linearizada, apenas 1,5% das ocorrências. Nessa posição, *mas* atua com função adversativa, em construções como a exemplificada em 37.

Exemplo 37 (*bfamfl35*)

*LIA: *ea mora lá pertinho //*

*LEO: *apesar que / **mora perto mas nũ é muito pertinho não** //*

A conjunção *mas* também é utilizada na fala como em unidade tonal dedicada (15,6%), principalmente em posição inicial de turno (18,2%) e de enunciado (35,5%), apresentando a função de Conector Discursivo, fazendo a ligação entre Comentários ou ente subpadrões em uma Estrofe.

Exemplo 38 (*bfammn24*)

*MIC: *o [/1] o método de ensino // bom // pra área específica que é o curso / diz que é um curso excelente // &he / me parece/ se eu não me engano / é relações públicas //*

*BRU: *ah / tá //*

*MIC: *que seria um curso bom / também // **mas** / igual eu tava te falando / essa / &meni [/2] essa mulher [/2] essa prima de Esmeralda / ela / fez publicidade / e ela falando comigo que ela sempre / teve vontade de &s [/1] ser / uma coisa na vida //*

5.3 A conjunção *que*

Que é a principal conjunção utilizada em construções subordinadas, e, como vimos, é muitíssimo utilizada na língua falada. Ainda assim, essa conjunção nem sempre apresenta a função sintática de subordinante, havendo casos em que seu valor pragmático é evidente.

Exemplo 39 (*bpubcv04*)

*ANT: *ô William // xá eu te falar / tô querendo fluoxetina // verdade// vinte &fi [/1] miligramas //*

*WIL: *vinte miligramas //*

*ANT: *é //*

*WIL: *tá //*

*ANT: *e outra coisa / Tiane teve aqui hoje //*

*WIL: *hoje //*

*ANT: ***que** ea falou que &vi [/1] ia / vim aqui pra te pagar / e tal //*

A conjunção *que* ocorre em posição inicial em 10,6% das ocorrências, abrindo turnos (4,7%) e enunciados (5,9%). Nessa posição, *que* tem principalmente uma função interrogativa (seja sozinho ou em locuções interrogativas), como mostrado no exemplo 40. Também é bastante comum o uso do *que* inicial (mas não somente) em expressões exclamativas (exemplo 41).

Exemplo 40 (*bfamcv13*)

*MNV: *é // foi // a perna começou a doer / né // começou a doer / eu falei / uai // **que** negócio é esse //*

Exemplo 41 (*bfamcv33*)

*HEL: ***que** lindo //*

*BAO: *Nossa / **que** lindo isso / gente //*

Quando a conjunção *que* está em posição articulada (17%), esta não funciona como um subordinante, pois não há relação de dominância sintática entre unidades prosódicas distintas. Sua função nestes casos é conectar unidades informacionais justapostas, marcando o encadeamento de ideias. Normalmente introduz unidades com teor explicativo (comumente Parentéticos, mas também unidades de Comentários Ligados e de Apêndice).

É a conjunção que menos ocorre em unidade tonal dedicada (apenas 1,4%). Esses dados indicam que a função principal dessa conjunção na fala é a subordinante ou a de introduzir unidades justapostas, com pouca frequência de usos pragmáticos, ainda que estes também existam, como já exemplificado. O exemplo 42 abaixo mostra estes dois usos do *que*: unidade tonal dedicada (em início de enunciado) e em posição articulada. Note-se que em posição articulada não há subordinação sintática com a unidade precedente.

Exemplo 42 (*bfammmn05*)

*CAR: *porque eu / também se fosse pela mãe / nã levaria não // mais pela avó // e outa também // **que** / quando nós fomos levar o papel do advogada lá pra assinar / **que** a advogada é que mexeu pra mim / ela nã queria assinar //*

A conjunção *que* é a única que aparece com maior frequência na posição linearizada (74% das ocorrências), configurando-se como a principal conjunção subordinativa. Em posição linearizada, a conjunção *que* realiza funções verdadeiramente sintáticas, como a introdução de orações relativas, pseudorrelativas e clivadas, entre outros tipos.

Que é um item extremamente produtivo na fala, participando de muitas estruturas, com diferentes valores. A Tabela 18 abaixo sumariza as principais construções com *que* no C-ORAL-BRASIL e sua respectiva frequência.

Tabela 18
Ocorrência das construções mais comuns com a conjunção *que* no C-ORAL-BRASIL
(valores percentuais relativos ao número total de ocorrências da conjunção)

Seq.	Construção	Freq.	%	Seq.	Construção	Freq.	%
1	ter que	617	8,25%	15	como que	34	0,46%
2	que que	406	5,43%	16	pra/para que	32	0,43%
3	do que	238	3,18%	17	assim que	29	0,39%
4	como é que	223	2,98%	18	que é que	28	0,37%
5	só que	143	1,91%	19	mesmo que	21	0,28%
6	que (exclamação)	131	1,75%	20	quanto que	18	0,24%
7	de que	101	1,35%	21	que nem	15	0,20%
8	por que	88	1,18%	22	em que	12	0,16%
9	qual que	64	0,86%	23	se bem que	9	0,12%
10	na hora que	62	0,83%	24	na época que	7	0,09%
11	por que que	57	0,76%	25	sempre que	6	0,08%
12	depois que	52	0,70%	26	bem que	4	0,05%
13	quem que	43	0,57%	27	ainda que	4	0,05%
14	que (interrogação)	41	0,55%	28	ainda bem que	4	0,05%

5.4 A conjunção *porque*

A conjunção *porque* ocorre especialmente em posição inicial (48,7%), abrindo o turno (15,7%) ou, principalmente, o enunciado (33%). Não introduz oração subordinada, mas frequentemente tem função de dar apoio à argumentação.

Exemplo 43 (*bfammn28*)

*REG: *mamãe desceu do carro também / já ia / eu / mãe / segura o João // **porque eu pus a mão** / minha filha // e eu senti o João nascendo //*

A posição articulada é a segunda mais relevante no uso da conjunção *porque*, com 33,7% das ocorrências.

Exemplo 44 (bfamdl05)

*ANE: &he / essa localização aqui / eu nũ acho ruim //

*RAQ: eu gostei também foi por causa do construído já ao redor / né //

*ANE: é //

*RAQ: é sim //

*ANE: tá tudo construído //

*CES: Anete // ele [/1] e' disse que é &d + final de novembro // é / **porque agora é rápido** / isso aqui //

O uso como conjunção subordinativa é o menos importante. *Porque* está em posição linearizada em 17,6% das ocorrências (184 casos). É relevante o fato de que, nesta posição, 100 ocorrências (54%) equivalem à expressão *é porque*. Destas, 56 ocorrências estão em posição inicial de turno (29 ocorrências) e de enunciado (27 ocorrências), 29 estão em posição articulada e 15 ocorrências de *é porque* estão em posição linearizada. Outras nove (5%) são ocorrências da expressão *mas é porque*, três em início de turno e seis em início de unidade tonal. Excluindo-se estes casos, restam apenas 75 ocorrências (41% do total de *porque* linearizado) da conjunção *porque* com a função subordinativa que se assemelha à da escrita.

Exemplo 45 (bfammn23)

*MEL: e / e aí / &he / a gente continuou essa aula / foi / por causa da indisciplina hhh / deu pra fazer assim mas cê [/1] pra você fazer uma coisa / pequena / pouca coisa / **gasta já a aula inteira porque eles não prestam atenção** / é uma bagunça / nts / é //

A conjunção *porque* é a mais utilizada com função pragmática, conforme já demonstrado por sua alta frequência em posição inicial, mas também por ser a mais frequente em unidade tonal dedicada (21,4%), cumprindo uma função de Conector Discursivo ou de Fático.

Exemplo 46 (bfamcv13)

*ATA: Mané / o' / eu fiquei feliz de encontrar [/2] ocê encontrou / dois gente boa aí / com cê / no quarto / batendo papo //

*MNV: não / é / uai //

*ATA: **porque / a pessoa ficar sozinha é ruim / né //**

Os dados apresentados corroboram em muitos aspectos as conclusões de Cresti (2005b) para a língua italiana falada, que mostrou o predomínio das conjunções coordenativas em posições iniciais de turno e enunciado e o predomínio da conjunção subordinativa *que* em posições linearizadas. A análise interlinguística confirma que a segmentação prosódica atua como nível estruturante da informação na fala e revela aspectos comuns entre as línguas românicas analisadas.

6. Considerações finais

Neste capítulo, apresentamos uma comparação da fala espontânea do PB a partir do *corpus* C-ORAL-BRASIL com a diamesia escrita, além de uma comparação com as línguas europeias representadas no C-ORAL-ROM. A comparação envolvendo o uso de verbos e nomes revelou algumas particularidades do PB, indicando uma prevalência muito maior de verbos na fala brasileira do que a observada nas demais línguas comparadas (português europeu, italiano, espanhol e francês).

As medidas relacionadas ao tamanho e à composição dos enunciados (simples e complexos) demonstraram a importância do enunciado como entidade de referência da fala. Além disso, são indicadores importantes do grau de interatividade das situações gravadas, que se reflete em uma grande acionalidade nos textos, o que deve ter como reflexo uma grande quantidade e variedade de valores ilocucionários nos textos do C-ORAL-BRASIL.

Principalmente, apresentamos aqui evidências em favor da função estruturante da prosódia em seu papel de segmentar o contínuo da fala em unidades de ação e de informação, possibilitando desse modo que a linguagem cumpra sua função comunicativa. Consideramos fundamental que os estudos originados a partir do C-ORAL-BRASIL (e possivelmente em outros recursos) levem em consideração a segmentação da fala em enunciados e unidades tonais, tendo em vista que essa segmentação é o recurso natural de envelopamento das estruturas linguísticas em outros níveis.

Anexo

Características dos textos do *subcorpus*
do C-ORAL-BRASIL etiquetado informacionalmente

(Continua)

Texto	Situação	Informantes		Duração h:min:s	Nº de palavras
		M	F		
Total		28	27	03:58:36	31.318
Conversações					
bfamcv01	amigos avaliam um campeonato de futebol organizado por eles e planejam o próximo	4	0	00:07:00	1.467
bfamcv02	senhoras conversam sobre os preparativos do casamento de uma parente	0	3	00:07:51	1.725
bfamcv03	amigos jogam sinuca	5	0	00:06:50	1.390
bfamcv04	amigos jogam "Imagem e Ação", após explicar as regras do jogo para uma das participantes	2	2	00:07:30	1.766
bpubcv01	funcionários de banco de sangue explicam como o sangue coletado é armazenado*	1	3	00:08:30	1.798
bpubcv02	reunião ordinária em uma sede regional de partido político	3	1	00:29:47	1.628
Diálogos					
bfamd101	colegas de apartamento fazem as compras do mês*	0	2	00:14:39	2.131
bfamd102	colegas de faculdade batem papo enquanto organizam o material de gravação	1	1	00:07:26	1.572

Texto	Situação	Informantes		Duração	Nº de (Conclusão) palavras
		M	F		
bfamd03	casal faz uma viagem de carro*	1	1	00:10:30	1.637
bfamd04	domésticas, mãe e filha, fazem a limpeza da cozinha após o almoço*	0	2	00:19:32	1.249
bfamd05	corretor de imóveis leva a irmã para visitar apartamento*	1	1	00:11:28	1.736
bpubd01	engenheiro e pedreiro trabalham em uma obra	2	0	00:26:08	1.568
bpubd02	cliente e vendedor interagem durante a compra de calçados*	1	1	00:15:45	1.438
M o n ó l o g o s					
bfamnn01	senhor narra história fantástica sobre uma cobra	7	10	01:05:40	10.213
bfamnn02	sobrinha de Carlos Drummond de Andrade conta histórias da família ao neto	2	0	00:05:02	1.086
bfamnn03	narrativa de “causos” divertidos para a família*	1	1	00:07:23	1.677
bfamnn04	senhora conta sua experiência no hospital após ter dado à luz no carro*	3	3	00:07:08	1.206
bfamnn05	senhora fala sobre a adoção da filha após a morte de sua filha biológica*	0	1	00:06:57	1.450
bfamnn06	pai conta seu percurso profissional à sua filha	0	2	00:09:52	1.580
bpubnn01	entrevista de avaliação sobre aulas de inglês na rede pública de ensino	1	1	00:10:02	1.600
		0	2	00:19:16	1.614

* presença de pequenas intervenções de um ou mais informantes que estavam fora da situação.