

Interface entre sintaxe e articulação informacional na fala espontânea: uma comparação baseada em *corpus* entre português brasileiro e italiano

Syntax-informational patterning interface in spontaneous speech: a corpus-based comparison between Brazilian Portuguese and Italian

Giulia Bossaglia

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brasil.
giulia.bossaglia@gmail.com

Resumo: Conforme a Teoria da Língua em Ato (CRESTI, 2000a), que fundamenta este artigo, a sintaxe no estudo da fala espontânea necessita ser estudada considerando a dimensão da articulação informacional, e, portanto, de forma diferente de como é tradicionalmente feita a análise sintática da escrita. Com o objetivo geral de aprofundar o estudo da sintaxe da fala e com base em *corpora* especificamente construídos para uma adequada análise desta diamesia, é aqui apresentado o primeiro mapeamento das orações completivas no *corpus* de fala de português brasileiro C-ORAL-BRASIL (RASO; MELLO, 2012), a fim de comparar esta primeira análise quantitativa com os dados disponíveis extraídos do *corpus* C-ORAL-ROM (CRESTI; MONEGLIA, 2005) para fala de italiano. A análise qualitativa das orações completivas nas duas línguas, conduzida sobre dois *subcorpora* do C-ORAL-BRASIL e do C-ORAL-ROM anotados informacionalmente, é apresentada em seguida.

Palavras-chave: Teoria da Língua em Ato; orações completivas; sintaxe da fala espontânea; português; italiano.

Abstract: According to Language into Act Theory (CRESTI, 2000a), on which this paper is founded, the syntax of spontaneous speech can't be analyzed without taking into account the informational patterning

dimension. Its analysis, therefore, must be different from the traditional one applied to written language. With the general aim of examining in depth spontaneous speech syntax and on the basis of adequate spontaneous speech *corpora*, this paper shows the first mapping of completive clauses in spoken Brazilian Portuguese *corpus* C-ORAL-BRASIL (RASO; MELLO, 2012), in order to make a quantitative analysis comparable with available data from spoken Italian C-ORAL-ROM *corpus* (CRESTI; MONEGLIA, 2005). The qualitative analysis of completive clauses in both spoken languages is then presented, conducted on two *sub-corpora* of C-ORAL-BRASIL and Italian C-ORAL-ROM presenting informational patterning annotation.

Keywords: Language into Act Theory; completive clauses; spontaneous speech syntax; Portuguese; Italian.

Recebido em 30 de agosto de 2014.

Aprovado em 18 de novembro de 2014.

1 A Teoria da Língua em Ato

Este estudo sobre sintaxe da fala espontânea fundamenta-se na Teoria da Língua em Ato, doravante, TLA (CRESTI, 2000a; RASO, 2012; MONEGLIA; RASO, 2014). Esta teoria destaca-se de outras abordagens do estudo da fala em vários aspectos teóricos e metodológicos, inclusive na própria concepção de sintaxe da fala, entendida como estritamente ligada à dimensão da articulação do fluxo da fala em unidades informacionais.

1.2 TLA: unidades de referência da fala

A TLA foi desenvolvida por Cresti (1995, 2000a, 2000b) como extensão da Teoria dos Atos de Fala de Austin (1962) e caracteriza-se por uma sólida verificação empírica com base nos *corpora* do LABLITA - *Laboratorio Linguistico del Dipartimento di Italianistica dell'Università di Firenze* (MONEGLIA, 2005), que levou mais tarde à construção dos *corpora* de fala espontânea de italiano, francês, espanhol e português europeu C-ORAL-ROM (CRESTI; MONEGLIA, 2005).

A característica mais inovadora da TLA é a inclusão do elemento prosódico na análise da fala espontânea, sendo que durante muito tempo, pelo

contrário, esta diamesia tem sido estudada apenas com base em transcrições e, portanto, a partir de categorias mais adequadas ao estudo da escrita.

É precisamente com base em critérios prosódicos que a TLA individualiza a unidade de referência da fala no enunciado (CRESTI, 2000a, 2000b), e não na cláusula (HALLIDAY, 1989; MILLER; WEINERT, 1998), ou na predicação com núcleo verbal (VOGHERA, 1992), ou na sentença (CHAFE; DANIELEWICZ, 1987).

De acordo com Austin (1962), o enunciado corresponde a um ato de fala, ou seja, a um ato locutivo (transmissão de material linguístico) que possui força ilocutiva (a ação que é realizada através do conteúdo linguístico transmitido).

O valor ilocutivo do enunciado é veiculado pela prosódia, que pode conferir autonomia pragmática a qualquer tipo de conteúdo locutivo, independentemente da sua autonomia semântica ou sintática, como se pode observar escutando os exemplos (1) a (3) extraídos do *corpus* C-ORAL-BRASIL:¹

(1) *TER: com a dona Deise //²

(2) *HEL:ahn //

(3) *JOR: e é um caso interessante nesse mercado /

Os perfis prosódicos de (1) e (2) (um sintagma preposicional e uma interjeição) contrastam com o de (3) (uma frase semântica e sintaticamente completa) por serem interpretáveis como unidades prosódica e pragmaticamente completas, ou seja, como enunciados. Pelo contrário, (3) não pode receber a mesma interpretação e não funciona como enunciado, constituindo uma parte não autônoma dele, como se mostra em (4):

(4) *JOR: e é um caso interessante nesse mercado / que muito deles me convidavam pra ser sócio dele //

(5) *JOR: que muito deles me convidavam pra ser sócio dele //

¹ Devido à importância dos fatores prosódicos para o estudo da sintaxe da fala, para cada exemplo disponibilizam-se os arquivos áudio online como documento complementar na página da *Caligrama* com o rótulo *audio.zip*.

² Cada exemplo extraído do C-ORAL-BRASIL é reportado incluindo a sigla de três letras (precedida por um asterisco) usada no *corpus* para indicar o falante.

De fato, no enunciado em (4), é a parte reportada em (5) que possui autonomia pragmática, pois carrega o núcleo prosódico da ilocução (é a unidade de comentário, veja *infra*).

Portanto, a TLA assume a existência de uma Interface Prosódica que (i) veicula o valor ilocutivo do enunciado, isto é, a autonomia pragmática dele, mas também (ii) sinaliza a divisão do fluxo da fala em enunciados, e até divisões internas a eles, como em (4) (CRESTI, 2000a; CRESTI, 2014; MONEGLIA, 2011). Os enunciados são segmentados com base em quebras percebidas como conclusivas (quebras terminais, “//” na transcrição), mas também podem ser segmentados internamente por quebras percebidas como não conclusivas (quebras não terminais, “/”).³ Define-se simples um enunciado formado por uma única unidade (como (1), (2)), que é necessariamente a de Comentário (COM), pois é aquela que carrega a ilocução, sem a qual não se daria a autonomia pragmática do enunciado. Definem-se como complexos os enunciados formados pela unidade de COM mais uma ou mais unidades tonais de outro tipo, como (4).

As unidades tonais diferentes de COM possuem valores informacionais que são também veiculados prosodicamente: de acordo com o modelo IPO (*Institute of Perception Research*) da Universidade de Eindhoven (HART; COLLIER; COHEN, 1990), a TLA assume que a f_0 (frequência fundamental) de um enunciado possui uma série de movimentos dentro das unidades. Entre estes, uns são voluntários (embora inconscientes) e têm relevância perceptiva, carregando determinadas funções informacionais, conforme a *Information Patterning Theory* (CRESTI, 1995; MONEGLIA; CRESTI, 2006; CRESTI; MONEGLIA, 2010; MELLO; PANUNZI; RASO, 2011).⁴

Por exemplo, o perfil prosódico da primeira unidade que compõe o enunciado complexo em (4) é o da unidade de Tópico (TOP), que tem a função de definir o campo de aplicação da ilocução veiculada pelo COM em virtude de uma relação de *aboutness* pragmática que existe entre as duas unidades.

Portanto, com base em critérios prosódicos e pragmáticos, a TLA conseguiu ampliar o repertório de unidades pertinentes ao estudo

³ A segmentação (assim como a transcrição) do fluxo da fala no *corpus* C-ORAL-BRASIL passou por vários processos de validação estatística interna, a fim de garantir a consistência da anotação das quebras pelos transcritores (MELLO et al., 2012).

⁴ Para além dos movimentos de f_0 , outras variáveis como duração, intensidade e alinhamento silábico ao andamento de f_0 são tomadas em conta na análise prosódica das unidades de fala.

da fala, tradicionalmente limitadas àquelas de tópico e comentário (CHAFE, 1976; HALLIDAY, 1989; KRIFKA, 2007). Com efeito, foram individualizadas várias unidades dotadas de funções, características prosódicas e distribuição próprias dentro do enunciado. Estas se dividem em (i) unidades *textuais*, que compõem o texto do enunciado ou, ainda que não o componham, são dirigidas à interpretação dele; e (ii) unidades *dialógicas*, que não compõem o texto do enunciado e são dirigidas ao interlocutor, tendo função de regular a interação entre os falantes. No Quadro 1 abaixo (adaptado de TUCCI, 2011, p. 90) ilustram-se sinteticamente as unidades informacionais individualizadas pela TLA:

Quadro 1 – Unidades informacionais segundo a TLA

	Tipo	Tag	Função
Unidades Textuais	Comentário	COM	Veicula a força ilocucionária do enunciado.
	Tópico	TOP	Estabelece o âmbito de aplicação da força ilocucionária.
	A p ê n d i c e d e comentário / tópico	APC/APT	Integra textualmente a unidade da qual é apêndice.
	Parentético	PAR	Dá instruções sobre como deve ser interpretado o enunciado ou parte dele.
	Introdutor locutivo	INT	Sinaliza que o que segue tem nível herárquico diferente daquele da enunciação (geralmente, é uma metailocução).
Unidades Dialógicas	Alocutivo	ALL	Individualiza o interlocutor, marca coesão social.
	Incipitário	INP	Sinaliza o começo do turno ou do enunciado.
	Expressivo	EXP	Fornece suporte emotivo para o ato de fala, marca coesão social.
	Conector discursivo	DCT	Sinaliza continuidade de uma sequência com a anterior.
	Fático	PHA	Sinaliza a abertura ou a manutenção do canal comunicativo.
	Conativo	CNT	Induz o interlocutor a cumprir ou desistir de certa ação.

Para além destas, é necessário mencionar também outros tipos de unidades que são importantes para o estudo da fala: unidades de escansão, comentários múltiplos e comentários ligados.

As unidades de escansão (SCA) são partes tonais de uma mesma unidade informacional que não pode ser realizada como unidade única devido ao tamanho do conteúdo locutivo, ou a razões de tipo expressivo ou de imperícia na fala (crianças, falantes de baixa diastratia):

(6) *TER: do lado da mãe /=SCA= da Fafica /=TOP= tudo pobrezim
//=COM=

Os comentários múltiplos (CMM) são unidades de comentário que apresentam entre si uma relação retórica particular e que são interpretadas de forma holística em virtude de uma específica padronização prosódica. A relação lógica entre os CMM pode ser de vários tipos (temporal, hipotética, de consequência, entre outros):⁵

(7) *CAR: se ele nũ morrer /=CMM= nũ tem problema não //COM=

Finalmente, os comentários ligados (COB) são comentários caracterizados por ilocuções homogêneas (da mesma classe) e enfraquecidas, não padronizadas mas simplesmente justapostas, no processo de o falante dar forma a seu pensamento. Os COB formam estrofes (CRESTI, 2009), ou seja, macro-enunciados, sequências terminadas de dimensão muito dilatada e por isso típicas de textos monológicos (onde a acionalidade diminui, em prol de um maior peso da semântica):

(8) *JOR: e assim eu fiquei dentro dessa outra multinacional por um período /=COB= trabalhando com som automotivo /=COB= &he /=TMT=⁶ produtos automotivos da rede de autopeça /=COB= hhh e ferramentas elétrica //COM=

Como se mostra na seguinte seção, a noção de sintaxe dentro do paradigma da TLA está estritamente ligada à estruturação do fluxo da fala em unidades informacionais.

⁵ O mesmo fenômeno pode acontecer com unidades de Tópico: é o caso das Listas de Tópicos (TPL) descritas em Mittmann (2012).

⁶ O tag TMT refere-se à tomada de tempo, ou seja, hesitação do falante.

1.3 TLA: sintaxe da fala espontânea

De acordo com a TLA, relações sintáticas e semânticas *stricto sensu* têm validade apenas dentro das unidades textuais, que são consideradas “ilhas”: o *output* final do enunciado resulta da combinação de tais ilhas e sua sintaxe corresponde à combinação das frases, dos sintagmas e até dos fragmentos que o podem compor (SCARANO, 2003; CRESTI, 2000a; BLANCHE-BENVENISTE, 2000; CRESTI, 2014; CRESTI; MONEGLIA, 2010). A relação entre as ilhas não é sintática: as unidades são ligadas umas às outras por relações informacionais específicas, veiculadas por específicos perfis prosódicos. Por exemplo, já se mencionou que o TOP define o campo de aplicação da ilocução veiculada pelo COM segundo uma relação de *aboutness* pragmática existente entre as duas unidades. Graças a um perfil prosódico específico, o enunciado é percebido como completo ainda que não haja uma relação estritamente sintática entre TOP e COM, como em (9), onde *o aluno* em TOP não pode ser o sujeito sintático do predicado em COM (papel já preenchido por *ele*):⁷

(9) *SHE: e o aluno /=TOP= ele é extremamente carente //COM=

A noção de sintaxe é analisada pela TLA de acordo com sua distribuição dentro do contínuo da fala, isto é, das ilhas/unidades textuais que o compõem. A TLA distingue, portanto, entre sintaxe (i) *linearizada*: estruturas de subordinação e coordenação próprias, dentro da mesma unidade informacional, como em (10); e (ii) *padronizada*: estruturas de subordinação e coordenação realizadas em mais do que uma unidade informacional (veja (11)), ou seja, ao longo de várias ilhas, cada uma das quais desenvolve uma função informacional diferente.⁸

(10) *PAU: tô achando que eu vou fazer ela com um metro //COM=

(11) *PAU: acho que com um metro e vinte /=TOP= ela fica boa
/=COM= né //PHA=

⁷ Mas também em ausência de predicação verbal um enunciado pode ser percebido como completo, cf. [6] na seção anterior.

⁸ Cresti (2014); Miller e Weinert (1998, p. 22) falam de *integrated* vs. *fragmented/unintegrated syntax*. Nos termos de Blanche-Benveniste (2000), a distinção entre sintaxe linearizada e padronizada corresponde à distinção entre *micro-* e *macro-syntax*.

Em (10) e (11) a completiva objetiva é realizada respectivamente em configuração linearizada e padronizada. Em (11), o predicado “principal” é o da oração completiva, porque carrega a função informacional de COM, enquanto o *verbum putandi*, que de acordo com uma análise puramente sintática seria o núcleo da regência, adquire função de TOP, ou seja, torna-se informação de *background* necessária para a definição do campo de aplicação do COM (cf. Quadro 1). Esta aquisição de função pragmática pela estrutura subordinativa apresentada em (11) não ocorre em (10), onde oração principal e subordinada são realizadas dentro da mesma unidade, mantendo sua relação sintática direta. Portanto, nos casos de sintaxe padronizada, a função das estruturas coordenativas e subordinativas é orientada pela organização pragmática (articulação informacional) do enunciado, que é sempre veiculada, lembre-se, pela prosódia. Cresti (2014) demonstra que testes de dependência sintática tradicionais (clivagem, coordenação negativa, coordenação afirmativa depois de resposta afirmativa) não resultam ou acarretam um *output* não natural, quando aplicados a estruturas de sintaxe padronizada.⁹

O mapeamento das orações completivas no *corpus* C-ORAL-BRASIL apresentado nas seções seguintes foi feito considerando a distinção entre estruturas linearizadas e padronizadas, ou seja, a interface entre sintaxe e articulação informacional, assim como é entendida no paradigma da TLA.

2 Orações completivas: uma comparação entre português e italiano falados

A escolha das orações completivas como objeto de estudo deste trabalho sobre sintaxe da fala espontânea deve-se à importância de tais estruturas sintáticas como estratégias subordinativas nas línguas românicas. Neste trabalho analisaram-se apenas as completivas explícitas, em duas etapas distintas: (i) uma análise quantitativa de tais orações no *corpus* C-ORAL-BRASIL I (seção informal; RASO; MELLO, 2012); (ii) uma análise qualitativa num *subcorpus* do C-ORAL-BRASIL anotado informacionalmente. Os resultados da etapa (i) foram comparados com os de estudos existentes (CRESTI, 2005, 2014) sobre as mesmas orações no italiano falado (IT), com base no respectivo *corpus* C-ORAL-ROM (CRESTI; MONEGLIA, 2005). A comparação dos resultados obtidos

⁹ Não nos é possível aprofundar o assunto nesta sede, para os detalhes consulte-se Cresti (2014, p. 393-400).

para o português brasileiro (PB) na etapa (ii) com os de italiano foi feita através de uma análise específica sobre um *subcorpus* do C-ORAL-ROM de italiano anotado informacionalmente (PANUNZI; MITTMANN, 2014) e comparável com o *subcorpus* utilizado do C-ORAL-BRASIL (CRESTI; RASO, 2012).

2.1 As completivas no *corpus* C-ORAL-BRASIL

2.1.2 O *corpus* C-ORAL-BRASIL

O C-ORAL-BRASIL (RASO; MELLO, 2012) é um *corpus* de fala espontânea informal de português brasileiro,¹⁰ representado na diatopia mineira, principalmente da região metropolitana de Belo Horizonte. O *corpus* é estruturado de forma a ser comparável com os *corpora* do projeto C-ORAL-ROM (CRESTI; MONEGLIA, 2005) para italiano, espanhol, francês e português europeu.

O C-ORAL-BRASIL é um *corpus* balanceado e representativo das principais tipologias interacionais da fala. Com efeito, os 139 textos das gravações que o compõem (com um total de 208.130 palavras) pertencem a dois contextos - familiar/privado (149.364 palavras) e público (48.766) - e são divididos, em quantidades aproximadamente iguais, entre monólogos, diálogos e conversações (ou seja, diálogos com mais de dois participantes). Gravações de alta qualidade acústica foram transcritas de acordo com o modelo CHAT (MACWHINNEY, 2000) implementado para a anotação das fronteiras prosódicas (MONEGLIA; CRESTI, 1997): o *corpus*, conforme os pressupostos da TLA, é segmentado em enunciados (34.167 no total) e unidades tonais. O C-ORAL-BRASIL apresenta, ainda, alinhamento texto-som através do *software* WinPitch (MARTIN, 2004), com o qual também é feita a análise prosódica dos textos, o que faz dele uma ferramenta realmente adequada ao estudo da fala, pois está sempre disponível o acesso simultâneo ao texto, ao áudio e ao espectrograma.

O *corpus* foi construído procurando alcançar um alto grau de variação diafásica, para que fosse realmente representativo da grande variedade situacional que caracteriza a fala espontânea: para este fim, foram diversificados significativamente os contextos situacionais das

¹⁰ A seção formal do C-ORAL-BRASIL encontra-se em fase de compilação.

gravações, que por sua vez têm um número médio de palavras (1.500) que garante a autonomia textual da interação (BIBER, 1993; RASO, 2012). O C-ORAL-BRASIL apresenta apenas anotação morfossintática através do *parser* PALAVRAS (BICK, 2000). Atualmente, é dotado de anotação informacional um *subcorpus* dele, que foi utilizado na segunda etapa deste trabalho e que é descrito na seção 2.2.1.

2.1.3 Coleta e tratamento dos dados

O mapeamento das completivas foi feito através da plataforma de busca *online* Db-CoM C-ORAL-BRASIL (MELLO et al., em preparação), onde se encontra o *corpus* anotado morfossintaticamente, junto com os arquivos áudio de cada enunciado. Foram procuradas orações subjetivas, objetivas e interrogativas indiretas explícitas. Portanto, a coleta dos dados foi realizada buscando as ocorrências de vários complementizadores: em primeiro lugar *que*, principal conjunção subordinativa na fala do PB (RASO; MITTMANN, 2012, p. 206). Com relação às interrogativas indiretas, foram buscadas as ocorrências de *se*, *como*, *porque*, *quando*, *onde* e construções interrogativas do tipo de *que que*, *o que que*, *por que que*, *qual que*, *quanto que*, *como que*, etc.¹¹ As conjunções mencionadas foram buscadas através dos filtros “lemma” ou “palavra em contexto” mais o atributo PoS de conjunção subordinativa.

Os dados assim recolhidos foram limpos manualmente para excluir ocorrências dos complementizadores com valores não completivos, e casos de subordinação aparente, como em (12) abaixo:

(12) *JAN: é muito comum a gente ver que / pô / na casa da minha avó tinha isso //

Somente *escutando* o enunciado em (12) pode-se observar que o verbo *ver*, apesar da presença do complementizador *que*, não está introduzindo uma subordinada objetiva, mas um discurso reportado, ou seja, uma reprodução mimética do que é falado por outra pessoa. Discurso reportado, exemplificação emblemática e pensamento falado

¹¹ Esta é a transcrição utilizada para as correspondentes formas *que é que*, *o que é que*, *quando é que*, etc. quando o verbo é não é pronunciado pelo falante (MELLO et al., 2012, p. 143).

são todas instâncias de subordinação aparente (MAIA-ROCHA; RASO, 2011; CRESTI, 2014).

Uma vez recolhidas as ocorrências das orações em estudo, os dados foram organizados separando os enunciados contendo orações completivas em configuração linearizada (dentro da mesma unidade textual) daqueles contendo estruturas subordinativas padronizadas (ao longo de duas ou mais unidades textuais).

Decidiu-se ainda separar os casos onde a configuração sintática fosse padronizada por quebras devidas a *retracting*, ou seja, reformulação (na transcrição: “[/n]”, onde *n* = número de palavras canceladas pelo falante; *tag*: EMP = *empty*). A padronização por *retracting* não é uma organização das unidades informacionais voluntária, mas um “acidente” devido ao fato de a fala ser espontânea, ou seja, planejada ao mesmo tempo em que é realizada. Dividiram-se, porém, os casos onde o enunciado apresenta apenas padronização por *retracting* (que chamaremos de *retracting* linearizados, veja (13)), dos que apresentam padronização por *retracting* mais padronização informacional (*retracting* padronizados, veja (14)):

- (13) *LUZ: porque /=DCT= eu só soube que eu ã [/6]=EMP= eu tive certeza absoluta que eu ã era daqui quando eu saí // =COM=
 (14) *LUA: é /=INP= eu acho que /=SCA= uma coisa também que ajudaria¹² nisso /=TOP= é [/1]=EMP= era se as [/1]=EMP= as coordenações nas escolas fossem melhor /=COM= né // =PHA=

2.1.4. Resultados

2.1.4.1 Linearização vs. padronização

A busca das completivas explícitas no C-ORAL-BRASIL levou a um total de 1.324 enunciados, representando aproximadamente 4% sobre o total do *corpus*.¹³ Excluindo os enunciados com *retracting*, a proporção entre ocorrências de completivas linearizadas e padronizadas

¹² A falante pronuncia efetivamente “ajudaria” em lugar de “ajudaria”. A transcrição registra este tipo de fenômenos.

¹³ De acordo com Cresti (2014), a subordinação explícita constitui aproximadamente 10% sobre o total de enunciados do *corpus* italiano. Desse valor, 7% é representado por completivas e relativas, e 3% por subordinadas adverbiais.

é, respectivamente, de 77,2% (919 enunciados) vs. 22,8% (272). Dentre as ocorrências com *retracting*, a proporção entre linearizadas e padronizadas é similar: 72,6% (96) vs. 27,4% (37). Considerando todas as ocorrências de completivas linearizadas e padronizadas incluindo aquelas com *retracting*, a proporção entre linearização e padronização das completivas fica em valores parecidos: 76,7% (1.015) linearizadas vs. 23,3% (309) padronizadas. O italiano também apresenta uma maior incidência de linearização por volta de valores percentuais parecidos: 72%, de acordo com Cresti (2014). As orações subjetivas, por um total de 90 ocorrências, constituem 6,8% de todas as completivas, a maioria sendo representada por orações objetivas e interrogativas indiretas.

2.1.4.2 Complementizadores

Com relação aos complementizadores, *que* é o mais utilizado (1.061 enunciados),¹⁴ seguido por *se* (148). O grupo restante de complementizadores é composto pelas expressões interrogativas do tipo “*X que*” (*por que que, quanto que, qual que, como que, o que que, etc.*) que apresentam ampla variedade com relação à base lexical (115).

Contudo, é necessário considerar que o C-ORAL-BRASIL não é anotado informacionalmente. Isto significa que, neste primeiro mapeamento das completivas, no grupo das padronizadas poderiam estar também completivas em configuração padronizada aparente, ou seja, devido à escansão (cf. seção 1.2 acima): sendo que as unidades de SCA são partes da mesma unidade informacional, uma completiva distribuída ao longo de mais unidades deste tipo é de *se* considerar, de fato, linearizada. Uma quantificação mais refinada de completivas linearizadas e padronizadas foi feita, para português e italiano, nos dois *subcorpora* do C-ORAL-BRASIL e C-ORAL-ROM italiano anotados informacionalmente.

¹⁴ O complementizador correspondente em italiano, *che*, foi o único analisado por Cresti (2014) para completivas explícitas. Em português, considerando os enunciados em que as completivas são introduzidas só pelo complementizador *que*, porém, a proporção entre linearização e padronização fica praticamente inalterada: 77,5% (812) vs. 23,5% (249).

2.2 As completivas nos *minicorpora* do C-ORAL-BRASIL e do C-ORAL-ROM italiano

2.2.1 Os dois *minicorpora* do PB e do IT da plataforma DB-IPIC

Os *minicorpora* do PB e do IT extraídos dos respectivos *corpora* C-ORAL-BRASIL e C-ORAL-ROM encontram-se disponíveis online na plataforma DB-IPIC (*Database for Information Patterning Interlinguistic Comparison C-ORAL-ROM*) do LABLITA¹⁵. Os dois *minicorpora* são compostos por 20 gravações cada, e foram construídos de forma a preservar a mesma estrutura dos *corpora* de referência no que diz respeito a (i) proporções de cada tipologia interacional (diálogos e conversações: 2/3; monólogos: 1/3), (ii) alta qualidade acústica, (iii) variação diafásica e diversificação dos falantes, e (iv) pressupostos teóricos da TLA, ou seja, os critérios de segmentação em unidades de fala e a anotação informacional. Isto garante que eles sejam realmente representativos dos *corpora* de referência. Além disso, os dois *minicorpora* são comparáveis um com o outro (GREGORI; PANUNZI, 2011, 2012; MITTMANN; RASO, 2011; RASO; MITTMANN, 2012; PANUNZI; MITTMANN, 2014), pois, apesar da leve diferença em termos de números de palavras, eles são perfeitamente equilibrados em termos de unidades de referência da fala, como se mostra no Quadro 2¹⁶:

Quadro 2 – Os *minicorpora* DB-IPIC

	<i>Minicorpus</i> PB	<i>Minicorpus</i> IT
Gravações	20	20
Palavras	29.909	32.589
Seqüências terminadas (enunciados e estrofes)	5.511	5.663
Enunciados	5.045	5.117

Além da transcrição, os *corpora* da plataforma incluem os arquivos de áudio, os metadados (informação sociolinguística e dados das gravações) de cada sessão, o alinhamento texto-som e a anotação PoS. No

¹⁵ Disponível em: <<http://lablita.dit.unifi.it/app/dbipic/>>.

¹⁶ Adaptado dos quadros disponíveis em: <http://lablita.dit.unifi.it/ipic/index_html>.

momento, o *minicorpus* do PB não possui anotação PoS na plataforma, mas somente no DVD que reúne todos os textos do C-ORAL-BRASIL¹⁷. Os textos dos *minicorpora* apresentam anotações manuais que pertencem aos níveis (1) da segmentação do fluxo da fala em enunciados e unidades tonais e (2) da análise do valor informacional de cada unidade tonal (cf. os *tags* no Quadro 1), de acordo com os pressupostos da TLA e da Teoria da Padronização da Informação.

2.2.2 Coleta e tratamento dos dados

O mesmo processo de busca das ocorrências dos complementizadores mencionados na seção 2.1.3 foi utilizado para a coleta dos dados nos dois *minicorpora*, através da interface de busca disponível na própria plataforma DB-IPIC, para o levantamento das completivas explícitas. Para italiano, além de *che*, principal conjunção subordinativa no italiano falado (CRESTI, 2005), procuraram-se os complementizadores interrogativos *se*, *quanto*, *come*, *perché*, *dove*, *quale*, etc., de forma a cobrir uma gama o mais possível variada e completa como para o PB. Antes de serem quantificados, também nesta segunda fase os resultados da busca passaram por um processo de limpeza manual, a fim de excluir eventuais dados não pertinentes (cf. seção 2.1.3).

2.2.3 Resultados

2.2.3.1 Linearização vs. padronização

A busca das completivas nos *minicorpora* levou a um total de 211 enunciados no PB, e 164 no IT. Se as proporções entre linearização e padronização no *corpus* C-ORAL-BRASIL podiam ser consideradas parecidas com as detectadas por Cresti (2014; cf. seção 2.1.4.1) no C-ORAL-ROM italiano, o mesmo não se pode observar depois de uma análise mais refinada sobre os *minicorpora* anotados. Com efeito, no PB as completivas linearizadas representam 83,9% (177 enunciados) do total, e as padronizadas apenas 16,1% (34). Por sua vez, no IT as linearizadas continuam representando a maioria sobre o total, porém com menos distância percentual com relação às padronizadas: 67,3%

¹⁷ Disponível em: <<http://www.c-oral-brasil.org/>>

(111) vs. 32,7% (54). No *minicorpus* de italiano, dentre as linearizadas, destacamos um pequeno grupo (9) de enunciados onde as completivas aparecem ao longo de unidades interrompidas por unidades de PAR, como se mostra no exemplo (15):

- (15) *GCM: te /=SCA= dirai di volta in volta /=i-COM= o di modulo in modulo /=PAR= quale portare //COM= [Trad.: *você /=SCA= dirá a cada vez /=i-COM= ou a cada módulo /=PAR= qual é de levar //COM=]*

Em (15), o comentário é interrompido (i-COM) por uma unidade de Parentético que se insere entre o predicado principal *dirai* e a interrogativa indireta *quale portare*. A função metalinguística do PAR (cf. Quadro 1) coloca ele num nível diferente do da ilocução do enunciado, e, portanto, o PAR não participa do padrão informacional, como demonstrado também pelo fato que a eliminação desta unidade através da edição do som não gera impressão de corte (TUCCI, 2004; RASO, 2012, p. 108). Por esta razão as 9 ocorrências de completivas realizadas ao longo de uma unidade interrompida por PAR foram computadas no grupo das linearizadas.

Como esperado, as orações subjetivas representam uma minoria à frente das objetivas e interrogativas indiretas: no *minicorpus* do PB há apenas 20 ocorrências, constituindo 9,5% sobre o total dos enunciados, mas no do IT as orações subjetivas são 31 (18,9% sobre o total).

2.2.3.2 Complementizadores

Como era esperado, nas duas línguas o complementizador mais utilizado é *que/che*: no PB representa 83,9% (177) sobre o total, no IT, 77,4% (120). O segundo, em termos de frequência, é o interrogativo *se*, seja no PB (11,4%, 24 ocorrências) seja no IT (13,5%, 21). No PB, as interrogativas indiretas introduzidas por expressões interrogativas do tipo “*X que*” são representadas em apenas 10 ocorrências. Também no IT os complementizadores atestados no *minicorpus* *perché*, *come*, *dove*, *quanto*, e *che* (interrogativo) usados na formação de interrogativas indiretas são os menos frequentes, constituindo o restante grupo de resultados (14 ocorrências).

2.2.3.3 Articulação informacional¹⁸

2.2.3.3.1 Completivas linearizadas

A configuração linearizada, como vimos, é a mais utilizada para as completivas, no PB e no IT. Em ambas as línguas, as completivas linearizadas aparecem de preferência em unidades de COM: no PB, 167 sobre o total de 177 linearizadas; no IT, 88 sobre 102. Outras unidades nas quais aparecem completivas são as de Tópico (PB: 5; IT: 5), Parentético (PB: 3; IT: 7), e Introdutor Locutivo (PB: 2; IT:1). Uma ocorrência de completiva linearizada em Apêndice de COM foi encontrada no *minicorpus* de italiano (cf. seção 2.2.3.3.2.1). É interessante notar que as unidades mencionadas (COM, TOP, PAR, INT) são as que podem receber modalização, e não é um caso que as completivas sejam uma das estratégias de modalização, isto é, de expressão do ponto de vista do falante sobre o conteúdo linguístico que está transmitindo, mais utilizadas na fala.¹⁹

2.2.3.3.2 Completivas padronizadas

À frente da maior incidência de estruturas sintáticas padronizadas em italiano, observa-se alguma homogeneidade com relação aos tipos de padrões informacionais detectados em ambos os *minicorpora*. No PB, os padrões mais utilizados são: combinações de CMM ou COB (vamos por comodidade referi-las com a abreviatura “CMM/CMM – COB/COB” 11 ocorrências) e TOP/COM (9), que é o principal padrão informacional da fala (MITTMANN, 2012). Seguem os padrões INT/COM (5), COM/DCT/COM (4), INT/TOP/COM (2), apenas uma ocorrência de PAR/PAR, uma de COM/APC e uma de TOP/INT.²⁰

¹⁸ Os *tags* da anotação informacional, que são usados também como abreviaturas dos nomes das unidades informacionais, são ilustrados no Quadro 1 acima.

¹⁹ Sobre o conceito de modalidade na fala espontânea, veja Mello, Carvalho e Côrtes. (2010), Ávila e Mello (2013), Ávila (2014).

²⁰ Pode ser que unidades dialógicas apareçam dentre os padrões informacionais. Uma vez que, porém, que elas não compõem o texto do enunciado, os padrões são referidos sem considerar eventuais unidades dialógicas presentes. Também é o caso, contudo, dos padrões com unidades interrompidas pela unidade textual de PAR, cf. exemplo (15).

Os padrões mais frequentes no *minicorpus* de italiano são, como para o PB, TOP/COM (21) e CMM/CMM – COB/COB, ainda que com apenas 9 ocorrências. Padrões que têm certa relevância são também INT/COM (7), INT/TOP/COM (7) e COM/APC (7). Alguns padrões informacionais textualmente mais complexos também são representados, embora apenas por ocorrências isoladas: TOP/INT/COB, TOP/DCT/COM, TOP/APT/COM, INT/TOP/COM/APC.

A maior complexidade dos padrões informacionais nos quais aparecem as completivas em italiano era em parte esperada, sendo que já tem sido verificado que o italiano apresenta maior complexidade textual ao nível micropragmático, ou seja, um maior uso de enunciados complexos e formados por unidades textuais, enquanto o PB faz um uso mais amplo de enunciados simples e de complexos formados, porém, por unidades dialógicas (PANUNZI; MITTMANN, 2014). Portanto, não é casual que apenas no *minicorpus* de italiano foram achados casos de unidades interrompidas, como o do exemplo [15] acima, onde unidades de COM são interrompidas por unidades de PAR.

2.2.3.3.2.1 Algumas observações sobre completivas padronizadas nos dois *minicorpora*

Como já referido, segundo a TLA, nos casos de padronização a sintaxe passa a ser “orientada” pelas funções pragmáticas, isto é, ligadas à ação que o falante cumpre com seu ato de fala: quando distribuídas ao longo de mais unidades informacionais, as estruturas sintáticas deixam de ser interpretáveis de acordo com análises tradicionais de dependência, e assumem funções pragmáticas específicas. Cresti (2005) observou que, além da padronização prosódica, podem-se individualizar algumas particularidades com relação à posição dos índices lexicais da subordinação, ou seja, no caso em estudo, dos complementizadores. Passamos, portanto, à ilustração de alguns padrões informacionais interessantes achados nos *minicorpora*.

TOP/COM

No padrão TOP/COM, o principal da fala (CRESTI, 1999, 2003; CRESTI; FIRENZUOLI, 1999; FIRENZUOLI; SIGNORINI, 2003; CRESTI; MONEGLIA, 2010; ROCHA et al., 2012; MITTMANN, 2012), a unidade de TOP define o campo de aplicação da ilocução veiculada

pela unidade de COM. As completivas que aparecem neste padrão informacional, seja no PB seja no IT, apresentam um tipo de distribuição dos índices regular: verbo da oração principal e complementizador em TOP, verbo da completiva em COM. Neste tipo de padrão, que Cresti (2014) chama de “Theme in Topic”, o núcleo da regência sintática passa à função pragmática de providenciar informação de *background* (Tema) para o predicado dependente em COM, que vira “principal” de um ponto de vista pragmático, sendo que é o que carrega a ilocução, logo a autonomia, do enunciado. Veja:

(16) *MAR: quindi vuol dire che /=TOP= uno era sbagliato //COM=
[Trad.: *então significa que /=TOP= umestava errado //COM=]*

(17) *LUZ: porque eu acho que no mesmo concurso /=TOP= cê nũ
pode fazer duas //COM=

COM/APC

Uma situação oposta àquela do padrão TOP/COM é representada por COM/APC: as unidades de Apêndice servem de integração textual às unidades de referência (que podem ser apenas a de TOP e a de COM). Nestes padrões, seja em italiano seja em português, o verbo principal está em COM,²¹ e complementizador e predicado da completiva em APC:

(18) *DAN: diglielo /=COM= che hanno fatto i cavallini //APC=
[Trad.: *fala para ele /=COM= o que fizeram os cavalinhos //APC=]*

(19) *MAI: nũ sei /=COB= né /=PHA= a imaginação hhh dum
[1]=SCA= dum animal /=COM= o que que pode ser /=APC=
né //PHA=

Nestes casos, o predicado principal carrega o valor ilocutivo, e a completiva serve apenas de integração textual. As unidades de Apêndice, de fato, não podem se considerar “ilhas” independentes como as outras unidades textuais (RASO; ULISSES, 2008; CRESTI, 2014).²²

²¹ No exemplo (19), a oração principal é distribuída ao longo de um COB e do COM, ou seja, é uma configuração levemente mais articulada do padrão COM/APC.

²² As unidades de Apêndice não podem ser modalizadas, e têm um estatuto especial

INT/COM e INT/TOP/COM

Em ambas as línguas, as completivas que aparecem no padrão INT/(TOP/)COM apresentam dois tipos de distribuição: (i) verbo principal e complementizador em INT, predicado da completiva em COM; ou (ii) verbo principal em INT, complementizador e predicado da completiva em COM.

- (20) *ELA: *perché quando tu racconti /=INT= che c' era questi americani /=CMM= che vi facevano la corte //CMM=* [Trad.: *porque quando você conta /=INT= que havia estes americanos /=CMM= que cortejavam vocês //COM*]
- (21) *TAM: *credo che /=INT= al di fuori /=TOP= si dovrebbe ricreare /=COB= dare queste possibilità //COM=* [Trad.: *acho que /=INT= do lado de fora /=TOP= se deveria recriar /=COB= dar estas possibilidades //COM=*]
- (22) *SHE: *vendo assim /=INT= que tem /=SCA= condição //COM=*
- (23) *SHE: *então /=INP= eu vejo que /=INT= é [1] /=EMP= querendo ou não /=TOP= gente /=ALL= a organização /=TOP= é a alma do negócio //COM=*

CMM/CMM-COB/COB

Nas completivas que aparecem em combinações de CMM ou COB, em ambas as línguas, a distribuição dos índices lexicais de subordinação é: verbo principal numa unidade, complementizador e completiva na outra. Nestes casos não é posta ênfase particular sobre uma unidade; pelo contrário, as unidades que compõem o padrão ou (1) se combinam para veicular a ilocução (é o caso dos CMM), ou (2) se justapõem, acrescentando ou reforçando uma ilocução, com um andamento parecido ao de um elenco (são os COB). Veja:

- (24) *TAM: *perché i giovani /=TOP= hanno voglia di capire /=COB= come avviene la politica /=COB= come avviene la gestione della cosa pubblica //COM=* [Trad.: *porque os jovens /=TOP= têm*

que não podemos aprofundar nesta sede. Veja, porém, o trabalho de Raso e Ulisses (2008) e bibliografia aí referida.

vontade de entender /=COB= como acontece a política /=COB= como acontece a gestão da coisa pública //COM=]

- (25) *OTT: se non condividono SR /=TOP= devono capire /=SCA= che quando si relazionano alla tre F /=CMM= si relazionano alla tre F //COM= [Trad.: *mas se não concordam com SR /=TOP= têm que compreender /=SCA= que quando se relacionam com a “três F” /=CMM= se relacionam à “três F” //COM=]*
- (26) *FLA: é /=CMM= me falaram /=CMM= que ele é muito <bom> //CMM=
- (27) *SIL: falou que a idade é assim mesmo /=COB= que ela vai ter que conviver com essas dores /=COB= e pronto //COM=

3 Conclusões

A partir dos resultados das duas etapas deste trabalho, podem-se fazer algumas generalizações de várias ordens. Em primeiro lugar, observou-se que as orações completivas explícitas são realizadas de preferência, no PB e no IT, em configuração linearizada, ou seja, mantendo o valor de dependência sintática própria. Contudo, observamos no PB uma incidência de estruturas linearizadas significativamente maior que no IT. Isto poderia ter a ver com o fato do PB, sobretudo na sua variedade mineira, que é a que mais está representada no C-ORAL-BRASIL, ser uma língua de ritmo mais marcadamente acentual que o IT (BARBOSA, 2000, p. 395). Isto significa que no PB o ritmo da fala é marcado pelos acentos tônicos e, no IT, pelas sílabas acentuadas: conseqüentemente, dentro de uma unidade tonal cabe mais material fonológico no PB do que no IT. A variedade mineira, além disso, é caracterizada por marcados fenômenos de rressilabificação (contrações, formas aferéticas, etc.) que também contribuem para a redução da quantidade de sílabas fonéticas com relação às fonológicas. Estes fatos poderiam acarretar, a princípio, uma maior incidência de sintaxe linearizada no PB.

A diferença nas proporções entre linearização e padronização das completivas parece ser a principal diferença entre as duas línguas, sendo que, em termos de (i) tipos de complementizadores e suas freqüências relativas, (ii) tipos (objetiva ou subjetiva) de oração, e (iii) tipos de padrões informacionais principais, o português brasileiro e o italiano aparentam certo grau de homogeneidade.

Com relação à articulação informacional, o italiano, como era esperado, apresenta um grau de complexidade textual maior, representado por um maior número e uma maior variedade de padrões informacionais nos quais aparecem as completivas.

Analisando alguns dos principais padrões nas duas línguas, observamos o comportamento dos índices lexicais de subordinação ao longo das unidades envolvidas e ilustramos como a padronização das completivas serve a determinadas estratégias (ênfase de algum elemento informacional, fornecimento de informação de *background*, entre outras) de organização *pragmática*, já não sintática, do fluxo da fala.

Agradecimentos

Este artigo é fruto de uma pesquisa (programa de pós-doutorado – PMPD II, 2014/2015, processo nº 21442) financiada pela FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), à qual agradeço.

Referências

AUSTIN, L.J. *How to do things with words*. Oxford: Oxford University Press, 1962.

ÁVILA, L. *Modalidade em perspectiva: estudo baseado em corpus oral do Português Brasileiro*. 2014. 253 f. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2014. Disponível em <<http://hdl.handle.net/1843/MGSS-9LZMMJ>>. Acesso em 30 ago. 2014.

ÁVILA, L.; MELLO, H. Proposta para um esquema de anotação da modalidade em um minicorpus oral do português brasileiro: desafios preliminares. In: LAPORTE, É.; SMARSARO, A.; VALE, O. de A. (Org.). *Dialogar é preciso: linguística para o processamento de línguas*. Vitória: PPGEL/UFES, 2013. V. 1, p. 31-44.

BARBOSA, P. A. “Syllable-timing in Brazilian Portuguese”: uma crítica a Roy Major. *DELTA*, São Paulo, v. 16, n. 2, p. 369-402, 2000. Disponível em <<http://www.scielo.br/pdf/delta/v16n2/a06v16n2.pdf>>. Acesso em 30 ago. 2014.

BIBER, D. Representativeness in corpus design. *Literary and Linguistic Computing*, Oxford, v. 8, n. 4, p. 243-257, 1993.

BICK, E. *The parsing system "PALAVRAS": automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus: Aarhus University Press, 2000. 412 p. Disponível em: <<http://beta.visl.sdu.dk/~eckhard/pdf/PLP20-amilo.ps.pdf>>. Acesso em 30 ago. 2014.

BLANCHE-BENVENISTE, C. *Approches de la langue parlée en français*. Paris: Ophrys, 2000.

CHAFE, W. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: LI, C. N. (Ed.). *Subject and topic*. New York: Academic Press. 1976. p. 25-55.

CHAFE, W.; DANIELEWICZ, J. Properties of written and spoken language. In: HOROWITZ, R.; SAMUELS, S.J. (Eds.). *Comprehending oral and written language*. New York: Academic Press, 1987. p. 83-113.

CRESTI, E. Speech act units and informational units. *Speech Acts and Linguistic Research: Proceedings of the Workshop*, July 15-17, 1994. New York: Nemo, 1995. p. 89-107.

CRESTI, E. Force illocutoire, articulation topic/comment et contour prosodique en italien parlé. *Faits de langue*, Le Mans, n. 13, p. 168-181, 1999.

CRESTI, E. *Corpus di italiano parlato*. Firenze: Accademia della Crusca, 2000a.

CRESTI, E. Enunciato e frase. In: TROVATO, S. (Ed.). *Atti del V Congresso SILFI*. Catania, 15-17 ottobre 1998. 2000b.

CRESTI, E. Illocution et modalité dans le comment et le topic. In: SCARANO, A. (Ed.). *Macrosyntaxe et pragmatique: l'analyse linguistique de l'oral*. Roma: Bulzoni, 2003. p. 133-182.

CRESTI, E. Enunciato e frase: teoria e verifiche empiriche. In: BIFFI, M.; CALABRESE, O.; SALIBRA, L. (Ed.). *Italia linguistica: discorsi di scritto e di parlato*. Scritti in onore di Giovanni Nencioni. Siena: Prolagon, 2005. p. 249-260.

CRESTI, E. La Stanza: un'unità di costruzione testuale del parlato. X CONGRESSO SILFI, 2009, Basileia. *Atti del X Congresso SILFI: Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione*. Basileia, 30.06-03.07-2008, 2009. p. 1-25.

CRESTI, E. Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses. In: RASO, T.; MELLO, H. (Ed.). *Spoken corpora and linguistic studies*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 2014. p. 365-410.

CRESTI, E.; FIRENZUOLI, V. Illocution et profils intonatifs de l'italien. *Revue Française de Linguistique Appliquée*, Paris, v. IV, n. 2, p. 77-98, 1999.

CRESTI, E.; MONEGLIA, M. (Ed.). C-ORAL-ROM. *Integrated reference corpora for spoken romance languages*. Amsterdam/ Philadelphia: John Benjamins, 2005.

CRESTI, E.; MONEGLIA, M. Informational Patterning Theory and the corpus-based description of spoken language. The compositionality issue in the topic-comment pattern. In MONEGLIA, M.; PANUNZI, A. (Ed.). *Bootstrapping information from corpora in a cross-linguistic perspective*. Firenze: Firenze University Press, 2010. p. 13-46.

CRESTI, E. ; RASO, T. *Text annotation of information units through IPIC*. 2012. <<http://lablita.dit.unifi.it/ipic/>>.

FIRENZUOLI, V.; SIGNORINI, S. L'unità informativa di topic: correlti intonativi. In: MAROTTA, G.; NOCCHI, N. (Ed.). *La coarticolazione: Atti delle XIII Giornate di Studio del Gruppo di Fonetica Sperimentale*, Pisa, 28-30 novembre 2002. Pisa: Edizioni ETS, 2003. p. 177-184.

GREGORI, L.; PANUNZI, A. DB-IPIC: an XML database for the representation of information structure in spoken language. In: MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *Pragmatics and prosody: illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 2011. p. 133-149.

GREGORI, L.; PANUNZI, A. DB-IPIC: an XML database for informational patterning analysis. In: MELLO, H.; PETTORINO, M.; RASO, T. (Ed.). *Speech and corpora: Proceedings of the VIIth GSCP International Conference*. Firenze: Firenze University Press, 2012. p.121-127.

HALLIDAY, M. A. K. *Spoken and written language*. Oxford: Oxford University Press, 1989.

HART, J. T.; COLLIER, R.; COHEN, A. *A perceptual study on intonation: an experimental approach to speech melody*. Cambridge: Cambridge University Press, 1990.

KRIFKA, M. Basic notions of information structure. In: FÉRY, C., FANSELOW, G.; KRIFKA, M. (eds.) *The notions of information structure*. Potsdam: Univ.-Verl., 2007. (Interdisciplinary Studies on Information Structure, 6). p. 13-55. Disponível em <<http://edoc.hu-berlin.de/oa/bookchapters/re08JlHtgchw/PDF/23dHYFKxbQaXY.pdf>>. Acesso em 30 ago. 2014.

MACWHINNEY, B. J. *The CHILDES project: tools for analyzing talk*. 3. ed. Mahwah: Lawrence Erlbaum Associates, 2000. 2 v.

MAIA-ROCHA, B.; RASO, T. A unidade informacional de introdutor locutivo no português do Brasil: uma primeira descrição baseada em corpus. *Domínios de Linguagem*, Uberlândia, v. 5, p. 327-343, 2011. Disponível em <<http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/12479/8070>>. Acesso em 30 ago. 2014.

MARTIN, PH. WinPitch corpus: a text to speech alignment tool for multimodal corpora. In: *Proceedings of the 4th International Conference on Language Resources Evaluation*. Lisbon, 26-28 may 2004, 2004. p. 537-540. Disponível em <<http://www.lrec-conf.org/proceedings/lrec2004/pdf/780.pdf>>. Acesso em 30 ago. 2014.

MELLO, H.; CARVALHO, J. M. R.; CÔRTEZ, P. O. Modalização na fala espontânea do português brasileiro: um primeiro mapeamento de índices morfolexicais. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 18, n. 2, p. 105-133, 2010. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2550/2502>>. Acesso em 30 ago. 2014.

MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *Pragmatics and prosody*. illocution, modality, attitude, information patterning and speech annotation. Firenze: Firenze University Press, 2011.

MELLO, H. et al. Transcrição e segmentação do corpus C-ORAL-BRASIL: critérios de implementação e validação. In: RASO, T.; MELLO, H. (Ed.). *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: Ed. UFMG, 2012. p. 125-176.

MELLO, H. et al. *Plataforma de busca online Db-CoM C-ORAL-BRASIL*. Em preparação.

MILLER, J.; WEINERT, R. *Spontaneous spoken language*. Oxford/New York: Clarendon Press, 1998.

MITTMANN, M. M. *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o tópico no Português do Brasil*. 2012. 248 . Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012. Disponível em < <http://hdl.handle.net/1843/LETR-97YMKT>>. Acesso em 30 ago. 2014.

MITTMANN, M. M.; RASO, T. The C-ORAL-BRASIL informationally tagged minicorpus. In: MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *Pragmatics and prosody*. illocution, modality, attitude, information patterning and speech annotation. Firenze: Firenze University Press, 2011. p. 151-183.

MONEGLIA, M. I corpora dell'italiano parlato di LABLITA: criteri di costituzione, unità di analisi e comparabilità dei dati linguistici orali. In: BURR, E. (Ed.). *Il parlato: teoria - corpora - linguistica dei corpora. Atti del VI convegno SILFI*. Duisburg, 28.06-02.07.2000, 2005. p. 213-232.

MONEGLIA, M. Spoken corpora and pragmatics. *Revista Brasileira de Linguística Aplicada*, Belo Horizonte, v. 11, n. 2, p. 479-519, 2011. Disponível em: <<http://www.scielo.br/pdf/rbla/v11n2/a09v11n2.pdf>>. Acesso em 30 ago. 2014.

MONEGLIA, M.; CRESTI, E. L'intonazione e i criteri di trascrizione del parlato adulto e infantile. In: BORTOLINI, U.; PIZZUTO, E. (Ed.). *Il progetto CHILDES Italia*. Pisa: Del Cerro, 1997. p. 57-90.

MONEGLIA, M.; CRESTI, E. C-ORAL-ROM: Prosodic boundaries for spontaneous speech analysis. In: KAWAGUCHI, Y.; ZAIMA, S.; TAKAGAKI, T. (Ed.). *Spoken language corpus and linguistics informatics*. Amsterdam: John Benjamins, 2006.p. 89-114.

MONEGLIA, M.; RASO, T. Notes on Language into Act Theory (L-Act). In: RASO, T; MELLO, H. (Ed.). *Spoken corpora and linguistic studies*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, 2014. p. 468-494.

PANUNZI, A.; MITTMANN, M. M. The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese. In: RASO, T; MELLO, H. (Ed.). *Spoken corpora and linguistic studies*. Amsterdam/Philadelphia: John Benjamins, 2014. p. 129-150.

RASO, T. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In: RASO, T.; MELLO, H. (Eds.). *C-ORAL-BRASIL I: corpus de referência do*

português brasileiro falado informal. Belo Horizonte: Ed. da UFMG, 2012. p. 91-123.

RASO, T.; MELLO, H. (Ed.). *C-ORAL-BRASIL I*: corpus de referência do português brasileiro falado informal. Belo Horizonte: Ed. da UFMG, 2012.

RASO, T.; MITTMANN, M. M. As medidas da fala. In: RASO, T.; MELLO, H. (Ed.). *C-ORAL-BRASIL I*: corpus de referência do português brasileiro falado informal. Belo Horizonte: Ed. da UFMG, 2012. p. 177-220.

RASO, T.; ULISSES, A. J. Tópico e apêndice no português do Brasil: algumas considerações. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 16, n. 1, p. 247-262, 2008. Disponível em: <<http://periodicos.letras.ufmg.br/index.php/relin/article/view/2489/2441>>. Acesso em 30 ago. 2014.

ROCHA, B. et al. O pronome lembrete e a teoria da língua em ato: novas perspectivas de análise. *Anais do Colóquio Brasileiro de Prosódia da Fala*, v. 1, n. 1, p. 27-31, 2012. Disponível em: <http://www.periodicos.letras.ufmg.br/index.php/anais_coloquio/article/view/1203/1354>. Acesso em 30 ago. 2014.

SCARANO, A. Les constructions de syntaxe segmentée: syntaxe, macro-syntaxe et articulation de l'information. In : _____ (ed.). *Macrosyntaxe et pragmatique : l'analyse linguistique de l'oral*. Roma: Bulzoni, 2003. p. 183- 203.

TUCCI, I. L'inciso: caratteristiche morfosintattiche e intonative in un corpus di riferimento. In: ALBANO LEONI, F.; CUTUGNO, F.; PETTORINO, M.; SAVY, M. (Ed.). *Atti del convegno nazionale "Il parlato italiano"*. Napoli, 13-15 febbraio 2003, Napoli: M. D'Auria Editore, 2004. p.1-14.

TUCCI, I. Illocution and modality in spoken Italian: performing a speech act through words and judging their semantic content. A corpus-based analysis. In: MELLO, H.; PANUNZI, A.; RASO, T. (Ed.). *Pragmatics and prosody. illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 2011. p. 83-109.

VOGHERA, M. *Sintassi e intonazione nell'italiano parlato*. Bologna: Il Mulino, 1992.