

# *Cláusulas Relativas na Fala Espontânea do Português do Brasil: um estudo exploratório baseado no corpus C-ORAL-BRASIL*

RELATIVE CLAUSES IN BRAZILIAN PORTUGUESE SPONTANEOUS SPEECH:  
AN EXPLORATORY STUDY BASED ON THE C-ORAL-BRASIL CORPUS

Cryсна Bonjardim da Silva **Carmo\***  
Heliana **Mello\*\***

**Resumo:** Este artigo sistematiza o resultado de investigações exploratórias sobre as cláusulas relativas encontradas no minicorpus de fala espontânea do português do Brasil (PB), etiquetado informacionalmente, extraído do corpus C-ORAL-BRASIL. A seleção dos dados enfocou a identificação dos enunciados nos quais as cláusulas relativas ocorrem. Para essa tarefa, utilizou-se a plataforma DB – IPIC (PANUNZI; GREGORI, 2011). Em seguida, para triagem das cláusulas relativas, empregou-se a ferramenta computacional *AntConc* (ANTHONY, 2013). Para verificação da estrutura informacional dos enunciados, adotou-se o programa *WinPitch* (MARTIN, 2004). Os resultados encontrados: (i) corroboram, em termos prosódicos, a diferença entre cláusulas relativas restritivas e relativas não restritivas – as primeiras ocorrem linearizadas em uma mesma unidade informacional e as últimas, padronizadas em mais de uma unidade informacional; e (ii) demonstram que no PB, ambas ocorrem preferencialmente na unidade de COM, e mesmo o padrão informacional preferencial das relativas não restritivas é resultado da

---

\* Mestre em Linguística pela Universidade Federal de Juiz de Fora (2005). Doutoranda em Estudos Linguísticos pela Universidade Federal de Minas Gerais (2013). Professor Assistente II\_UNEB. Contato: crysnabonjardimsc@gmail.com.

\*\* Doutora em Linguística pela City University of New York (1997). Professor Titular\_UFGM. Contato: heliana.mello@gmail.com.

combinação com uma unidade informacional de COM. Além disso, o artigo apresenta uma descrição de cláusulas relativas através de um modelo diagramático que permite alcançar: (i) as relações internas entre cláusula relativa e principal; e (ii) a estrutura informacional definidora da estrutura interna dos enunciados – escopo de realização das cláusulas relativas.

**Palavras-chave:** Fala espontânea. Cláusulas relativas. Linguística de Corpus.

**Abstract:** This paper systematizes the results found in an exploratory study about relative clauses found in an informationally tagged, balanced sample extracted from the C-ORAL-BRASIL corpus. The data selection focused on the identification of utterances which contain relative clauses which was carried through the DB-IPIC platform (Panunzi & Gregori, 2011). The sorting out of relative clauses was done with the AntConc computational toolset (Anthony, 2013). In order to verify the informational structure of the chosen utterances, the *WinPitch* software was employed (Martin, 2004). The results found: (i) prosodically corroborate the differences between restrictive and non-restrictive relative clauses, whereby the former occur as linearized structures within one informational unit, whereas the latter are patterned through more than one informational unit; and (ii) demonstrate that in Brazilian Portuguese, both relative clause types prefer the Comment information unit as their default carrier. The paper also presents a description of relative clauses through a diagrammatic representation model that allows showing the relationship between main and relative clause, as well as informational structure within an utterance that stands for the realization scope of relative clauses.

**Keywords:** Spontaneous speech. Relative clauses. Corpus Linguistics.

## Introdução

Este artigo apresenta um estudo prospectivo sobre as cláusulas relativas no âmbito da fala espontânea do português do Brasil. É de se notar que as investigações sobre cláusulas relativas em diversas línguas, tradicionalmente, têm privilegiado a diamesia escrita. Mesmo aqueles estudos que usam dados

de fala, a exemplo daqueles voltados para as línguas indígenas<sup>1</sup>, o fazem majoritariamente através de dados transcritos, a rigor, dados escritos, sem considerar a influência da prosódia na segmentação das unidades linguísticas no fluxo da fala. Contudo, para se estudar verdadeiramente dados orais, é preciso observar, além das informações de natureza estritamente lexicais, sintáticas, morfológicas e semânticas, a informação prosódica, pois é ela o *locus* analítico que determina as unidades da fala e, por consequência, a suborganização de informação sintática, como é o caso das cláusulas relativas. Sendo assim, para a execução de tal tarefa, é preciso um corpus de fala espontânea, arquivado precipuamente para o estudo da fala, e que leve em conta a representatividade diafásica, bem como um quadro teórico-metodológico que permita a adequada exploração das particularidades oferecidas pela organização linguística presente na oralidade. É o que se apresenta a seguir.

## 1 A Teoria da Língua em Ato

A Teoria da Língua em Ato (CRESTI, 2000) é um quadro teórico de base pragmática, herdeira da *Teoria dos Atos de Fala* (AUSTIN, 1962). Dessa forma, reconhece nos *atos de fala*, ações realizadas através do dizer, como a principal função realizada através da linguagem. Como teoria de forte base empírica, foi elaborada a partir da compilação e observação sistemática de *corpora* de fala espontânea ao longo dos últimos 30 anos.

Dentre as propostas inovadoras da Teoria da Língua em Ato (doravante, TLA) está a centralidade da informação prosódica para o estudo da fala. É através das quebras prosódicas que se dão: (i) a identificação linguística do *enunciado* (a menor unidade pragmaticamente autônoma identificável no fluxo da fala); (ii) a especificação e realização dos diversos atos de fala; e (iii) a *estruturação informacional* da fala.

---

<sup>1</sup> Em muitos estudos a preocupação central se restringe ao estudo do *tom* – traço suprasegmental que diz respeito à variação da altura do som de determinado fonema ou sílaba emitidos durante a fala e que tem consequências semânticas em algumas línguas.

Segundo a TLA, conforme Raso (2012), para se investigar qualquer fenômeno linguístico na fala, é preciso considerar a existência de três níveis hierárquicos na sua expressão, quais sejam: (i) *unidade ilocucionária*; (ii) *estrutura informacional*; e (iii) *estrutura sintática*. Tais níveis, ainda segundo Raso (2012, p. 92), precisam ser individualizados como unidades e considerados simultaneamente, no que concerne às suas restrições internas. No caso da análise sintática, foco deste artigo, ela só pode ser realizada dentro da unidade informacional.

Seguindo a notação proposta pela TLA, usa-se uma *barra dupla (/ /)* para marcar uma quebra prosódica terminal, que sinaliza a conclusão de um enunciado; e uma *barra simples (/)* para marcar uma quebra prosódica não terminal, ou seja, que ocorre no interior de um enunciado. Este último tipo de quebra prosódica sinaliza as *unidades tonais*. Essas unidades são responsáveis por definir a estrutura interna dos enunciados e veicular as diferentes *unidades informacionais* (cf. RASO, 2012), das quais destacamos o Comentário (COM), unidade responsável por carregar a *força ilocucionária*. Para ilustrar, observemos o exemplo que se segue:

(1) bfammn01 72: [MAI] só até aí que eu sei o caso / que ele me contou / /

Em (1), temos um *enunciado complexo* marcado por uma quebra prosódica terminal (/ /) no final, e uma quebra prosódica não terminal (/), que o divide em duas unidades tonais. Enunciados realizados por uma única unidade tonal são classificados como enunciados simples. Neste caso a unidade que o compõe, necessariamente, é um Comentário.

## 2 O Corpus de Investigação: C-ORAL-BRASIL

O C-ORAL-BRASIL (RASO; MELLO, 2012) é um corpus de fala espontânea de português brasileiro, representativo da variação diafásica e diatópica mineira. O corpus foi compilado segundo as diretrizes estabelecidas pelo consórcio C-ORAL-ROM (CRESTI; MONEGLIA, 2005), que estuda as quatro maiores línguas românicas europeias: espanhol, francês, italiano e português europeu.

O C-ORAL-BRASIL contém 139 textos de aproximadamente 1.500 palavras, cada um, totalizando 208.130 palavras. Desses textos, 75% textos são do domínio familiar/privado (159.364 palavras) e 25% são do domínio público (48.766 palavras). Cada um desses domínios tem um terço de gravações de caráter monológico, um terço de caráter dialógico e um terço de caráter conversacional. O corpus, da mesma forma que o C-ORAL-ROM, conta não só com as gravações e suas respectivas transcrições, mas também com os arquivos de alinhamento texto-som obtidos através do programa *WinPitch* (MARTIN, 2004)<sup>2</sup>. As transcrições são segmentadas em enunciados e unidades tonais conforme previsto pela TLA e possuem tanto anotações prosódicas quanto anotações morfossintáticas. Além disso, foi constituído um minicorpus representativo do C-ORAL-BRASIL com as seguintes especificações: 20 textos, 29.909 palavras e 5.478 enunciados. Este minicorpus representativo do C-ORAL-BRASIL foi etiquetado informacionalmente segundo a TLA e é dele que extraímos os dados analisados no presente estudo. Para tanto, utilizamos a plataforma DB – IPIC<sup>3</sup>.

### 3 Articulação Sintática e Informacional da Cláusula Relativa

A pesquisa tipológica translinguística (KEENAN; COMRIE, 1977; DE VRIES, 2002), levando em conta a ordem das palavras, indica que há quatro tipos de estratégias sintáticas de cláusulas relativas: (i) *pós-nominal*: a cláusula relativa segue o núcleo (*inglês*); (ii) *pré-nominal*: a cláusula relativa precede o núcleo (*turco*); (iii) *circumnominal*: o núcleo ocorre dentro da cláusula relativa (*banbara*); e (iv) *correlativa*: um núcleo interno é seguido por uma correlativa (*hindí*).

A estratégia de formação sintática convencional das cláusulas relativas no português do Brasil (doravante PB) obedece ao padrão pós-nominal de

---

<sup>2</sup> Download disponível em <<http://www.winpitch.com/>>.

<sup>3</sup> Disponível em: <<http://bit.do/cWvg2>>. DP – IPIC (*Information Structure Database*) é uma plataforma, desenvolvida pelo LABLITA – Laboratório de Pesquisa do Departamento de Humanidades da Universidade de Florença, que contém diferentes tipos de dados e metadados: cada sessão contém áudio, seção de metadados, transcrição, e alinhamento de texto-som.

realização, ou seja, a cláusula relativa segue o núcleo. Este núcleo corresponde ao *antecedente* da cláusula. Sintaticamente, no PB, tradicionalmente, se considera a cláusula relativa àquela formada por uma oração com forma verbal finita introduzida pelo pronome relativo. Este pronome é responsável pela conexão entre as orações principal e subordinada. Na tradição, tais cláusulas equivalem a adjetivos, sendo, portanto, nomeadas orações subordinadas adjetivas, funcionando assim como adjuntos adnominais complexos. Desse modo, as relativas subordinam-se a qualquer termo da cláusula anterior cujo núcleo tenha valor nominal. Observemos o exemplo a seguir:

(2) O *moço*<sub>SN</sub> [*que*<sub>Pro Rel</sub> *trabalha na agência dos Correios*]<sub>Rel</sub> me entregou este pacote.

No exemplo (2), temos a cláusula relativa, *que trabalha na agência dos Correios*, composta pela forma verbal finita *trabalha* (e seu complemento: o sintagma preposicional *na agência dos Correios*) que é introduzida, por sua vez, pelo pronome relativo *que*. A função dessa cláusula é acrescentar informação específica sobre o sintagma nominal antecedente: *moço*. Tal padrão é característico da escrita. No entanto, para se estudar as cláusulas relativas na fala espontânea, é preciso considerar obrigatoriamente dois fatos: (i) a estrutura informacional da fala e (ii) a natureza “fragmentada” de sua sintaxe.

A estrutura informacional da fala, como salientado na seção 2, tem como unidade de referência, o enunciado, cuja composição é dada por uma ou mais unidades informacionais (doravante UI). Estas unidades apresentam funções pragmáticas distintas: *textuais* e *dialógicas* (cf. CRESTI, 2000; RASO, 2012)<sup>4</sup>. As primeiras compõem o texto do enunciado ou são dirigidas à sua interpretação e as últimas regulam a interação junto ao interlocutor. Tais UI

---

<sup>4</sup> Elenco das unidades informacionais:

*Textuais*: Comentário – COM; Tópico – TOP; Apêndice de Comentário – APC; Apêndice de Tópico – APT; Parentético – PAR; Introdutor Locutivo – INT; Comentários Múltiplos – CMM.

*Dialógicas*: Incipitário – INP; Conativo – CNT; Fático – PHA; Alocutivo – ALL; Expressivo – EXP; Conector Discursivo – DCT.

compõem diversos *padrões informacionais* dentro da locução (CRESTI, 2000), entretanto a única unidade necessária e suficiente para realização do enunciado, conforme já anunciado na seção 1, é o Comentário (COM).

A fala é caracterizada por uma forte redução sintática, tanto no que diz respeito à variedade quanto à frequência de cláusulas. Segundo Cresti (2014, p. 375), pelo menos 35% dos enunciados produzidos nas línguas românicas (italiano, francês, espanhol, português europeu e brasileiro) não possuem verbo: a maior parte é realizada por *chunks* linguísticos. Se, de um lado, esses *chunks* não podem ser definidos sempre como sentenças, por outro, eles ainda assim funcionam comunicativamente. Sintaticamente, os *chunks* são encadeados em ilhas informacionais para responder a uma função pragmática. Nesse sentido, sua combinação, dentro de uma locução, não segue necessariamente regras sintáticas para gerar sentenças bem-formadas. Da mesma forma, semanticamente, esses *chunks* não constituem uma entidade composicional resultante em uma proposição com valor de verdade. Cada *chunk* possui uma representação semântica cujo escopo é a unidade informacional.

Em outros termos, esses *chunks* linguísticos funcionam como “ilhas sintático/semânticas” (cláusulas, frases, coleção de fragmentos) dentro das unidades informacionais. Sendo assim, não existiriam, de acordo com Cresti (2014, p. 374), restrições fortes sobre a composição do “recheio” linguístico de uma UI. Então, o que pareceria uma “descontinuidade” da estrutura sintagmática, à qual se está acostumado na escrita, é a realização adequada do fenômeno sintático na fala. Assim, relações sintático/semânticas básicas, tais como regência, escopo da predicação, modificadores, quantificadores, modalidade e relações de subordinação e coordenação se efetivariam dentro dessas ilhas que são as unidades informacionais. Sendo assim, pode-se concluir que o *output* linguístico final da ilocução é, na verdade, uma padronização de ilhas sintático/semânticas que constituem uma unidade pragmática, isto é, o enunciado.

Conforme indicado acima, a sintaxe da fala propriamente dita tem seu escopo, em princípio, dentro de uma única unidade informacional que corresponde a uma ilha sintático-semântica. Tal configuração é denominada *sintaxe linearizada*. Em um enunciado complexo, a um padrão informacional complexo pode haver correspondentemente um padrão sintático. A tal configuração se denomina *sintaxe padronizada*.

No que diz respeito às cláusulas relativas, os dados observados indicam que há ocorrências em que o núcleo modificado e a cláusula relativa se realizam dentro de uma mesma unidade informacional, a que denominamos *cláusula relativa linearizada*. Por outro lado, quando o núcleo modificado e a cláusula relativa ocorrem em unidades informacionais distintas, tem-se uma *cláusula relativa padronizada*. Sendo assim, as unidades informacionais envolvidas nesse último tipo de cláusula pode compor padrões informacionais e prosódicos variados (CRESTI, 2014, p. 383).

Tanto a estruturação linearizada, assim como a padronizada, são apresentadas abaixo, através de discussão de exemplos extraídos do corpus C-ORAL-BRASIL<sup>5</sup> que indicam diversas possibilidades configuracionais para a estruturação sintática em cotejamento com a estruturação informacional:

#### **Cláusula relativa linearizada:** ocorrências

(3) bfamdl01,176,FLA, ”cê ni quer comprar **um trenzinho** *que espirra* pro seu banheiro não // =COM

Em (3), tem-se um enunciado simples formado apenas pela unidade informacional de Comentário (COM), marcada pela quebra prosódica terminal (/ /). Dentro do COM, existe uma cláusula principal *cê ni quer comprar um trenzinho pro seu banheiro não* e uma cláusula relativa introduzida pelo pronome relativo, seguido de uma forma verbal finita, *que espirra*.

(4) bfammn05,69,CAR, ”meu filho é especial /=CMB\_r= **a minha filha** *que foi* é especial /=CMB\_r= mas você é muito especial // =CMM

Em (4), tem-se um enunciado complexo formado por três unidades informacionais, das quais duas são Comentários Múltiplos Ligados (CMB) e uma Comentário Múltiplo (CMM). Tais unidades configuram padrões informacionais que resultam da quebra do isomorfismo entre unidade tonal e unidade informacional. Grosso modo, isso ocorre quando uma única entidade

---

<sup>5</sup> Para efeitos didáticos, nos exemplos extraídos do corpus C- ORAL-BRASIL, o SN está destacado em negrito, o pronome e a cláusula relativa estão destacados em itálico. Já a etiqueta da unidade informacional está em letras maiúsculas.

linguística concluída apresenta mais de uma unidade nuclear (RASO, 2012, p. 113-120). Entretanto, a ilha sintático-semântica que nos interessa está na segunda unidade informacional. Neste segundo CMB, há a cláusula principal *a minha filha é especial* e a cláusula relativa: *que foi*.

(5) bfammn01,55,MAI, "o **bote** *que ea deu nele* /=SCA= derrubou ele /  
=CMM= tão forte que ela era //CMM="

Em (5), tem-se um enunciado complexo formado por dois CMM. Dentro do primeiro CMM, onde ocorre uma *unidade de escansão* (SCA)<sup>6</sup> – na qual a unidade informacional é realizada em mais de uma unidade tonal, há uma cláusula principal *o bote derrubou ele* e uma cláusula relativa introduzida pelo pronome relativo seguido de uma forma verbal finita: *que ea deu nele*.

(6) bpubcv01,79,EMM, "esperamos < que esse > **novo programa** <  
*que > vai vim* /=TOP= ele /=TOP="

Em (6), tem-se um enunciado complexo formado por duas unidades informacionais de Tópico (TOP). Além do COM, esta é a única unidade informacional que carrega um núcleo prosódico que marca o valor funcional da unidade (RASO, 2012, p. 107). Dentro do primeiro TOP deste enunciado, há a cláusula principal *esperamos* e uma oração relativa introduzida pelo pronome relativo seguido de uma forma verbal finita, respectivamente: <*que esse*> *novo programa* < *que > vai vim*.

(7) bfamdl04,230,SIL, "a gente vai folgar **amanhã** *que é* /=SCA= *segunda*  
/=COB= *terça-feira* *cê vai* /=SCA= *po estágio* //COM="

---

<sup>6</sup> No entanto, o perfil prosódico da unidade informacional é realizado na última unidade tonal, enquanto a(s) anterior (es) possuem um perfil neutro, sem valor funcional. A SCA pode acontecer, dentre outras, por razões fisiológicas, uma vez que o conteúdo locutivo da unidade seja muito grande e não é possível realizá-lo em uma única unidade tonal ou por imperícia do falante. Apesar de constituírem partes tonais diferentes de uma mesma unidade informacional, elas são sempre composicionais sintática e semanticamente (RASO, 2012, p. 113-114).

Em (7), tem-se um enunciado complexo formado por um COB e um COM, ambos escansionados. A cláusula relativa se encontra no COB, em que a oração principal é *a gente vai folgar amanhã* e a cláusula relativa é *a que é segunda* - pronome relativo seguido de um predicativo nominal.

**Cláusula relativa padronizada:** ocorrências

(8) bfamdl02,73,BAL, "que eu dou **um exemplo de porta** /=TOP=  
*que é excelente* /=COM

Em (8), tem-se um enunciado complexo marcado por uma quebra não terminal (/) e por uma quebra prosódica terminal (/ /), havendo a delimitação das unidades informacionais de TOP e COM. No TOP, tem-se a cláusula principal *eu dou um exemplo de porta* contendo o SN referente *um exemplo de porta*. Já em COM, tem-se a cláusula relativa introduzida pelo pronome relativo seguido de uma forma verbal finita: *que é excelente*. A estrutura informacional TOP+COM contém, então, a estrutura sintática padronizada oração principal-cláusula relativa.

(9) bfammn01,72,MAI, "só até aí que eu sei **o caso** /=COM=  
*que ele me contou* /=APC

Em (9), tem-se um enunciado complexo marcado por uma quebra prosódica não terminal (/) e por uma quebra prosódica terminal (/ /), havendo a delimitação das unidades informacionais de COM e Apêndice de comentário (APC). No COM, tem-se a cláusula principal *só até aí que eu sei o caso* contendo o SN referente *o caso*. Já em APC, tem-se a cláusula relativa introduzida pelo pronome relativo seguido de uma forma verbal finita: *que ele me contou*. A estrutura informacional COM+APC contém, então, a estrutura sintática padronizada oração principal-cláusula relativa.

(10) bfamcv01,56,EVN, "tem **o SESC** /=COB=  
*que é bom pa caramba* / /  
=COM

Em (10), tem-se um enunciado complexo marcado por uma quebra não terminal (/) e por uma quebra prosódica terminal (/ /), havendo a

delimitação das unidades informacionais de COB e COM. No COM, tem-se a cláusula principal *tem o SESC* contendo o SN referente *o SESC*. Já em COM, tem-se a cláusula relativa introduzida pelo pronome relativo seguido de uma forma verbal finita: *que é bom pa caramba*. A estrutura informacional COB+COM contém, então, a estrutura sintática padronizada oração principal-cláusula relativa.

(11) bfamnm02,143,DFL,”que **a Dodora** /=i-COB= *que era a mamãe* /  
 =PAR= *ia ficar < muito > feliz* /=SCA= *de morar* /=SCA= *perto da*  
*dona Terezinha* /=COB= *que era minha* &#x2013; b [ / 2 ] =SCA= *minha avó* / /  
 =PAR

Em (11), tem-se um enunciado complexo marcado por três quebras não terminais (/) e por uma quebra prosódica terminal (/ /), havendo a delimitação das unidades informacionais de COB, PAR, COB e PAR. No primeiro COB, tem-se o SN referente *a Dodora*. A cláusula relativa, introduzida pelo pronome relativo seguido de uma forma verbal finita, *que era a mamãe*, está no primeiro PAR. Já no segundo COB, tem-se a cláusula principal *ia ficar muito feliz de morar perto da dona Terezinha*. A estrutura informacional COB+PAR+COB contém, então, a estrutura sintática padronizada oração SN referente-cláusula relativa-principal.

(12) bpubcv01,206,MAR,”aqui /=TOP= **fica o plasma** /=CMB= *que*  
*é aquele ali* /=CMB= *< e aqui no fundim* /=TOP= **fica a plaqueta** /  
 =CMB= *que é aquela > que < cê vê >* / /=COM

No exemplo (12), tem-se outro enunciado complexo formado por seis unidades informacionais, respectivamente: TOP, CMB, CMB, TOP, CMB, COM. Nele, encontram-se a *coordenação* de duas relativas subordinadas, estabelecida pela conjunção ‘e’ no segundo TOP: a primeira relativa se encontra distribuída nos dois primeiros CMB do enunciado. O SN referente *o plasma* está posto no primeiro CMB e o pronome relativo seguido da forma verbal finita no segundo CMB *que é aquele ali*, formando o padrão: CMB+CMB. Já a segunda relativa está distribuída da seguinte maneira: O SN referente *a plaqueta* está no terceiro CMB e o pronome relativo seguido da forma verbal

finita está no COM *que é aquela que cê vê*, a última unidade informacional do enunciado. Juntos formam o padrão: CMB+COM.

Sumarizando: como se pode atestar nos exemplos (3) a (7), as cláusulas relativas linearizadas ocorreram nas seguintes unidades informacionais: COM (3), CMB (4), CMM (5) TOP (6), COB (7). Já nos exemplos (8) a (12), referentes às cláusulas relativas padronizadas, encontram-se os seguintes padrões: TOP+COM (8), COM+APC (9), COB+COM (10), COB+PAR+COB (11), CMB+ CMB (12) e CMB+ COM (12).

Observando esses resultados, perceber-se a tendência das cláusulas relativas ocorrerem na unidade informacional de Comentário. Mesmo nos padrões configurados pelas cláusulas relativas padronizadas, o Comentário ou as unidades desdobradas a partir dele (COB, CMM, CMB), foram o mais comum. Tal fato aponta para a relevância da unidade informacional que carrega a força ilocucionária – COM – na realização de cláusulas relativas no PB.

#### 4 As Cláusulas Relativas no Minicorpus do C-ORAL-BRASIL

Tendo-se em vista tanto a caracterização linearizada quanto a padronizada para as cláusulas relativas, conforme exemplificadas na seção 4, procedeu-se à quantificação dessas estruturas no minicorpus do C-ORAL-BRASIL. Os resultados obtidos são explicitados nos Quadros 1, 2 e 3, a seguir:

**Quadro 1** – Distribuição das cláusulas relativas no minicorpus

Total de enunciados <b>5478</b>	
Relativas linearizadas (restritivas)	<b>73</b>
Relativas padronizadas (não-restritivas)	<b>31</b>
Total de relativas	<b>104</b>

Quanto à distribuição das cláusulas relativas linearizadas por unidades informacionais, temos o Quadro 2:

**Quadro 2** – Distribuição das relativas linearizadas por unidades informacionais

COM	51
COB	6
CMM	5
CMB	3
TOP	8
<b>Total</b>	<b>73</b>

No caso das ocorrências das cláusulas relativas padronizadas, é preciso considerar os padrões informacionais resultantes da combinação entre as unidades informacionais, isto é, a oração principal estará em uma unidade e a relativa em outra, compondo, assim, um padrão informacional. Foi observado que o SN referente se encontra sempre na primeira unidade informacional, à esquerda, enquanto a cláusula relativa se encontra na segunda UI, à direita, respeitando assim a estratégia de formação sintática das cláusulas relativas no PB que obedece ao padrão pós-nominal de realização, ou seja, a cláusula relativa segue o núcleo modificado. Os dados indicaram a seguinte distribuição (Quadro 3):

**Quadro 3** – Distribuição dos padrões informacionais das relativas padronizadas

TOP + COM	8
TOP + APT + COM	1
TOP + TOP + COM	1
TOP + CMB	1
TOP + COB	1
COM + APC	1
COM + PHA	1
COM + PAR + COM	1
COB + COM	4
COB + PAR	3
COB + COB	2
COB + PAR + COB	1
CMB + COM	2
CMB + CMB	1
CMM + APC	1
PAR + COB	1
INT + COB + COM	1
<b>Total</b>	<b>31</b>

No PB, como podemos observar no Quadro 2, as cláusulas relativas restritivas (64 em 72) ocorrem preferencialmente na unidade de COM, ou em

unidades resultantes de seus desdobramentos: COB, CMM, CMB. E, mesmo os padrões informacionais formados pelas cláusulas relativas não restritivas, como podemos observar no Quadro 3, são resultado da combinação de uma outra unidade informacional com a de COM (31 em 31). Os resultados encontrados aqui vão de encontro aos dados do italiano, que mostram uma tendência de as cláusulas relativas não restritivas ocorrerem padronizadas nas unidades informacionais com o TOP. Portanto, a tendência encontrada no minicorpus do PB é que as cláusulas relativas tanto restritivas quanto não restritivas ocorrem preferencialmente no COM, unidade informacional portadora da ilocução, responsável pela definição do tipo de ato de fala presente no enunciado. Contudo, ainda ressaltamos que a análise apresentada aqui parte apenas do minicorpus. O próximo passo na nossa pesquisa é verificar se esta tendência se mantém no corpus C-ORAL-BRASIL inteiro.

## **5 Uma Proposta Diagramática para a Descrição Sintático/Semântica e Informacional das Cláusulas Relativas na Fala Espontânea do Português do Brasil**

Para descrever as cláusulas relativas na fala espontânea, tendo em vista a observação do comportamento das cláusulas relativas no minicorpus, apresentamos uma proposta diagramática, baseada no *pensamento diagramático* de Hoffmann (2013), em lugar de uma formalização lógico-sintática, uma vez que este aparato descritivo não dá conta das informações linguísticas suprassegmentais advindas da informação prosódica.

### *5.1 O pensamento diagramático*

Segundo Hoffmann (2013, p. 106), os diagramas formam um subgrupo de ícones, os quais não devem ser reduzidos às formas gráfica ou pictórica, nas quais a *similaridade* é a marca que os liga aos objetos que representam (fotografia, pegadas, planta de uma casa). Na realidade, diagramas são signos que representam relações. Ou seja, a diferença dos diagramas, comparando-os a outros ícones, é que eles são elaborados segundo regras e convenções que são definidas em um sistema de representação. Dessa forma, tanto equações algébricas quanto sentenças linguísticas podem ser compreendidas

como diagramas. Entretanto, o autor reserva o termo diagrama para as representações espaciais das relações (HOFFMANN, 2013, p. 111).

Sendo assim, o *pensamento diagramático* de Hoffmann (2013) é definido como a construção de representações externas baseada em regras e convenções de um sistema de representação. Em outros termos, o *diagrama* é uma representação externa de relações que são construídas segundo regras e convenções de (e por meio de) elementos e relações disponíveis entre elementos de um em dado sistema de representação. Tal sistema representacional fornece os meios e as restrições na construção de diagramas, bem como as suas possibilidades de manipulação. Importante destacar que essa noção está sustentada na Semiótica de Charles S. Peirce (apud HOFFMANN, 2013).

Para Hoffmann (2013), ainda, o pensamento diagramático pode ser conceitualizado como um caso de cognição distribuída (*distributed cognition*). Já que, como sistema representacional, o diagrama tem uma realidade independente, pois não se pode definir suas regras e os significados de seus elementos arbitrários. Entretanto, são ferramentas presentes na cultura, as quais têm-se à disposição no pensamento e na ação. Dessa forma, esse tipo de representação facilita processos individuais ou sociais em situações que são complexas para serem conduzidas exclusivamente por meios cognitivos internos ou formais. Assim, tomamos os diagramas para representar as relações entre estrutura prosódica e estrutura sintática.

Como o diagrama representa relações que obedecem a regras gerais, o autor afirma que se podem realizar experimentos, anotar resultados, repetir experimentos a partir de relações diagramáticas já constituídas, uma vez que, observados os mesmos preceitos, devam se obter os mesmos resultados (HOFFMANN, 2013, p. 107). Para ele, o pensamento diagramático é uma ferramenta para produzir conhecimento, *não apenas para solucionar problemas de lógica*. Entretanto, ressalva que existem limitações no uso do pensamento diagramático, mas esta limitação pode ser substituída ou complementada com o pensamento verbal-conceitual.

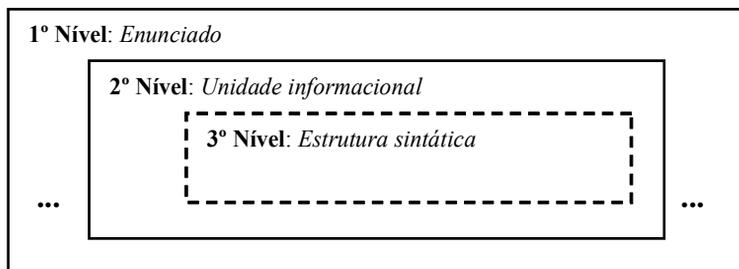
Assim, uma sequência de palavras como *O moço que trabalha na agência dos Correios me entregou este pacote* poderia ser interpretada como um ícone, se a entendermos como a representação de uma relação. No caso, estabelecida pelo verbo *entregar* ao selecionar, pelos menos, três argumentos: *um agente, um*

tema e um alvo. Se lermos *O moço que trabalha na agência dos Correios me entregou este pacote* como um signo que representa uma relação que é realizada com regras da gramática do português do Brasil, temos um diagrama. Embora, ressalte Hoffmann (2013), a linguagem diária não seja necessariamente um sistema tão consistente, a exemplo dos sistemas axiomáticos da matemática.

No caso das cláusulas na fala espontânea do PB, temos, então, dois sistemas de regras a considerar: as regras de formação e restrições sintáticas das cláusulas e as regras e restrições informacionais da prosódia. Ignorada nos estudos da diamesia escrita, o componente prosódico envolve um conjunto de elementos suprasegmentais que engloba acento, ritmo, tom, tempo e entonação, cujos correlatos auditivos constituem-se de melodia, intensidade e velocidade da fala. Como ambos expressam relações, podem, portanto, serem tomados como diagramas.

Assim, conforme a seção 1 deste artigo, ao analisar o fenômeno linguístico na fala espontânea, faz-se necessário observar sua unidade de referência, (i) o enunciado, e a hierarquia de sua expressão: (ii) unidade informacional; (iii) estrutura sintática. Ou seja, as informações suprasegmentais (i) e (ii) vêm antes da segmento linguístico (iii). Em outros termos, temos *diagramas contidos em diagramas*. De outro modo: o enunciado contém ( $\supset$ ) a unidade informacional que contém ( $\supset$ ) a estrutura sintática. Para expressar essa hierarquia, em termos gráficos, apresentamos o seguinte esquema no qual as informações (i) e (ii) estão expressas em linhas contínuas e a (iii) em uma linha traçada:

Enunciado  $\supset$  Estrutura informacional  $\supset$  Estrutura sintática



**Diagrama 1** – Estrutura do enunciado na fala

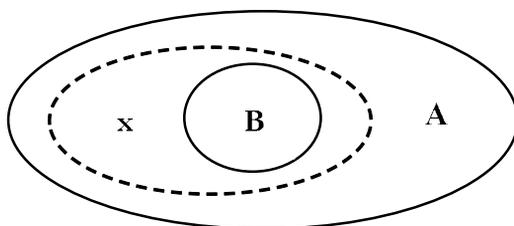
Esses níveis que compõem o enunciado, como já salientado, precisam ser individualizados como unidades e considerados simultaneamente, no que concerne às suas restrições internas. Dessa forma, um aparato descritivo precisa dar conta dessa dupla especificidade. Devido ao caráter desse trabalho, a nossa descrição foca o 3º nível – o sintático.

Posto isso, considerando que (i) a cláusula relativa na fala espontânea é um tipo de ilha sintático/semântica envelopada dentro de um *padrão informacional*, cuja composição se dá por uma ou mais *unidades informacionais* do tipo textual; e que (ii) as relações sintático/semânticas básicas se efetivariam dentro de ilhas sintático/semânticas no interior das unidades informacionais (cf. seção 4), propõe-se sua conceituação nos seguintes termos:

(1) *Cláusula relativa linearizada*

É um tipo de cláusula em que há um SN (x) complexo que contém ( $\supset$ ) uma cláusula relativa restritiva (B) como um termo sintático adjunto, ligada por um pronome relativo. A função dessa cláusula relativa restritiva é delimitar semanticamente este SN (x) por meio de uma informação pressuposta. Este SN está contido ( $\subset$ ) na cláusula principal (A) como um de seus argumentos sintáticos. Logo, a cláusula relativa restritiva (B) está contida na cláusula principal (A). Em termos prosódicos, tanto a cláusula principal quanto a cláusula relativa restritiva estão linearizadas sintaticamente dentro de uma mesma unidade informacional (UI) no enunciado – simples ou complexo:

Relativa linearizada =  $x \supset B / x \subset A /$  Logo,  $B \subset A$



**Diagrama 2 – Relativa linearizada**

A seguir, seguem exemplos de cláusulas relativas linearizadas, identificadas dentro das unidades informacionais nos enunciados retirados do minicorpus do C-ORAL BRASIL, conforme a descrição proposta:

13. tem [ **creme** [*que*<sub>ProRel</sub> + *ni dá com o cabelo não*] <sub>CRel</sub> ]<sub>SN</sub> // COM

Em (13), o SN complexo *creme que ni dá com o cabelo não* contém uma cláusula relativa restritiva *que ni dá com o cabelo não* como um termo sintático adjunto. Essa cláusula delimita semanticamente este SN por meio de uma informação pressuposta: *creme que dá com o cabelo*. Como o SN *creme* possui um alcance potencial de referentes no mundo (*creme que dá* e *creme que não dá com o cabelo*), a *restrição* da cláusula relativa incide apenas sobre o conjunto de *creme que não dá com o cabelo*. Tal proposição deve ser verdadeira dentro do conjunto dos referentes da cláusula. Isso porque este SN está contido na cláusula principal *tem creme que ni dá com o cabelo não* como um de seus argumentos sintáticos. Portanto, a sua negação implicaria na negação de toda a cláusula principal. Logo, a cláusula relativa restritiva está contida na cláusula principal como parte de um de seus argumentos. Em termos prosódicos, tanto a cláusula principal quanto a cláusula relativa restritiva estão linearizadas sintaticamente dentro de uma mesma UI, o COM, dentro do enunciado simples marcado pela quebra prosódica terminal (... //).

14. [ **o bote** [*que*<sub>ProRel</sub> + *ea deu nele*] <sub>CRel</sub> ]<sub>SN</sub> derrubou ele // CMM

Em (14), o SN complexo *o bote que ea deu nele* contém uma cláusula relativa restritiva *que ni ea deu nele* como um termo sintático adjunto. Essa cláusula delimita semanticamente este SN por meio de uma informação pressuposta: *o bote que ela não deu nele*. Como o SN *bote* possui um alcance potencial de referentes no mundo (*o bote que ela deu, o bote que ela tentou dar, e o bote que ela não deu nele*), a *restrição* da cláusula relativa incide apenas sobre o conjunto de *o bote que ea deu nele*. Tal proposição deve ser verdadeira dentro do conjunto dos referentes da cláusula. Isso porque este SN está contido na cláusula principal (*o bote que ea deu nele derrubou ele*) como um de seus argumentos sintáticos. Portanto, a sua negação implicaria na negação de toda a cláusula principal. Logo, a cláusula relativa restritiva está contida na cláusula principal

como parte de um de seus argumentos. Em termos prosódicos, tanto a cláusula principal quanto a cláusula relativa restritiva estão linearizadas sintaticamente dentro de uma mesma UI, o CMM, dentro do enunciado simples marcado pela quebra prosódica terminal (... //).

15. ea mordia e picava com [**o ferrão** [*que ea tem no cabo também*]<sub>CRel</sub> ]<sub>SN</sub> /  
/=COM

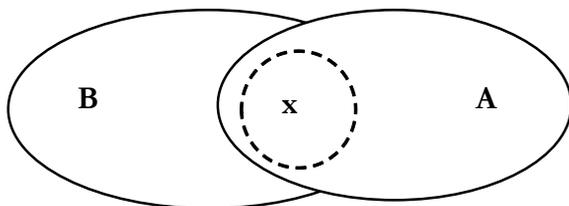
Em (15), o SN complexo *o ferrão que ea tem no cabo também* contém uma cláusula relativa restritiva *que ea tem no cabo também* como um termo sintático adjunto. Essa cláusula delimita semanticamente este SN por meio de uma informação pressuposta: *o ferrão que ea não tem no cabo*. Como o SN *ferrão* possui um alcance potencial de referentes no mundo (*ferrão que ea tem e o ferrão que ea não tem no cabo*), a restrição da cláusula relativa incide apenas sobre o conjunto de *ferrão que ea tem no cabo*. Tal proposição deve ser verdadeira dentro do conjunto dos referentes da cláusula. Isso porque este SN está contido na cláusula principal (*ea mordia e picava com o ferrão que ea tem no cabo também*) como um de seus argumentos sintáticos. Portanto, a sua negação implicaria na negação de toda a cláusula principal. Logo, a cláusula relativa restritiva está contida na cláusula principal como parte de um de seus argumentos. Em termos prosódicos, tanto a cláusula principal quanto a cláusula relativa restritiva estão linearizadas sintaticamente dentro de uma mesma UI, o COM, dentro do enunciado simples marcado pela quebra prosódica terminal (... //).

## (2) *Cláusula relativa padronizada*

É um tipo de cláusula (B) que está numa relação de interseção ( $\cap$ ) semântica com a cláusula principal (A), através de um pronome relativo cuja função intraoracional permite recuperar anaforicamente o conteúdo semântico de um SN antecedente posto na cláusula principal (A), entretanto não está contida ( $\not\subset$ ) dentro desta cláusula principal como um de seus argumentos sintáticos. A função dessa cláusula relativa não restritiva é adicionar uma informação suposta sobre um SN da cláusula principal, o qual pode estar declarado ou elíptico. Em termos prosódicos, a cláusula relativa não restritiva ocorre numa unidade informacional diferente daquela da cláusula principal,

e seus elementos podem ocorrer padronizados em diferentes unidades informacionais (UIs) dentro de um enunciado complexo (... /... //):

Relativa padronizada =  $B \cap A = \{X\}$  / Entretanto,  $B \not\subset A$



**Diagrama 2:** Relativa Padronizada

Abaixo, seguem exemplos de cláusulas relativas padronizadas, identificadas dentro de diferentes unidades informacionais nos enunciados retirados do minicorpus do C-ORAL BRASIL, conforme a descrição proposta:

16. tem o **[SESC]**<sub>SN</sub> / COB [*que é bom pa caramba*]<sub>CRel</sub> // COM

Em (16), a cláusula relativa padronizada *que é bom pa caramba* está numa relação de interseção semântica com a cláusula principal *tem o SESC*, através do pronome relativo *que*, cuja função intraoracional, recupera anaforicamente o conteúdo semântico de um SN antecedente posto na cláusula principal: **o SESC**. Entretanto, a cláusula relativa padronizada não está contida dentro da cláusula principal como um de seus argumentos sintáticos. Ela é um termo aposto. A função dessa cláusula relativa não restritiva é adicionar uma informação suposta sobre um SN da cláusula principal: *o SESC*. Portanto, ela pode ser negada sem implicar no valor de verdade da cláusula principal. Em termos prosódicos, a cláusula relativa não restritiva ocorre na UI de COM e a cláusula principal na UI de COB dentro de um enunciado complexo. Portanto, seus elementos ocorrem padronizados em UIs diferentes.

17. que eu dou [**um exemplo de porta**]<sub>SN</sub> / TOP [*que é excelente*]<sub>CRel</sub> /  
/ COM

Em (17), a cláusula relativa padronizada *que é excelente* está numa relação de interseção semântica com a cláusula principal *que eu dou um exemplo de porta* através do pronome relativo *que*, cuja função intraoracional, recupera anaforicamente o conteúdo semântico de um SN antecedente posto na cláusula principal: *um exemplo de porta*. Entretanto, a cláusula relativa padronizada não está contida dentro da cláusula principal como um de seus argumentos sintáticos. Ela é um termo aposto. A função dessa cláusula relativa não restritiva é adicionar uma informação sobre um SN declarado (*um exemplo de porta*) da cláusula principal. Portanto, ela pode ser negada sem implicar no valor de verdade da cláusula principal. Em termos prosódicos, a cláusula relativa não restritiva ocorre na UI de COM e a cláusula principal na UI de TOP dentro de um enunciado complexo. Portanto, seus elementos ocorrem padronizados em UIs diferentes.

18. só até aí que eu sei [**o caso**]<sub>SN</sub> / COM [*que ele me contou*]<sub>CRel</sub> // APC

Em (18), a cláusula relativa padronizada *que ele me contou* está em uma relação de interseção semântica com a cláusula principal *só até aí que eu sei o caso* através do pronome relativo *que*, cuja função intraoracional recupera anaforicamente o conteúdo semântico de um SN antecedente posto na cláusula principal: *o caso*. Entretanto, a cláusula relativa padronizada não está contida dentro da cláusula principal como um de seus argumentos sintáticos. Ela é um termo aposto, pois a função dessa cláusula relativa não restritiva é adicionar uma informação suposta sobre um SN declarado *o caso* da cláusula principal. Portanto, ela pode ser negada sem implicar no valor de verdade da cláusula principal. Em termos prosódicos, a cláusula relativa não restritiva ocorre na UI de APC e a cláusula principal na UI de COM dentro de um enunciado complexo. Portanto, seus elementos ocorrem padronizados em UIs diferentes.

Uma vez apresentadas as definições para as cláusulas relativas linearizada e padronizada na fala espontânea via formato diagramático, acreditamos que conseguimos cobrir descritivamente as possibilidades da realização dessas cláusulas, haja vista a observação de suas ocorrências no minicorpus. Afinal, o modelo diagramático permite a visualização não só da relação entre as cláusulas principal e relativa, como também os elementos sintático/semânticos e informacional que compõem cada uma delas. Especialmente, a

cláusula relativa padronizada, dada à variabilidade de composição de seus padrões informacionais (vide Quadro 3) e a distribuição de seus elementos dentro das UI. Como pode ser constatado nos exemplos<sup>7</sup> a seguir:

19. *o livro que eu adotei* / TOP *não compra* // COM

20. *todas essas que cê tá vendo aqui hoje* /TOP *foram coletadas hoje* // COM

21. *que a Dodora* /COB *que era a mamãe* /PAR *ia ficar muito feliz de morar perto da dona Terezinha* /COB *que era minha minha avó* // =PAR

Em (19), temos o padrão TOP+COM, no qual a cláusula relativa padronizada ocorre antes da cláusula principal: *o livro que eu adotei* + *não compra*. O que chama a atenção nessa ocorrência é o fato do SN referente não estar declarado na cláusula principal. No entanto, sabemos que este argumento elíptico trata-se do SN *o livro* posto na cláusula relativa padronizada, ou seja: *não compra* [*o livro*]. Isso confirma a relação de interseção semântica estabelecida entre as duas cláusulas.

Em (20) temos o padrão TOP+COM, no qual a cláusula relativa padronizada também ocorre antes da cláusula principal: *todas essas que cê tá vendo aqui hoje* + *foram coletadas hoje*. Nesse exemplo, o SN referente não está declarado na cláusula principal. No entanto, sabemos que o argumento elíptico trata-se do SN *todas essas* pertencente à cláusula relativa padronizada, ou seja: [*todas essas*] *foram coletadas hoje*. Isso confirma a relação de interseção semântica entre as duas cláusulas.

Em (21), temos o padrão COB+PAR+COB, no qual os elementos da cláusula estão distribuídos assim: SN referente *a Dodora* no primeiro COB, cláusula relativa padronizada *que era a mamãe* no PAR e a cláusula principal *ia ficar muito feliz de morar perto da dona Terezinha* no segundo COB. Segundo essa definição apresentada de cláusula relativa padronizada, os seus elementos podem ocorrer padronizados em diferentes unidades informacionais (UIs) dentro de um enunciado complexo. Aqui a relativa padronizada está distribuída respectivamente no primeiro COB (*a Dodora*) seguido do PAR (*que era a mamãe*).

---

<sup>7</sup> As cláusulas em análise estão grafadas em itálico.

O SN a Dodora da cláusula relativa trata-se do SN elíptico que compõe o argumento da cláusula principal, posta no segundo COB: [a Dodora] ia ficar muito feliz de morar perto da dona Terezinha. Isso confirma a relação de interseção semântica entre as duas cláusulas.

## Considerações Finais

Após investigação no minicorpus C-ORAL-BRASIL, têm-se os seguintes resultados: o estudo (i) corrobora a diferença entre cláusulas relativas restritivas e relativas não restritivas em termos prosódicos: as primeiras ocorrem linearizadas em uma mesma unidade informacional e as últimas, padronizadas em mais de uma unidade informacional; (ii) demonstra que no PB, tanto relativas restritivas (65 em 73) quanto não restritivas (31 em 31) ocorrem preferencialmente na unidade de COM (ou unidades resultantes de seus desdobramentos: COB, CMM, CMB). Mesmo o padrão informacional preferencial das relativas não restritivas é resultado da combinação do Comentário com outra unidade informacional.

Além disso, para descrever as cláusulas relativas na fala espontânea, tendo em vista o seu comportamento no minicorpus, foi preciso buscar um modelo alternativo de descrição, uma vez que a formalização lógico-sintática, como aparato descritivo, não dá conta de informações linguísticas suprasegmentais advindas da informação prosódica. Sobretudo no que concerne à cláusula relativa padronizada, devido à variabilidade de composição de seus padrões informacionais e à distribuição de seus elementos dentro das unidades informacionais. Assim, apresentamos o modelo diagramático que consegue alcançar: (i) as relações internas entre a cláusula relativa e a principal – na relativa linearizada a relação é de estar contida sintaticamente, enquanto na relativa padronizada a relação é de interseção semântica; e (ii) a informação prosódica definidora da estrutura interna dos enunciados – escopo de realização das cláusulas relativas.

## Referências

ANTHONY, L. *Developing AntConc for a new generation of corpus linguists*. Disponível em: <<http://bit.do/cWxvj>>. Acesso em: 11 maio 2015.

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

CRESTI, E. *Corpus di Italiano parlato*. v. 1. Firenze: Accademia della Crusca, 2000. p.41-166.

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. (Org.). *Spoken corpora and linguistic studies: problems and perspectives*. Amsterdam; Philadelphia: John Benjamins, 2014.

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

DE VRIES, M. *The syntax of relativization*. Amsterdam: Universiteit van Amsterdam; LOT, 2002.

HOFFMANN, M. H. G. *Cognição e pensamento diagramático*. In: QUEIROS, J.; MORAIS, L. *A lógica de diagramas de Charles Sanders Peirce: implicações em ciência cognitiva, lógica e semiótica*. Juiz de Fora: Editora UFJF, 2013. p. 105-137.

KEENAN, E. L.; COMRIE, B. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, v. 8, n. 1, p. 63-99, 1977.

MARTIN, P. WinPitch LTL II, a Multimodal Pronunciation Software. *Proceedings of InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems*. Venice, 2004.

PANUNZI, A.; GREGORI, L. DB-IPIC: an xml database for the representation of information structure in spoken language. In: MELLO, H. R.; PANUNZI, A.; RASO, T. (Org.). *Pragmatics and prosody: illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 2011. p. 133–150.

RASO, T. O C-ORAL-BRASIL e a Teoria da Língua em Ato. In: RASO, T.; MELLO, H. (Org.). *C-ORAL-BRASIL*. Corpus de referência da fala espontânea informal do português do Brasil. Belo Horizonte: UFMG, 2012.

RASO, T.; MELLO, H. (Org.). *C-ORAL-BRASIL*. Corpus de referência da fala espontânea informal do português do Brasil. Belo Horizonte: UFMG, 2012.

Recebido em: 03/04/2015

Aceito: 10/07/2015