doi: http://dx.doi.org/10.5007/1984-8412.2017v14n2p2061

RELATIVE CLAUSES IN SPONTANEOUS SPEECH: A DEFINITION BASED ON THE LANGUAGE INTO ACT THEORY

CLÁUSULAS RELATIVAS NA FALA ESPONTÂNEA: UMA DEFINIÇÃO BASEADA NA TEORIA DA LÍNGUA EM ATO

ORACIONES DE RELATIVO EN EL HABLA ESPONTÁNEA: UNA DEFINICIÓN BASADA EN LA TEORÍA DE LA LENGUA EN ACTO

Crysna Bonjardim da Silva Carmo*
Federal University of Minas Gerais
Bahia State University

ABSTRACT: This paper presents a definition for relative clause in the context of Brazilian Portuguese spontaneous speech based on the Language into Act Theory (CRESTI, 2000). This theory defines the utterance as the reference unit for speech, i.e., the smallest pragmatically autonomous linguistic unit (the speech act). In this framework, the dependency relations established by syntax can be altered by informational structure, i.e., syntax is restricted to the limits of the informational units delimited by prosody, thus forming syntactic/semantic islands. This study made use of the C-ORAL BRASIL spontaneous speech corpus (RASO; MELLO, 2012). Based on the definition proposed here, the following conclusions were reached concerning relative clauses: (i) they show the semantics of restriction of a *denotation set* within an underlying *reference set*; (ii) they appear linearized within a simple or compound utterance; (iii) they occur in isolation or as sentence components within utterances.
KEYWORDS: Spontaneous speech. Relative clauses. Language into Act Theory. Corpus Linguistics.

RESUMO: Este artigo apresenta uma definição de cláusula relativa para a fala espontânea do português brasileiro baseada na Teoria da Língua em Ato (CRESTI, 2000). Essa teoria postula o enunciado como a unidade de referência da fala, ou seja, como a menor unidade linguística pragmaticamente autônoma (ato de fala). Nesse quadro, as relações de dependência estabelecidas pela sintaxe podem ser alteradas pela estrutura informacional da prosódia, ou seja: a sintaxe está restringida aos limites das unidades informacionais delimitadas pela prosódia, configurando-se em ilhas sintático-semânticas. Para realizar este estudo, utilizou-se um *corpus* de fala espontânea, o C-ORAL BRASIL (RASO; MELLO, 2002). A partir da definição de relativa postulada, chegou-se aos seguintes resultados acerca dessas cláusulas: (i) apresentam a semântica de restrição de um *conjunto denotado* dentro de um *conjunto de referência* subjacente; (ii) ocorrem linearizadas dentro do enunciado simples ou complexo; (iii) ocorrem isoladas ou como componentes de sentenças dentro dos enunciados.
PALAVRAS-CHAVE: Fala espontânea. Cláusulas relativas. Teoria da Língua em Ato. Linguística de *Corpus*.

* PhD student at the Federal University of Minas Gerais (UFMG) – Graduate Program in Linguistic Studies (POSLIN). Assistant professor at the Bahia State University (UNEB). E-mail: crysnabonjardimsc@gmail.com.

RESUMEN: Este estudio presenta una definición de la oración de relativo para el habla espontánea del portugués brasileño basada en la Teoría de la Lengua en Acto (CRESTI, 2000). Esta teoría postula el enunciado como la unidad de referencia del habla, o sea, como la menor unidad lingüística pragmáticamente autónoma (acto de habla). En este enfoque, las relaciones de dependencia establecidas por la sintaxis pueden ser modificadas por la estructura informacional de la prosodia, o sea, la sintaxis está restringida a los límites de las unidades de información delimitadas por la prosodia, configurándose en islas sintáctico-semánticas. Para realizar este estudio, fue utilizado un corpus de habla espontánea, el C-ORAL-BRASIL (RASO; MELLO, 2012). Partiéndose de la definición postulada este estudio, se llegó a las siguientes conclusiones sobre esas oraciones: (i) presentan la semántica de restrición de un conjunto denotado dentro de un conjunto de referencia subyacente; (ii) ocurren de forma lineal dentro del enunciado simple o complejo; (iii) ocurren aisladas o como componentes de oraciones dentro de los enunciados.
PALABRAS CLAVE: Habla espontánea. Oraciones de relativo. Teoría de la Lengua en Acto. Lingüística de *Corpus*.

1 INTRODUCTION

Relative clauses are still part of linguistics research agenda, particularly when it comes to spontaneous speech, as written language remains the focus of traditional approaches. Even studies dedicated to speech data analysis focus on the transcription, which is itself written data, thus disregarding the influence of prosody on speech segmentation. Before the 20th century, there was obviously no technology allowing adequate research on speech data, given that the simultaneous observation of text and acoustic signal is paramount for the study of speech, which has only become possible thanks to the methodology brought about by the advent of computers and Corpus Linguistics. Furthermore, the concept of relative clause itself is still under debate, as traditional definitions that are available do not take the nature of spontaneous speech into consideration and are imprecise in their classification of relative clauses as restrictive and non-restrictive. Considering these aspects, as well as the fact that dependency relationships established through syntax can be altered by information structure, i.e., that the prosodic information affects the segmental information in spoken language, the present study proposes a definition for clausal relativization based on the *Language into Act Theory* – a theoretical and methodological framework for the study of spontaneous speech – and on data from a spontaneous speech minicorpus of Brazilian Portuguese – the C-ORAL-BRASIL. The theory and the data used in this study are presented below.

2 THE LANGUAGE INTO ACT THEORY (L-ACT)

The Language into Act Theory (L-AcT; CRESTI, 2000) deals with spontaneous speech taking the structuring role of prosody into consideration. Spontaneous speech is understood as speech whose planning and execution takes place simultaneously (NENCIONI, 1983). Thus, the concept does not encompass speech derived from written texts, such as read speech, plays, movies, soap operas, etc.

L-AcT is an extension of the *Speech Act Theory* (AUSTIN, 1962); nevertheless, L-AcT is a *corpus-driven* theory. Considering the transient nature of speech – which results from its environment, the soundwaves, and its channel, the air (RASO, 2013) –, L-AcT builds corpora transcribed with semi-orthographic criteria (see RASO, MELLO, 2009; MELLO et al., 2012). In other words, in this framework, a morphosyntactic codification is adopted to capture the typical features of speech already in the transcription process. With this methodology, L-AcT tries to avoid that the transcription process be influenced by written language and to capture ongoing speech phenomena.

Thus, based on prosodic criteria, L-AcT defines the *utterance* as the reference unit for spontaneous speech, i.e., as the smallest, pragmatically autonomous linguistic unit that, carrying illocutionary force, corresponds to a speech act. In other words, in taking prosodic criteria into consideration, L-AcT breaks with that tradition of grammatical studies that regards the sentence (CHOMSKY, 1970) or the clause (CHAFE, 1988) as the basic unit for language. Ultimately, such postulates require a verbal nucleus to be present in the unit. There are, however, verbless speech acts, which are composed of linguistic chunks having communicative value, e.g.,

fragments, interjections, adverbs and some types of clauses, such as nominal ones. Examples (1) through (6), taken from the C-ORAL-BRASIL minicorpus, are all instances of utterances, since they are pragmatically autonomous units that can, therefore, be interpreted in isolation, despite the difference among them in terms of locutive content.

(1) bfamcv01, 213: eu acho que a gente deve chamar os < times > legais //[1]
I think that we should invite the cool teams //

(2) bfamcv03, 10: vou jogar na quatro //
I'll shoot for the four [-ball] //

(3) bfamcv03, 9: < não > //
no //

(4) bfamcv03, 19: uhn //
hum //

(5) bfamcv03, 51: < nts > //
tsk //

(6) bfamcv04, 416: pra todo mundo //
for everybody //

From (1) through (6), there are a complex sentence, a simple sentence, an adverb, an interjection with negative value, a click expressing disappointment, and a prepositional phrase. Each of these utterances have specific prosodic profiles. Hence, they are identified as speech acts. This is not the case, however, with the sequence below.

(7) bfamcv01, 144, GIL: pra próxima taça /
for the next tournament /

The sequence in (7) does not constitute a pragmatically autonomous unit, i.e., it is not perceived as an utterance. In fact, the sequence *pra próxima taça* is part of an utterance, shown in (8) below, whose second part, *a gente tem que também mandar um e-mail*, is perceived as a linguistic act. This is because this part of the utterance carries the prosodic nucleus of the *illocution*, i.e., the speech act. Therefore, the final sequence of the utterance is interpreted as pragmatically autonomous.

(8) bfamcv01, 144: pra próxima taça / a gente tem que também mandar um e-mail //
for the next tournament / we should also send an e-mail //

Thus, (8) consists of an utterance composed of two parts. Each of these parts possesses a different prosodic profile determining the function within the utterance. In other words, in addition to supplying the locutive content with pragmatic autonomy, prosody is also responsible for segmenting the utterance in the stream of speech.

Regarding speech segmentation, prosody marks two types of breaks: those perceived as conclusive and that delimit utterances, the so-called *terminal breaks* ("//" in the transcription), and those perceived as non-conclusive and that mark the internal structure of utterances in terms of *information units* (IUs) called non-terminal breaks ("/" in the transcription). Utterances composed of one IU alone are referred to as *simple utterances*. These are necessarily composed of a *Comment* IU, which is responsible for the pragmatic

[1] Utterances from the C-ORAL-BRASIL minicorpus are introduced by an identification label. For example, "bfamcv02" provides the following information: the language (b = Brazilian Portuguese), the context (fam = family/private, pub = public), interactional typology (cv = conversation, dl = dialogue, mn = monologue). The two digits indicate the order of the text in its session. Angled brackets indicate overlapping speech. Simple slashes indicate non-terminal prosodic breaks, while double slashes indicate terminal prosodic breaks.

autonomy and interpretability, since this unit carries the illocutionary force. Utterances made up of more than one IU are referred to as *compound utterances*, which are composed of one or more IUs together with the Comment. Examples (9) and (10) show, respectively, a compound utterance made up of two IUs (TOP-COM) and a simple utterance made up of a *Comment* IU.

(9) bfamcv01, 144: pra próxima taça /=TOP=a gente tem que também mandar um e-mail //=COM (same as Ex. 8)

(10) bfamcv01, 8: < eles são piores do que o > Durepox //=COM
they are worse than the Durepox [a football team] //

L-AcT emphasizes that prosodic breaks have two simultaneous levels, a prosodic one and an informational one. At the prosodic level, terminal and non-terminal breaks are defined in terms of *tone units*. At the informational level, breaks are defined in terms of *information units*. As mentioned above, IUs make up the internal structure of utterances, which are internally delimited by non-terminal prosodic breaks.

Information units are characterized by prosodic profiles that have specific functions within the utterance and are defined considering their position in relation to the COM – the unit responsible for carrying the illocution. Such IUs are identified based on their F_0 (fundamental frequency)[2] contour. Resulting from voluntary movements, IUs show perceptual prominence and carry informational values, contrasting with involuntary movements caused by contextual, micromelodic features. It is because of the intentional and voluntary nature of their production that IUs are perceived by the hearer. Thus, based on the perceptual phonological model of the IPO (*Institute for Perception Research*) approach, L-AcT implements the *Information Patterning Theory*, which identifies the following prosodic profiles as characteristic of IUs.

a) *Root*: it carries the illocution and defines the COM unit only.
b) *Prefix*: it precedes units of the root type and defines the TOP unit.
c) *Suffix*: it follows units of the root type and defines the Appendixes of TOP and COM.
d) *Postfix*: it may precede or follow units of the root type but cannot occur at the beginning of utterances; it defines the Parenthetic.

Table 1 shows the information units according to L-AcT.

Table 1: Information units according to L-AcT

	Type	**Tag**	**Function**
Textual Units	Comment	COM	carries the illocutionary force of the utterance
	Topic	TOP	supplies the domain of application for the illocutionary force
	Appendix of comment/topic	APC/APT	integrates the text of the unit of which it is the appendix
	Parenthetic	PAR	provides instructions as to how the utterance of part of it should be interpreted
	Locutive Introducer	INT	signals that what follows it (usually, a meta-illocution) pertains to a hierarchical level different from that of the enunciation
Dialogic	Allocutive	ALL	singles out the addressee, marks social cohesion

[2] It describes the number of complete cycles by the vocal folds per unit time (usually seconds) during phonation.

Units	Incipit	INP	signals the beginning of a turn or an utterance
	Expressive	EXP	supplies emotional support for the speech act, marks social cohesion
	Discourse Connector	DCT	signals the continuity of a sequence in relation to a previous one
	Phatic	PHA	signals the opening or the maintenance of the communicative channel by the speaker
	Conative	CNT	pushes the interlocutor to accomplish or abandon certain action

Source: Bossaglia (2015, p. 312)

In summary, *textual units* either compose the text of the utterance (COM, TOP, APC, APT) or are directed to its interpretation (INT, PAR), while *dialogic units* regulate the interaction (ALL, CNT, DCT, EXP, INP, PHA). Dialogic units correspond to what is traditionally referred to as discourse markers. Other IUs exhibit certain particularities, as they infringe the isomorphism between the prosodic and information patterns that foresees the attribution of a pragmatic-informational value to each unit within an utterance. These IUs are the *Multiple Comment* (CMM), the *Bound Comment* (COB), and the *Scanning Unit* (SCA).

Multiple Comments are melodically patterned and holistically interpretable; in other words, they are not interpretable as separate information units, since they establish a logic relationship that can be causal, conditional, temporal, etc. This patterning is characteristic of lists and comparisons among others. Bound Comments are not melodically patterned; hence, their interpretation is not holistic. They make up the so-called *stanzas* (CRESTI, 2009), i.e., units containing more than one illocution marked by a continuative prosody, indicating that that piece of discourse is not concluded and that it will only be when the last illocution be accomplished. This pattern is characteristic of monologic texts. Examples (11) and (12) show patterns with CMMs and COBs respectively.

(11) bfamcv01, 66: mas a < gente tenta > fazer reunião /=*CMM*= galera nũ < comparece > //=*CMM*
but we try to arrange meetings / the guys don't turn up //

(12) bfamcv01, 94: a gente podia fazer a taça aqui /=*COB*_r= todo mundo vai adorar /=*COB*_r= e tal //=COM
we could have the tournament here / everybody will love it / and stuff //

Scanning Units occur when an IU is realized with more than one tone unit, which may be either due to the extension of the locutive content or due to issues related to expressive needs or inaccuracy on the part of the speaker. Scanning units show a neutral prosodic profile; only the final unit of the sequence carries informational value, since it carries the nucleus of the illocution. It is important to stress that the occurrence of SCAs is restricted to textual units and that they are syntactically compositional.

(13) bfamcv01, 134: *acho < que a gente > tem* /=*SCA*= que olhar direito //=COM
I think we have / to look carefully //

As may be noted, L-AcT extends the inventory of speech units in terms of information structuring, an inventory that had been restricted to Topic and Comment units only. This way of structuring spoken discourse into information units is paramount for the syntactic understanding of spontaneous speech postulated by L-AcT and, consequently, for the definition of *clausal relativization* that will be presented in this paper.

2.1 THE SYNTACTIC AND INFORMATIONAL ARTICULATION ACCORDING TO L-ACT

According to L-AcT, the notion of syntax is articulated with prosodic structure. Thus, syntax, in its strict sense, is confined within the IU, which comprises a genuine "syntactic/semantic island". Consequently, the utterance's final output is the product of the *combination* of many linguistic elements distributed in the IU, which may consist of sentences, phrases, clauses or different fragments – *linguistic chunks*. This combination, which takes place between "islands", has an informational nature, i.e., it is pragmatically oriented and should not be confused with syntactic compositionality. That is, the syntactic relations of predication, regency, modification, subordination, and coordination are confined within each IU – or syntactic/semantic island. Note, for example, the relationship between the information units of TOP and COM below.

(14) bfamcv01, 11 < porque o Durepox > /=*TOP*= pelo menos jogava bola //=*COM*
because the Durepox / at least knew how to play well //

In (14) there are two information units, TOP and COM, separated from one another by a non-terminal prosodic break ("/" in the transcription). These two units make up the utterance, which is delimited by a terminal prosodic break ("//" in the transcription) in the stream of speech. According to L-AcT, the relationship between TOP and COM in (14) is not established via the apparent predication in which the NP *o Durepox* (in TOP) would be the syntactic subject of the predicate *pelo menos jogava bola* (in COM). According to Cresti (2011), the TOP functions as a linguistic representation of a pragmatic/contextual prominence affecting the interpretation of the locutive content in COM. This TOP-COM relationship is promoted by prosody, as prosody marks the informational value of each unit – the Prefix prosodic profile of the TOP is not the same as that of the Root prosodic profile of the COM. Thus, the TOP unit always functions as an anacoluthon and a semantic island in relation to the COM. Put differently, a syntactic relation in the TOP-COM pattern, as traditionally claimed, only becomes possible if the prosodic level is disregarded.

Accordingly, based on the notion of IU as a syntactic/semantic island, L-AcT recognizes two types of syntactic relation in speech: *linearized syntax* and *patterned syntax*.

(i) *Linearized syntax* describes the relations of coordination and subordination that are established inside the same information unit; relations thus established are properly syntactic. Example (15) shows the elements of a relative clause in COM, i.e., '…N + QUE finite verb… //IU', highlighted in italics.

(15) bfamdl01, 176: cê nũ quer comprar *um trenzinho que espirra* pro seu banheiro não //=COM=
don't you want to buy a little thing that sprays for your bathroom //

(ii) *Patterned syntax* describes the relations of coordination and subordination that are established across different IUs; the nature of such relations is informational and not properly syntactic. Example (16) shows the elements of a relative clause realized in two information units (COM-APC), i.e., '… N /IU + QUE finite verb… //IU', highlighted in italics.

(16) bfammn01, 72: só até aí que eu sei *o caso* /=COM= *que ele me contou* //=APC=
only up to that point do I know the story / that he told me //

Thus, the concept of relative clause in spontaneous speech, in this paper, takes into consideration information structure in speech as defined by L-AcT.

2.2 RELATIVE CLAUSES IN SPONTANEOUS SPEECH: A DEFINITION BASED ON L-ACT

The concept of relative clause traditionally describes a subordinate clause that restricts the reference of a nominal element within a given predication (ROCHA LIMA, 1992; BECHARA, 2004; CUNHA & CINTRA, 2001; NEVES, 2000; PERINI, 2004). Most linguistic approaches identify two strategies for relative clauses: restrictive and non-restrictive ones. Despite their similar morphosyntactic configuration, '… N + QUE finite verb…', at the semantic level restrictive relative clauses delimit the referent,

while non-restrictive ones do *not* delimit the referent[3], as they only add information about it. In this context, only restrictive relative clauses fit the traditional concept of relative clause.

Considering the above observations, either the concept of relative clause should be loosened or the relativization status of non-restrictive clauses should be reconsidered (CRISTOFARO, 2003; KEENAN & COMRIE, 1977). This study opts for the second alternative, considering (i) the cognitive difference between the two types of relative clauses, given that one 'delimits' and the other do not, (ii) the insufficiency of the linguistic signifier regarding the available forms, given that morphosyntactic structures are repeated to express distinct meanings, viz. the phenomenon of homonym (SALOMÃO, 1999). Furthermore, for the definition of relativization in spontaneous speech, the traditional approach does not take information structure into consideration. Therefore, this study considers L-AcT's postulates regarding the workings of syntax in speech (linearization and patterning) as well as the observation of relative clauses in the C-ORAL-BRASIL minicorpus in order to propose a definition for clausal relativization.

Accordingly, the data from the minicorpus show that (i) relative clauses occur both in linearized and patterned syntactic contexts, (ii) relative clauses that show the semantics of restriction in spontaneous speech exhibit linguistic items that evoke a *reference set* that serves as *background* for the interpretation of the N that is restricted/delimited by the relative clause, which has been called *denotation set*[4]. This operation is accomplished via *inference* (LEVINSON, 2007)[5]. Thus, the following conclusions have been reached:

· The relative clause shows the semantics of restriction only when it delimits the interpretation of a referent (NP) that is underspecified in terms of *denotation set*. Thus, the computation of its truth condition is dependent on there being a *reference set* evoked at an underlying semantic level of which the underspecified referent is a subset – just as other elements may be. Otherwise, the semantics of the clause is a *non-restrictive* one. Thus, it does not establish the interpretation of the NP in terms of *denotation set*, considering its specification. Hence, the computation of its truth condition is not dependent on there being a *reference set*.

Accordingly, the interpretation of the clause as being restrictive is dependent on the following semantic-linguistic parameters, given that they establish the underlying reference set:

(i) *The distributive interpretation associated with quantifiers* – distributive phrases yield the following interpretation: "for each of the elements in a set (*x*, *y*, *z*…) there are distinct elements introduced by a quantifier respectively associated with the variables of this set (x, y, z…)".

(ii) *The interpretation of the assumption of the existence of the reference* – referential phrases yield the following interpretation: "for each referential N there is a potential referent whose value of truth or falsehood may be confirmed".

Thus, this study develops the *clausal relativization test*, which consists in verifying whether the clause showing the relativization syntax, within a simple or compound utterance, delimits the interpretation of a referent (NP) in terms of denotation set or not, considering its degree of specification. If it does not delimit the interpretation, it is not a relative clause – here this clause will be referred to as *informative clause*, since it syntactically resembles the relative clause, but it is semantically different. To perform this test, one must check for the presence of quantifying expressions and the presupposition of existence of the referent establishing the reference set. Examples (17) through (20)[6] illustrate what has been said.

[3] In writing, the two strategies are distinguished by punctuation; in speech, by that which the tradition identifies as "prosodic pause". This concept cannot be confounded with prosodic break. For further details, see O papel da pausa na segmentação prosódica de corpora de fala (RASO, MITTIMANN, OLIVEIRA, 2015).

[4] The notions of denotation set and referent set have been derived from Perini (1981) and Lyons (1977).

[5] That is, the deduced implicit content (reference set) is accessed by means of the linguistic expressions from the syntactic surface through the relative clause (denotation set).

[6] Clauses that do not establish the denotation and reference sets in the test are marked with the empty set symbol (Ø).

(17) bfammn02, 51: papai foi *o irmão que mais deu apoio a ele* //=COM
dad was the brother who mostly supported him //

Referent: o irmão (the brother)
Denotation set: irmão que mais deu apoio (brother who mostly supported)
Reference set: set of brothers
Type of clause: relative clause linearized in COM
The relative clause *que mais deu apoio* (who mostly supported) establishes the reference set *irmãos* (brothers) whose possible elements show features that are different from the denotation set *irmão que mais deu apoio* (brother who mostly supported) by means of the adverbial quantifier *mais* (mostly). Therefore, the clause *irmão que mais deu apoio* is true, since there exists a set of brothers.

(18) bfamdl01, 176: cê nũ quer comprar *um trenzinho que espirra* pro seu banheiro não // COM (same as Ex. 15)

Referent: um trenzinho (a little thing)
Denotation set: trenzinho que espirra (little thing that sprays)
Reference set: set of "little things"[7] for the bathroom
Type of clause: relative clause linearized in COM
The relative clause *que espirra* (that sprays) establishes the reference set *trenzinhos para o banheiro* (little things for the bathroom), whose possible elements show features that are different from the denotation set *trenzinho que espirra* (little thing that sprays), by means of presupposition of existence. Therefore, the clause *trenzinho que espirra* is true, since there exists a set of little things for the bathroom.

(19) bfammn06, 33: e *esse caso* /=TOP= *que acontecia* /=APT= marcava muito // COM
and this episode / that was happening / impacted a lot //

Referent: esse caso (this episode)
Denotation set: Ø
Reference set: Ø
Type of clause: informative clause patterned in TOP-APT
The clause *que acontecia* (that happened) does not establish an underlying reference set (other cases?). Therefore, there is no restriction by the denotation set or opposition between the sets. Thus, the clause *que acontecia* only adds information about a specified referent: *esse caso* (this episode). So, its content does not need to be computed in terms of truth condition.

(20) bfamdl02, 73, BAL: que eu dou *um exemplo de porta* /=TOP= *que é excelente* //=COM
I give an example with door / which is excellent //

Referent: *um exemplo de porta* (an example with door)
Denotation set: Ø
Reference set: Ø
Type of clause: informative clause patterned in TOP-COM
The clause *que é excelente* (which is excellent) does not establish an underlying reference set (examples with door?). Therefore, there is no restriction by the denotation set or opposition between the sets. Thus, the clause *que é excelente* only adds information about a specified referent: *um exemplo de porta* (an example with door). So, its content does not have to be computed in terms of truth condition.

To show the syntactic, semantic and informational difference between these two types of clauses (i.e., *linearized relative clause* and

[7] The expression designates "toilet cleaners", which may be solid, jellylike or liquid.

patterned informative clause) in BP spontaneous speech, the following formalization[8] is proposed:

(i) Linearized relative clause

At the syntactic level, it is a type of subordinate clause (B) that functions as adjunct of a NP (x) to which it is linked by the relative pronoun (QUE) inserted in its locutive content. In the utterance, the elements of the relative clause may occur either in a matrix clause (A) – $(NP + [QUE + B]) \subset A(NP + [QUE + B]) \subset A$//; therefore, $B \subset AB \subset A$ // – or isolated from a matrix clause – $(NP + [QUE + B])\ \emptyset$ //) $(NP + [QUE + B])\ \emptyset$ //). Semantically, the function of the relative clause is to restrict/delimit the N (x) within an underlying reference set which is possibly composed of other similar elements. At the prosodic level, the antecedent N and the relative clause occur syntactically linearized within the same information unit within a simple or compound utterance:

Linearized relative clause = $B \subset x\ /\ x \subset A/\ Therefore, B \subset A$ $B \subset x\ /\ x \subset A/\ Therefore, B \subset A$

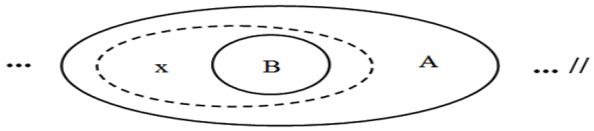

Figure 1: Linearized relative clause

The examples below show linearized relative clauses found in the C-ORAL-BRASIL minicorpus fitting the description here proposed:

(21) bfamdl04, 18: tem [*creme* [*que* $_{RelPro}$ + *nũ dá com o cabelo não*]$_{RelC}$]$_{NP}$ // COM
there is [hair conditioner [that + is not good for the hair]] //

(22) bfamdl01, 177: [*trenzim* [*que* $_{RelPro}$ + *espirra*]$_{RelC}$]$_{NP}$ // COM
[little thing [that + sprays]] //

In (21), the NP *creme que nũ dá com o cabelo não* has a restrictive relative clause, *que nũ dá com o cabelo não*, that functions as adjunct. Since the N *creme* is underspecified, for it has candidate referents (hair conditioner that is good vs. hair conditioner that is not good), the restriction of the relative clause only applies to the element of the reference set that corresponds to *creme 'que não dá com o cabelo'* (hair conditioner *that is not good for the hair*), i.e., the denotation set. At the syntactic level, this NP functions as argument of the matrix clause *tem creme...* At the prosodic level, the items of the relative clause are syntactically linearized within the same information unit, the COM, in a simple utterance.

In (22), the NP *trenzim que espirra*[9] has the restrictive relative clause *que espirra* as adjunct. Since the N *trenzim* is underspecified, i.e. it has candidate referents (*trenzim* that sprays, that melts, that hangs – or the reference set), the restriction of the relative clause only applies to the element of the set *que espirra* (that sprays), i.e., the denotation set. At the syntactic level, it functions in isolation, i.e., without a matrix clause; nevertheless, it exhibits informational autonomy – it is an illocution. At the prosodic level, the items of the restrictive relative clause are syntactically linearized within the same information unit, the COM, in a simple utterance.

[8] The formalization shown here was first developed in a previous study. For more details, see Carmo and Mello (2016, forthcoming).

[9] The expression tremzinho/trenzim designates a type of disinfectant spray.

(ii) *Patterned informative clause*

At the syntactic level, it is a type of appositive clause (B) that semantically intersects the antecedent N (x) of the matrix clause (A) through the relative pronoun (QUE) inserted in its locutive content. The pronoun allows for the anaphoric interpretation of the semantic content of N. However, B is not a subset of A ($B \not\subset AB \not\subset A$) as a syntactic element. In the utterance, the elements of the informative clause can be distributed, apposed within a matrix clause – ($NP\ /[QUE + B]) \subset ANP\ /[QUE + B]) \subset A$ // $\Longrightarrow B(\not\subset)A \Longrightarrow B(\not\subset)A$ //) – or isolated from the matrix clause – ($NP\ /\ [QUE + B])\ \not\subset NP\ /\ [QUE + B])\ \not\subset$ //). Semantically, B only adds information to N ($\subset)\subset)$ A. At the prosodic level, the antecedent N and the informative clause are patterned across different information units within a compound utterance.

Patterned informative clause = $B\ (\cap)\ A = \{x\} B\ (\cap)\ A = \{x\}$ / However, $B\ \not\subset AB\ \not\subset A$

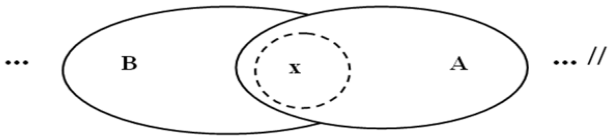

Figure 2: Patterned informative clause

The examples below show informative clauses found in the C-ORAL-BRASIL minicorpus according to the description proposed here.

(23) bfamcv01, 56: tem [*o SESC*]$_{NP}$ / COB [*que é bom pa caramba*]$_{RelC}$ // COM
there is [SESC] [Social Service of Commerce] / *[which is really good]* //

(24) bfamdl02, 30: [*um cuidado*]$_{NP}$ / *que cês têm que tomar*]$_{RelC}$ //COM
[something] / [which you have to be cautious about] //

In (23), the patterned informative clause *que é bom pa caramba* semantically intercepts the matrix clause *tem o SESC* by means of the relative pronoun *que*, which picks up through anaphora the semantic content of the antecedent NP *o SESC* inserted in the matrix clause. However, the informative clause is not a subset of the NP in question as an adjunct. It is rather an appositive within the matrix clause. The informative clause does not delimit the NP, which is already specified; i.e., it does not establish the reference set. Therefore, its function is only to add an extra piece of information about the NP in the matrix clause. At the prosodic level, the informative clause occurs in COM, and the antecedent NP occurs in the COB information unit, in a compound utterance. Thus, its elements are syntactically patterned across different information units.

In (24), the informative clause *que cês têm que tomar* semantically intercepts the referent NP *um cuidado* by means of the relative pronoun *que*, which picks up through anaphora the semantic content of N. However, the patterned informative clause is not a subset of the NP in question as an adjunct. It is an appositive. Thus, the patterned informative clause does not delimit the NP, but it adds a piece of information about the NP; i.e., it does not establish a reference set. Syntactically, the informative clause functions in isolation, i.e., without a matrix clause, although it has informational autonomy; it is an illocution. At the prosodic level, the informative clause occurs in COM and the antecedent NP occurs in TOP as part of a compound utterance. Therefore, they are syntactically patterned across different information units.

For the analysis of the data, this study follows the methodology described below.

3 METHODOLOGY

This study adopts the methodological framework of Corpus Linguistics (MELLO, 2014). Accordingly, a balanced sample of the C-ORAL-BRASIL corpus, the so-called *minicorpus*, is used.

3.1 THE C-ORAL-BRASIL MINICORPUS AND THE IPIC DATABASE

The minicorpus used in this study is derived from the C-ORAL-BRASIL (RASO; MELLO, 2012), a balanced corpus of spoken Portuguese, specifically of the variety spoken in the state of Minas Gerais (Belo Horizonte and metropolitan area). Each of the sessions in the corpus contains audio, transcription and metadata files, as well as text-to-speech alignment. The C-ORAL-BRASIL minicorpus is also organized in the same manner, but it features information structure annotation in accord with terminal and non-terminal prosodic breaks that define the utterance and its internal structure, a central prerequisite of the *Information Patterning Theory* (CRESTI, 2000; MONEGLIA; CRESTI, 2006). The minicorpus has 20 texts, 31,318 words and 5,512 terminated sequences (utterances). It follows the same architecture of the C-ORAL-BRASIL, which means that, regarding context of interaction, it is divided into two domains: family/private and public. Each of these domains is further divided into three interactional types: conversations, dialogues and monologues (RASO, 2012).

The C-ORAL-BRASIL minicorpus can be accessed online through the search platform of the IPIC Database[10] (PANUNZI; GREGORI, 2011). The platform was developed at the LABLITA lab (Research Unit at the Humanities Department of the University of Florence) and it also hosts the Italian informal section of the C-ORAL-ROM[11] corpus and its minicorpus.

4 RELATIVE CLAUSES IN THE C-ORAL-BRASIL MINICORPUS: OUTCOMES

The IPIC platform was used for the survey of relative clauses. Through IPIC's interface, a search was run on the C-ORAL-BRASIL minicorpus to retrieve the utterances hosting the subordinator *que*. Out of the 5,512 utterances of the minicorpus, 1,821 had the subordinator *que*. However, only 148 occurrences of *que* comprised instances of relative pronouns. These data were classified and grouped in spreadsheets according to their syntactic informational structure (linearized or patterned). Thus, two groups were created: (1) relative clauses syntactically linearized within one information unit alone, and (2) informative clauses syntactically patterned in more than one information unit. The counts in Table 2 show the frequency of each group.

Table 2:

Total	Distribution of clauses according to informational syntax in the minicorpus
123	Linearized relative clauses
25	Patterned informative clauses

[10] Information Structure Database: http://lablita.dit.unifi.it/app/dbipic/.

[11] The C-ORAL-ROM is a multilingual speech corpus of the four main European Romance languages: Italian, European Portuguese, French and Spanish. Resulting from a consortium among four European universities, it is coordinated at the LABLITA and funded by the European Union. For more about the C-ORAL-ROM, see Cresti and Moneglia (2005).

148	Occurrences

Nevertheless, both linearized and patterned clauses may occur as part of or isolated from a matrix clause within an utterance. Hence, the following syntactic contexts:

· *Complex syntactic context*: the relative or informative subordinate clause occurs in a matrix clause as one of its arguments. This matrix clause, in terms of locutive content, is distributed inside a simple or compound utterance: "…N [QUE + finite verb…]$_{SubCl}$ + Matrix Cl.…".

(25) bfamdl01,176: cê nũ quer comprar$_{MatrixCl}$ um *trenzinho*$_{N}$ [*que espirra*]$_{SubCl}$ pro seu banheiro não // - *Relative clause* (same as Ex. 15)

(26) bfamcv01,56: tem$_{MatrixCl}$ o *SESC*$_{N}$ [*que é bom pa caramba*]$_{SubCl}$ // *Informative clause* (same as Ex. 23)

· *Simple syntactic context*: the relative or informative subordinate clause does not occur as part of a matrix clause. It functions by itself as the locutive content inside a simple or compound utterance. Hence, it will be referred to as *relativized NP*: "…N [QUE + finite verb…] SubCl – Ø ...".

(27) bfamdl01, 177: *trenzim*$_{N}$ [*que espirra*] $_{SubCl}$ // *Relative clause* (same as Ex. 22)

(28) bfamdl02, 30: um *cuidado*$_{N}$ / [*que cês têm que tomar*] $_{SubCl}$ // - *Informative clause* (same as Ex. 24)

Table 3 shows the frequencies of the informational and syntactic contexts found in the minicorpus.

Table 3:

Total	Distribution of relative clauses according to syntactic-informational context
97	Linearized in complex syntactic context
26	Linearized in simple syntactic context
21	Patterned in complex syntactic context
04	Patterned in simple syntactic context
148	**Occurrences**

As Table 3 shows, 97 occurrences of linearized relative clauses and 21 occurrences of patterned relative clauses were found exhibiting complex syntactic structure, i.e., these are part of a matrix clause "… N [QUE + finite verb…]$_{SubCl}$ + Matrix Cl…". Nonetheless, 30 occurrences of isolated relative clauses with informational function were also found (26 linearized and 4 patterned). They show a simple syntactic structure, i.e., they are not part of a matrix clause: "… N [QUE + finite verb…] SubCl – Ø ...".

As for the semantics of the relative clauses found in the C-ORAL-BRASIL minicorpus, the *test of clausal relativization* was applied and the results are shown in Table 4.

Table 4:

Total	Distribution of clauses according to semantic types
123	Linearized relative clauses
25	Patterned informative clauses
148	**Occurrences**

Table 4 shows that relative clauses occur syntactically linearized inside the utterance, including both those clauses that occur in complex and simple syntactic contexts. These clauses correspond to relative clauses, as they exhibit the semantic function of restriction. Informative clauses, however, occur syntactically patterned inside the utterance, involving both the clauses that occur in complex and simple syntactic contexts. These do not correspond to relative clauses, as they do not exhibit the restrictive semantic function, even though, as highlighted above, they present a syntactic structure like that of relative clauses.

This finding corroborates the postulate by Cresti (2014) regarding relative clauses in Italian: *the restrictive relative clause* corresponds to a syntactic/semantic island, occurring linearized inside only one information unit in the utterance, which can be simple or compound. As for *the non-restrictive relative clause*, it corresponds to the combination of at least two syntactic/semantic islands, occurring in a patterned configuration, i.e., distributed in different information units, each of which realizing its own informational function within a compound utterance.

5 FINAL REMARKS

This study presented the analysis of the relative clauses found in the C-ORAL-BRASIL minicorpus. The data show that, out of the total of 5,512 utterances of the minicorpus, 148 show the syntactic structure of a relative clause. Out of these 148 occurrences, 123 relative clauses occur linearized inside the same information unit either in a simple or compound utterance, whereas 25 of them occur in a patterned configuration across information units inside compound utterances. Nonetheless, out of the 123 linearized relative clauses, 97 occur in complex syntactic context (NP [QUE + finite verb] + Matrix Cl.), while 26 occur in simple syntactic context (NP [QUE + finite verb] Ø). As for the 25 patterned relative clauses that were found, 21 occur in complex syntactic contexts, while 4 occur in simple syntactic contexts.

To sort out the semantics of relative clause – traditionally identified as restrictive and non-restrictive relative clauses – this study proposes the *test of clausal relativization*, which verifies whether a given clause establishes an underlying *reference set* for its interpretation, considering the degree of specification of the referent. The outcome achieved through the application of this test showed that only syntactically linearized relative clauses establish the underlying reference set, considering the presence of quantifiers and/or presupposition of existence. Conversely, syntactically patterned clauses do not establish this underlying set for their interpretation. In light of this finding, it is assumed that, in spontaneous speech, only linearized relative clauses, identified with the semantics of restriction, comprise instances of relativization. Patterned relative clauses, despite their syntactic similarity with linearized clauses, are not actually relative, since the semantics and information structure of such clauses are distinct from those of true relative clauses. Hence, they came to be referred to as *informative clauses*. The results presented in this paper are consistent with linguistic studies that argue that clausal relativization encompasses only the traditionally defined restrictive clauses.

REFERENCES

BECHARA, E. *Moderna gramática portuguesa*. 37ed. Rio de Janeiro: Lucerna, 2004.

BOSSAGLIA, G. Orientação pragmática da sintaxe na fala: uma análise corpus-based da subordinação completiva e adverbial no português do Brasil. *Domínios de Lingu@gem*, v. 9, n. 5 (dez. 2015), p. 309-335, 2015. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/29246> Acesso em: 15 de jul. de 2016.

CARMO, C; MELLO, H. Cláusulas relativas na fala espontânea do português do Brasil: um estudo exploratório baseado no corpus C-Oral-Brasil. *Signum: Estudos da Linguagem*, v. 19, n. 2, 2016. Disponível em: < http://www.uel.br/revistas/uel/index.php/signum/article/view/25097 > Acesso em: 2 jan. 2016.

CINTRA, L.; CUNHA, C. *Nova gramática do português contemporâneo*. 3.ed. Rio de Janeiro: Nova Fronteira, 2001.

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

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*. 1ed. Amsterdam/Philadelphia: John Benjamins, 2014. p. 365-410.

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

CRISTOFARO, S. *Subordination*. Oxford: University Press, 2003.

KEENAN, E. Relative clauses. In: SHOPEN, T. (Ed.) *Language Typology and Syntactic Description*. Cambridge: University Press, 1985, v.2. p. 206-236.

KEENAN, E. L.; COMRIE, B. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, n. 8, p. 63-99, 1977.

HART, J; COLLIER, R; COHEN, A. *A Perceptual Study on Intonation:* An Experimental Approach to Speech Melody. Cambridge: CUP, 1990.

KATO, M. A. et al. As construções-Q no português brasileiro falado: perguntas, clivadas e relativas. In: KOCH, Ingedore V.G (Org.). *Gramática do Português falado VI*. 2.ed.Campinas, SP: Editora da UNICAMP, 2002. p. 309-374.

LYONS, J. *Semantics* 1. Cambridge: Cambridge University Press, 1977.

MELLO, H. What Corpus Linguistics can offer Contact Linguistics: The C-oral-Brasil corpus experience. *PAPIA*, São Paulo, v.24, n.2, p. 407-427, Jul/Dez 2014. Disponível em: < http://revistas.fflch.usp.br/papia/article/view/2204> Acessado em: 15 de jul.2016.

NEGRÃO, Esmeralda V. Forma Lógica e quantificação. In: MULLER, A. l., NEGRÃO, E.V.;

NEVES, M. H.M. *Gramática de usos do português*. São Paulo: 2000. p. 71-95

PANUNZI, A; GREGORI, L. 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, 2012. p. 133-150.

PERINI, M. A. *Gramática descritiva do português*. 4.ed. São Paulo: 2004.

PERINI, M. A. *Um aspecto da interpretação do tópico em português*. Publicação do Curso de Letras do Centro de Ciências Humanas e Letras das Faculdades Integradas de Uberaba. Série Estudos - 7, Uberaba, MG, 1981.

RASO, T. Fala e escrita: meio, canal, consequências pragmáticas e linguísticas. *Domínios de Lingu@Gem*, v. 7, p. 12-46, 2013. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/23730> Acesso em: 20 jul. 2016.

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

SALOMÃO, M.M.M. A questão da construção do sentido e a revisão da agenda dos estudos da linguagem. *Veredas*, v.2, Juiz de Fora: EDUFJF, 1999.

Recebido em 13/09/2016. Aceito em 24/10/2016.