

Relative clauses in a spoken corpus of European Portuguese: identifying the factors determining their variation

Elisabeth Aßmann
Assmann@em.uni-frankfurt.de
Goethe University, Frankfurt (Germany)

Esther Rinke
Esther.Rinke@em.uni-frankfurt.de
Goethe University, Frankfurt (Germany)

ABSTRACT: This paper investigates a range of factors determining variation in European Portuguese (EP) relative clauses on the basis of a corpus study. From a more general perspective, we aim at empirically confirming or disconfirming generalizations formulated on the basis of introspection or standard language grammars. In addition, we seek to identify preferences within the “envelope of variation” in actual language use. We will focus on three variable aspects in EP relative clauses: i) the choice of the relativizer, ii) non-canonical relative clauses (preposition chopping and resumption), and iii) subject-object asymmetries and related information structural aspects. The results confirm that *que* is the default relativizer in spoken EP, and that animacy determines the choice of the relativizer in free relative clauses and headed relative clauses introduced by a preposition. Resumption is attested in various types of relative clauses and seems to be favoured in headed indirect object relative clauses. The relevance of other factors such as indefiniteness or presentational matrix verbs, however, cannot be confirmed. Chopping occurs in restrictive as well as non-restrictive relative clauses and with different prepositions, e.g. temporal prepositions, grammaticalized (“verbal”) prepositions but also in oblique and indirect object relative clauses. Finally, the corpus data provide evidence for a subject-object asymmetry that relates to information structural aspects.

KEYWORDS: Relative clauses, European Portuguese, corpus study, syntactic variation

1. Introduction

As in many languages, relative clauses in European Portuguese can be of various types. Depending on the realization of an antecedent, we distinguish between headed and free relative clauses. Headed relative clauses can be differentiated with respect to their semantic contribution to the reference of the noun phrase, e.g. they can either be restrictive or non-restrictive

(appositive). The various types of relative clauses differ with respect to the types of relativizers that can be employed. There are a number of factors that have been proposed to determine the choice of the relativizer (*que* vs. *o qual* vs. *quem* vs. *cujo*), namely the type of relative clauses (free vs. headed; restrictive vs. non-restrictive), semantic properties of the head noun (human vs. non-human; specific vs. non-specific reference), the determiner accompanying the head noun (definite vs. indefinite article), or the case and the syntactic function of the antecedent (e.g. subject/object vs. indirect object, prepositional phrases). However, it is not always clear how these factors actually interact in the spoken language and whether all the factors are of the same relevance in the vernacular. One aim of this paper is therefore to analyze how the different factors mentioned in the literature determine the choice of the relativizer in different types of relative clauses.

In addition, we want to focus on non-canonical types of relative clauses, namely resumptive relative clauses (including an element doubling the relativizer) and chopping relative clauses (where the preposition of a complex relativizer is skipped). According to the literature, chopping and resumption also depend on a number of factors, which are not yet well understood. It is assumed that the chopping strategy is less likely with oblique and direct object relative clauses than with other prepositional relative clauses and it is assumed to be linked to A' movement and information structural aspects (Cohen 1990). With respect to resumption, it is assumed that island environments favour the occurrence of resumptive elements (Veloso 2007), but also other factors such as indefiniteness of the antecedent, a presentational verb selecting the head noun, and indicative mood of the verb inside the relative clause (Veloso 2013). However, in a previous corpus study by Veloso (2007), some of the potential factors have been investigated, but not all of them could actually be proved to be relevant. In our corpus study, we include more data and more of the potential factors in order to enlarge the empirical basis and to widen the perspective. We put the different factors to test in order to find out whether they can be shown to determine chopping or resumption in EP.

The third aspect that will be investigated is related to the well-known subject-object asymmetry that has been identified in a number of studies

on processing and language acquisition (Lipka et al. 2000, Friedmann et al. 2009, Costa et al. 2011), where it has been shown that subject relative clauses are preferred over object relative clauses in processing and that they occur earlier in acquisition. In recent work, we have suggested that the subject-object asymmetry is related to information-structure, namely to the fact that a head noun modified by a relative clause is typically an entity newly introduced into the discourse whereas the relativizer is a topic about which the rest of the relative clause adds a comment (Rinke & Aßmann 2017). Our corpus study confirms that a subject-object asymmetry on these two levels (head noun and relativizer) also exists in actual language use.

The paper is structured as follows. In section 2, we present the background of the study and identify the relevant research questions that will guide our investigation. Section 3 provides information about the methodology and the results of the corpus study. In section 4, we discuss the results with respect to our research questions. Section 5 summarizes and concludes.

2. Background and research questions

2.1. Types of relative clauses and choice of the relativizer

Depending on the realization of an antecedent, relative clauses (RC) divide into headed and non-headed relative clauses. A headed RC (1a.) modifies a noun, the antecedent, while a free RC occurs without a head noun and represents the argument by itself (1b.):

- (1) a. Ligou [[um homem]_{antecedente} que não conheço.]_{RC}
called a man rel not I.know
'A man called whom I don't know.'
- b. Todos sabemos [o que temos de fazer.]_{RC}
all we.know the rel we.have of to.do
'We all know what we have to do.'

Headed relative clauses traditionally divide into restrictive relative clauses (RRCs) and non-restrictive or appositive relative clauses. RRCs restrict the extension of the concept expressed by the antecedent, i.e. they are intersective modifiers, whereas non-restrictive relative clauses (NRRCs)

represent a proposition on their own and have a parenthetical character (cf. Heim & Kratzer 1998:88, Brito & Duarte 2003:670)¹.

Headed relative clauses and free relative clauses differ with respect to the occurrence of the relativizer *que*. In headed relative clauses (1a.), *que* occurs without a determiner, whereas in free relative clauses, *que* is accompanied by the masculine singular definite article “o” (1b.). When the argument represented by the free relative clause is referring to a [+human] referent, the relativizer of free relative clauses is *quem* instead of *o que*.

Animacy also plays a role in headed relative clauses in the sense that *quem* occurs with [+human] antecedents in relative clauses introduced by a preposition. However, as shown in (2), other relativizers like *o qual* and *que* are also possible in this context.

- (2) Ligou o homem [com que] / [com o qual] / [com quem] falaste na festa.
called the man with rel / with the rel / with rel you.spoke in.the party
'The man with whom you talked at the party called.'

In fact, *que* is possible in virtually all contexts and sometimes even the only possible relativizer, namely in restrictive subject and object relative clauses.

- (3) a. Vi o homem que / *o qual / *quem roubou a tua carteira.
I.saw the man rel stole the your wallet
'I saw the man who stole your wallet.'
b. O homem que / *o qual / *quem vi ontem é o ladrão.
the man rel I.saw yesterday is the thief
'The man that I saw yesterday is the thief.'

In non-restrictive subject and object relative clauses (4a.-b.) and in restrictive and non-restrictive indirect object (4.c), genitive, possessive (4.d) or prepositional (4e.) relative clauses, *que* alternates with *o qual*.

- (4) a. A minha avó, [que] / [a qual] faz anos amanhã, mora em Lisboa.
the my grandmother rel / the rel makes years tomorrow lives in Lisbon
'My grandmother, whose birthday is tomorrow, lives in Lisbon.'

¹ There have been reasons to believe that this binary differentiation does not do justice to the reality of the semantics of relative clauses: Cinque (2008) argues for a further subdivision of the class of non-restrictive RCs, while Kleiber (1987) proposes to distinguish “specifying” RCs from the restrictive RCs in the stricter sense. For the present investigation, however, this detailed distinction is not important.

- b. A minha avó, [que] / [a qual] vou visitar em Lisboa, faz anos amanhã.
the my grandmother rel /the rel I.go to.visit in Lisbon makes years tomorrow
'Tomorrow is my grandmother's birthday, whom I will visit in Lisbon.'
- c. O idiota [a que] / [ao qual] / [a quem] emprestei esse livro
the idiot to rel / to.the which / to whom I.lent that book
nunca mais mo devolveu.
never more me.it returned
'The idiot to whom I lent that book never gave it back.'
- d. Encontrei o rapaz [de que] / [do qual] / [de quem] conheço as primas.
I.met a boy of rel / of.the which / of whom I.know the cousins
'I met a boy whose cousins I know.'
- e. É esta a caneta [com que] / [com a qual] escrevi a carta.
is this the pen with rel / with the which I.wrote the letter
'This is the pen with which I wrote the letter.'

Which factors determine the choice between *que* and *o qual* in these contexts is not yet well understood. It seems to be clear that *que* is the more frequent form in colloquial speech whereas *o qual* may be preferred in formal registers. Veloso (2013:2091) proposes that *o qual* is inherently specific (as can be seen by the definite determiner preceding the Q-morpheme). She assumes that, because of this specification, this relativizer is always possible in NRRCs, because their antecedent is necessarily [+specific]. However, as shown in example (5), RRCs with non-specific head nouns and the subjunctive mood marking non-specificity also allow for *o qual* as a relativizer. Therefore, it is questionable whether *o qual* is indeed inherently specific.

- (5) No meu prédio não mora nenhum homem do qual o meu cão não tenha medo.
in.the my building not lives none man of.the which the my dog not has.subj fear
'There does not live any man in my building whom my dog is not afraid of.'

In genitive relative clauses, the relativizer *cujo* may occur in addition to *que*, *o qual* and *quem*. However, as shown in (6), in the contexts where *cujo* occurs, it alternates with *o qual* and *quem* but not with *que*.

- (6) Encontrei um rapaz [cujas primas] / [*as primas de que] / [as primas do qual] /
I.met a boy whose cousins / the cousins of rel / the cousins of.the which /
[as primas de quem] conheço do instituto.
the cousins of whom I.know of.the school
'I met a boy whose cousins I know from school.'

In example (6), the relative clause is introduced by a complex consisting of the genitive attribute and the relativizer: [*cujas primas*]/ [*as primas do qual*]/ [*as primas de quem*]. It seems that *que* cannot be part of a complex relativizer whereas *o qual* and *quem* can be and *cujo* always is. *Que* is only admitted in genitive relative clauses if the genitive attribute is stranded (cf. 7; see also Veloso 2013: 2083, fn. 33)².

- (7) Nunca mais funcionou bem o telefone [de que] / [do qual] o gato roeu o fio.
never more worked well the telephone of rel / of.the which the cat bit the wire
'The telephone whose wire the cat gnawed at never worked well again.'

Finally, the relativizer *quanto* occurs exclusively with quantified antecedents, e.g. the universal quantifier *tudo*, "everything" (cf. ex. 8a., from Brito & Duarte 2003: 664). Also forms of *todo*, "every/all X" and *tanto*, "so much/many X" are possible (cf. examples 8b. from Brito & Duarte 2003:664, and 8c. from Veloso 2013:2100).

- (8) a. Pensei muito acerca de tudo quanto disseste.
I.thought much about of everything how.much you.said
'I thought a lot about everything that you said.'
- b. Ela trouxe todos quantos encontrou.
she brought all.masc.pl how.many.masc.pl she.found
'She brought all that she found.'
- c. Vou comprar tantos quantos me apeteça.
I.go to.buy so.many how.many to.me like.subj
'I will buy as many as I like.'

In 8b. and 8c., *quantos* is supposed to head a nominal group erased by ellipsis. For this reason, it agrees in number and gender with the referent, exactly as the quantificational antecedent *todos* and *tantos*, respectively. Given its intrinsic quantificational character, *quanto* cannot occur in appositive RCs.

Table 1 gives a schematic overview of the distribution of relativizers in EP canonical relative clauses.

² It seems to be the case that the acceptability of these complex relativizers in restrictive relatives is subject to individual variation: while Veloso (2013) considers them grammatical, Peres & Mória (1996) find them odd, exactly as an anonymous reviewer.

TABLE 1:³ Possible relativizers in EP.

RC		relativizer		<i>o que</i>	<i>o qual</i>	<i>quem</i>	<i>cujo</i>	<i>onde</i> ⁴	<i>quanto</i>	<i>que</i>
		free	+hum	-	-	+	-	-	-	-
				+	-	-	-	+	-	-
SU/DO	RRC	+hum	-	-	-	-	-	-	+	+
		-hum	-	-	-	-	-	-	+	+
	NRRC	+hum	-	+	-	-	-	-	-	+
		-hum	-	+	-	-	-	-	-	+
IO/OBL	RRC	+hum	-	+	+	-	-	-	+	+
		-hum	-	+	-	-	-	+	+	+
	NRRC	+hum	-	+	+	-	-	-	-	+
		-hum	-	+	-	-	-	+	-	+
GEN ⁵	RRC	+hum	-	+	+	+	-	-	+	+
		-hum	-	+	-	+	-	-	+	+
	NRRC	+hum	-	+	+	+	-	-	-	+
		-hum	-	+	-	+	-	-	-	+

2.2. Non-canonical relative clauses: resumption and preposition chopping

In addition to the above mentioned canonical relativization structures, there are two non-canonical relativization strategies worth citing, the *estratégia resumptiva* (resumptive strategy) and the *estratégia cortadora* (chopping strategy) (cf. Tarallo 1985, Brito 1995, Brito & Duarte 2003: 666, among many others). Although they do not form part of the standard

³ SU = subject; DO = direct object; IO = indirect object; GEN = genitive; OBL = oblique case; free = free relative clause; RRC = restrictive relative clause; NRRC = non-restrictive relative clause; P = preposition; N = noun. The morphosyntactic feature matrixes refer to the antecedent's characteristic. Note that there exist some other, more complex relativizer-like forms in EP, so-called relative "pro-forms" (Velooso 2013:2101) such as *onde* (where), *como* (how), *quando* (when):

- a. *Visitei uma casa onde tinha vivido o Lorca.*
 I.visited a house where had lived the Lorca
 "I visited a house where Lorca had lived."
- b. *Não gostei do modo como ele se comportou.*
 not I.liked.of.the way how he REFL behaved
 "I did not like the way he behaved."
- c. *Os alunos contestaram a data para quando o exame foi marcado.*
 the students questioned the date for when the exam was set
 "The students questioned the date for which the exam was set."

Como and *quando* have only recently been admitted as part of the relativizer paradigm in EP (cf. Peres & Mória 1995, Mória 2001, among others). For the purpose of this paper, these complex forms are not relevant because, on the one hand, they do not show any particularities (in the case of *onde*) and, on the other hand, some of them do not occur in our data base (*como*, *quando*).

⁴ The distribution of *onde* refers only to oblique cases, not to the indirect object.

⁵ The use of *que* in genitive RCs is restricted to non-complex contexts, i.e., where the genitive attribute is stranded, cf. exs. (6) and (7).

language, these relativization strategies are considered to occur frequently in colloquial speech, irrespective of the speaker's education or social status (cf. Brito 1995, Brito & Duarte 2003, Veloso 2013), and can even occur in journalistic and literary texts (Peres & Mória 1995)⁶. Apart from EP, these structures are reported to be particularly frequent in Brazilian Portuguese (BP) as well as in African Portuguese (AP) varieties (cf. Brito 1995, Silva & Lopes 2007, Alexandre & Hagemeyer 2013).

In resumptive relative clauses, a resumptive element, which is coreferential with the antecedent and the relativizer, occurs in a position inside the relative clause where one would expect a gap:

- (9) Que é **uma pronúncia** cantada **que** eu própria que sou de cá
 what is a pronunciation sung rel I myself rel I.am of here
 não **a** sei muito bem dizer.
 not cl3.sg.acc.fem I.know very well to.say
 'Which is a singing pronunciation that I myself, being from here, don't know how to say it very well.' (i.e. don't know how to imitate it)
 (Brito & Duarte 2003:667)

Alexandre (2000) reports that resumption is possible in subject, direct and indirect object, oblique, and genitive relative clauses. The resumptive elements can be stressed (strong) and unstressed (clitic) pronouns, demonstratives, adverbs, or even the whole antecedent. Most authors agree that the more complex the construction, the easier it is to find a resumptive element inside the RC (Brito 1995, Peres & Mória 1995, Brito & Duarte 2003, Veloso 2013). In particular, complexity in form of an island environment favours resumption, as has been argued for EP as well as several other languages (McCloskey 1990, Prince 1990, Suñer 1998, Bianchi 2004, Veloso 2007, among others). Example (10), taken from Veloso (2007:2), illustrates such an island context:

- (10) Há coisas fantásticas que eu nem sei como é que elas me acontecem.
 there.are things fantastic rel I not.even know how is that they to.me happen
 "There are fantastic things that I don't even know how they happen to me."

⁶ However, some authors see a link between education and the chosen relativization strategy, as the quantity of canonical structures increases over time (cf. Valente 2008, Duarte 2011, Alexandre & Hagemeyer 2013). Duarte (2011) therefore considers canonical RCs to be a structure of late acquisition.

In (10), the relative clause contains a *wh*-question introduced by *como*, “how”. Within this *wh*-island, the resumptive element *elas* occurs, which is coreferent with the antecedent *coisas fantásticas*. The assumption that island contexts favour resumption has been justified by assuming that in resumptive RCs, movement of the relativizer to its clause-initial position is somehow not possible⁷ and a complementizer-like element (*que*) is therefore introduced into the relative clause CP.

Apart from structural complexity, Veloso (2013:2130ff) gives four more conditions which are said to favour resumption:

- a. the indefiniteness of the antecedent;
- b. a presentational verb introducing the head noun (e.g. *haver* “to exist”, *ser* “to be” and *ter* “to have”);
- c. an assertive value of the RC (i.e. indicative mood);
- d. the complexity of the relativizing constituent and its syntactic function.

These generalizations are based on a small corpus study, which the author conducted and discussed in earlier work (cf. Veloso 2007). In the other part of that work, however, where the author investigates RCs in a spontaneous speech corpus, generalization b., i.e. the antecedent being a complement of a presentational verb, does not hold, as the majority of RCs that show one of these verbs is formed in a canonical way. With respect to the choice of the relativizer, Veloso (2013) states that *que* is not the only possible element in resumptive relatives, but that also more complex relativizers, like *onde* “where”, can co-occur with a resumptive element, like *lá* “there”.

In preposition chopping constructions, complex relativizers are “cut” and substituted by mere *que* (11).

- (11) A pessoa **que** te falei é minha amiga. (≈ canonical: a pessoa de que te falei)
the person rel to.you I.spoke is my friend
‘The person that I talked to you about is my friend.’
(Brito & Duarte 2003:666)

⁷ Cf., for example, Alexandre (2000) for EP, Suñer (1998) for Spanish. There is, however, an ongoing debate regarding (non-)movement in resumptive relative structures. See e.g. Boeckx (2003) for an overview.

The chopping strategy of relativization involves simplification on two levels: first, the preposition is deleted; second, the most underspecified *que* (in contrast to e.g. *o qual*) is used systematically (Veloso 2013:2128). According to Veloso (2013), the chopping strategy is especially frequent if the relativized constituent has a temporal or local value (cf. 12);

- (12) Porque, cada dia **que** nós íamos,... ≈ cada dia em que
 because every day rel we went...
 'Because, every day that we went there...'
 (C-ORAL-ROM via Veloso 2013:2128)

the involved preposition is of pure grammatical nature, i.e. *de* "of" or a "to", selected by verbs such as *falar* "to speak", *gostar* "to like", or *precisar* "to need" (cf. 13):

- (13) ...há coisas na cultura Americana que eu realmente não gosto.
 there.is things in.the culture American rel I really not like
 'There are things in the American culture that I really don't like.'

In other words, prepositions may only be deleted in relative clauses when they do not add new information to the meaning of the verb with which they co-occur, as Cohen (1990:112) argues: for example, in the expression *falar de*, "to talk of/about", the preposition is very likely to be deleted, while this is not the case for *falar a favor de*, "to speak in favour of", where the preposition is more complex.

The chopping strategy is argued to be less frequently employed for the following contexts: indirect objects (14a.), oblique objects whose preposition has a semantic value (14b.), as well as oblique complements which are relativized but modify themselves a noun inside the RC (*o sítio* in (14c.) that is relativized but also serves as a modifier of *à procura* (*à procura do sítio*) – examples (14) are taken from Veloso (2013:2128-2129):

- (14) a. Nós estávamos com uns portugueses que vir a Lisboa ao fim-de-semana
 we were with some Portuguese rel to.come to Lisbon at.the weekend
 não dá muito jeito. ≈ portugueses a quem
 not gives much knack
 'We were with some Portuguese for whom coming to Lisbon at the weekend is not very convenient.'

- b. Continente é a marca que os portugueses mais confiam
Continente is the brand rel the Portuguese more trust
na sua categoria. ≈ a marca na qual / em que
in.the its category
'Continente is the brand the Portuguese trust most.'
- c. lam-nos levar ao sítio que nós andávamos à procura. ≈ ao sítio de que
they.went-us to.bring to.the place rel we went to.the search
'They brought us to the place that we were looking for.'
(examples from C-ORAL-ROM via Veloso 2013:2129)

2.3. Subject-object asymmetries

Regarding information-structural aspects of relative clauses, many authors have argued that antecedent and RC are in an aboutness-relationship insofar as the antecedent is the topic and the RC a proposition about this topic (Schachter 1973, Kuno 1976). However, Lambrecht (1994: 129-130) argues that this only holds necessarily for the relation between the proposition expressed by the RC and the referent of the head noun, while the head noun itself may well be a focus expression in the discourse.

In previous work (Rinke & Aßmann 2017), we have argued along these lines that the head noun preferentially represents an entity that is newly introduced into the discourse. As such, it tends to occur in object position instead of subject position, at least in restrictive relative clauses, where the relative clause then serves to specify an entity that is newly introduced into the discourse. However, the relativizer itself, which introduces the relative clause and refers back to the head noun, represents a topic to which the rest of the relative clause adds a comment (cf. also Lambrecht 1994). Hence, we argue that the subject-object asymmetry is found on two levels – with respect to the relativizer and with respect to the head noun. We interpret this distribution (head noun = focus / relativizer = topic) as being one possible explanation for the well-known subject-object asymmetries in relativization. It has been repeatedly shown in previous studies that in language processing or acquisition, subject relatives are preferred over object relatives (Gouvea 2003 for BP; Lipka et al. 2000 for German; for acquisition, among others: Costa et al. 2011 for EP; Friedmann et al. 2009 for Hebrew). With respect to processing, for instance, it has been argued that subject A'-dependencies, i.e. subject RCs, are easier to understand and/or to process than object A'-dependencies, i.e. object RCs. The question is whether this asymmetry can be confirmed in spoken language.

2.4. Research questions

On the basis of what has been said so far, we will focus on the following research questions. The first bundle of questions relates to the choice of the relativizer in different contexts:

1. How systematic is the distribution of *que* and *quem* with respect to animacy in free and headed relative clauses?
2. To what extent does specificity determine the distribution of *que* and *o qual* in restrictive and non-restrictive relative clauses? Is it the case that the antecedent of *o qual* is always specific?
3. How systematic is the distribution of the different forms with respect to the syntactic function of the relative clause? E.g. can we confirm that subject and object relative clauses differ in a systematic way from other types of relative clauses?
4. Which forms of relativizers occur as complex relativizers in the corpus? Is *que* excluded from these contexts? Can we identify factors that determine the distribution of the different forms in complex relativizers?

The second aspect to be investigated concerns the conditions determining the occurrence of non-canonical relative clauses.

5. Can we confirm that island environments favour the occurrence of resumptive relative clauses (Veloço 2007)?
 - 5a. What is the role of other factors such as definiteness of the antecedent, occurrence of a presentational verb or indicative mood (Veloço 2013)?
 - 5b. Do we find relativizers different from *que* in resumptive relative clauses?
6. Is preposition chopping more likely with local and temporal RC and with certain types of prepositions (“purely grammatical prepositions”)? Is it less likely with oblique and indirect object relative clauses than with other prepositional RC?
7. Is chopping only attested in restrictive relative clauses?

The third part of the investigation focuses on the subject-object asymmetry and information-structural aspects:

8. Can we identify a subject-object asymmetry? Are subject relative clauses indeed more frequent than object relative clauses in spontaneous speech?
 - 8a. Do we find evidence that the head noun typically relates to new information focus?
 - 8b. Do restrictive and appositive relative clauses behave alike with respect to the subject-object asymmetry and information structural aspects?

In the next chapter, we will present the methodology used for this study before we will provide answers to the questions in section 4 by presenting the results of our corpus study.

3. Methodology

The data were taken from two corpora: the *Syntax-oriented Corpus of Portuguese Dialects* (*Corpus Dialectal para o Estudo da Sintaxe*, CORDIAL-SIN), provided by the Universidade de Lisboa (Martins 2000-), and the *Perfil Sociolinguístico da Fala Bracarense* (BRAGA), provided by the Universidade do Minho (Barbosa 2011-2014). Both corpora consist of interview transcripts, annotated in different ways, from which we manually extracted relative clauses and coded them for morphological and syntactic criteria.

The CORDIAL-SIN corpus aims at delivering an empirical basis for the study of dialectal variation in contemporary European Portuguese. It offers a 600.000 word corpus, including interviews from more than 200 localities in the Portuguese territory, including the archipelagos of Madeira and the Azores. As informants, elderly locals were chosen who were born and raised in the respective locality and had little education. The transcriptions are downloadable in four different annotations: normalized orthographic transcripts; verbatim transcripts, including standard linguistic expressions as well as pauses, hesitations, etc.; morphological annotations, providing POS as well as inflectional items; and syntactic annotations (cf. Martins 2000-).

The BRAGA corpus concentrates on the variety of contemporary Portuguese spoken in the city of Braga, in the north of Portugal. The collected data, 80 interviews with a length of 60 minutes each, are controlled for three sociolinguistic variables, gender, age and education. The transcripts follow semiorthographic rules (Barbosa 2011-2014).

As for the morphosyntactic annotation for the relative clauses we collected out of the transcripts, we mostly chose criteria from the corpus *Atlante Sintattico d'Italia, ASIt*, which cover a wide range of properties, but we also created new criteria. A selection of the tags we used is as follows:

- properties of the relative clause such as restrictive/appositive/free;
- properties of the relativizer, e.g. its syntactic function;
- properties of the antecedent such as definiteness, gender, number, animacy;
- properties of clitics within the RC such as Case, doubling, position with respect to the verb;
- properties of the verb in the relative clause as well as in the matrix clause, e.g. Mood, inflection, negation.

For a complete listing of the tags, cf. the ASIt corpus (Benincá & Poletto 2007) and the ASCRP corpus (Rinke 2016-). The manually extracted clauses were coded according to the previously determined criteria, afterwards a native speaker checked the coding with respect to the correct interpretation of the sentences. In this way, we collected 1563 sentences of the CORDIAL-SIN and 350 sentences of the BRAGA corpus, giving us a total of 1913 coded sentences.

In an ongoing collaboration with the ASIt people, the tagged sentences were finally uploaded to a searchable online databank, which forms part of the project website ASCRP, “Annotated corpus of the DFG project *Synchronic and diachronic analysis of the syntax of Italian and Portuguese relative clauses*” (cf. Rinke 2016-; Agosti et al. 2016).

It has to be mentioned that the corpora that served as the data-base for our study contain information on dialectal variation: the CORDIAL-SIN corpus provides data from all over the Portuguese territory, including the archipelagos of the Azores and Madeira, while the BRAGA corpus is concentrated on the variety of the city and the surrounding area of Braga, in

the north of continental Portugal. Interestingly, though, dialectal variation did not play any role with respect to the phenomena discussed here. Therefore, it is not mentioned in the following sections. All the affirmations are equally valid for the Portuguese dialects, at least concerning our empirical findings.

4. Results of the corpus investigation

4.1. The distribution of different types of relativizers

Our corpus consists of a total of 1913 relative clauses, extracted from the spoken speech corpora mentioned in the previous section. Among them, we find 353 free relative clauses and 1560 headed relative clauses (1279 restrictive and 281 appositive relative clauses).

In a first step, we focussed on the distribution of *que* and *quem* with respect to animacy in free relative clauses and in headed relative clauses introduced by a preposition. Table 2 gives an overview over the occurrence of the relativizers *que* and *quem* with respect to the animacy of the respective referent in free as well as in headed (i.e. restrictive and non-restrictive) relative clauses.

TABLE 2: Distribution of the relativizers (*o*) *que* and *quem* with respect to animacy

	free [+ human]	free [-human]	headed (+ prep) [+ human]	headed (+ prep) [-human]
(<i>o</i>) <i>que</i> ⁸	17 (10.6%)	191 (99.5%)	0	45 (97.8%)
<i>quem</i>	143 (89.4%)	1 (0.5%)	6	1 (0.2%)
Σ	160 (100%)	192 (100%)	6 (100%)	46 (100%)

As expected, we find a division of labour in free relative clauses: there are only 17 examples with *o que* referring to [+human] referents in free relative clauses with 143 examples of *quem* in the same context. With non-human referents, there is only one example with *quem* (15b.). In all other sentences with non-human referents, we find the relativizer *o que*.

In example (15a.), the referent is human and realized by the relativizer *quem*. Example (15b.) shows the only example of a free relative clause with an inanimate, non-human referent, namely a *stone-like element* used to mark where one owner’s field ends and the other one’s starts.

⁸ In combination with a quantifier as an antecedent, *que* can be accompanied by the masculine singular determiner *o*. In such contexts, *o que* generally refers to non-human referent.

- (15) a. Há **quem** diga que isso foi feito ali por causa disso .
there.is who says.subj that this was done there for reason that
'There are people who say that this was done because of that.'
- b. Há **quem** esteja já feito, enterram com cimento
there.is who is.subj already made they.bury with cement
e depois está ali pregado.
and then is there nailed
'There is one which is already finished, they bury it with cement and then it is fixed there.'
(Context: marking fields of different owners with stone-like markings)
(ASCRP, Alcochete)

In (16), we provide examples of free relative clauses with *o que*. In (16a.) the referent is non-human, in (16b.) the referent is human.

- (16) a. Pois à noite **o que** está em cima é estrelas.
because at.the night the rel is in top is stars
'Because what is above at night are stars.'
(ASCRP, Alvor)
- b. E **o que** trabalha por sua conta, esse já é proprietário.
and the rel works for his account this already is owner
'And whoever works on his own account, that one is already the owner.'
(ASCRP, Enxara do Bispo)

In headed relative clauses, *quem* also occurs with [+human] antecedents. In our corpus, there are 5 examples of *quem* accompanied by a preposition in restrictive relative clauses (e.g. 17a.). In one case, *quem* occurs with a pronominal antecedent and without a preposition (17b.). We will come back to example (17b.) below. There is only one example with a non-human antecedent and a relative clause introduced by *quem* (17c.).

- (17) a. Porque a pessoa a **quem** ela foi contar contou-me a mim.
because the person to whom she was to.tell told-me to me
'Because the person to whom she told it was me.' (ASCRP, Braga)
- b. Depois sou eu **quem** cozo a louça.
after I.am I who I.boil the dishes
'Then it is me who bakes the dishes.' (ASCRP, Graciosa)
- c. Porque Nossa Senhora, quando nasceu o Menino Jesus,
because our lady when was.born the child Jesus
a urze riaga foi com **quem** tapou-o também
the heather 'riaga' was with who she.covered-him also
'Because Our Lady, when Baby Jesus was born, it was 'riaga' heather with which she covered him as well.' (ASCRP, Outeiro)

However, (17c.) is an exception and probably a mistake. In headed relative clauses with [-human] antecedents introduced by a preposition, *que* is employed (cf. 18).

- (18) porque é um curso **em que** uma pessoa entra com uma média muito baixa
because is a course in rel a person enters with an average very low
'Because this is a course in which you enter with a very low mark.' (ASCRP, Braga)

To sum up, we found that the generalization widely holds that in free relative clauses as well as in headed relative clauses introduced by a preposition, a [+human] referent leads to a realization of the relativizer as *quem* whereas a [-human] referent relates to (*o*)*que*. However, this division of labour seems not to be strict, at least not with respect to free relatives (cf. 16b).

In a second step, we were interested whether we can detect a correlation between a specific interpretation of the antecedent and the occurrence of the relativizer *o qual* in the database. However, *o qual* is virtually non-existent in our corpus. All of its functions seem preferentially to be expressed by other relativizers, especially by *que*. There is only one example with *o qual* in a total of 1913 sample sentences, cf. (19).

- (19) A gente fazia-lhe uma cavidade aqui na madeira para a boca encaixar
the people made-him a cavity here in.the wood for the mouth to.fit
e era tudo pregado era com cravetes,
and it.was all nailed it.was with spikes
o qual desta forma é mais fácil, com menos despesa.
the which of.this form is more easy with less expenses
'We made a cavity here in the wood so that the mouth piece would fit and everything
was fixed with spikes, which is easier this way and less expensive.'
(context: building some agricultural tool)
(ASCRP, Graciosa)

In (19), *o qual*, heading an appositive RC, refers to the whole proposition. In this context, specificity does not play a role. Therefore, it is not possible to make a statement about a potential specificity restriction on its antecedent. *Que* occurs with all sorts of antecedents. In general, the antecedent is interpreted as specific, also because of its combination with the relative clause that turns the indefinite noun phrase into a specific entity (20a.). But

there are also some cases in which *que* refers to a non-specific antecedent with the finite verb in the relative clause in subjunctive mood (20b.).

- (20) a. Era para aproveitar a farinha **que** ficava.
was for make.use the flour rel stayed
'This was for making use of the flour that was left over.'
(ASCRP, Outeiro)
- b. Mas não houve assim nenhum sítio **que** me fizesse confusão.
but not had such no place rel me make.subj confusion
'But there was no such place that would confuse me.'
(ASCRP, Braga)

In a third step, we concentrated on the question whether subject and object relative clauses differ in a systematic way from other types of relative clauses. According to the literature, subject and object RCs are different from other types of relative clauses insofar as they are more restricted with respect to which kind of relativizers they allow: restrictive subject and object RCs are said to only allow for *que*, while in their non-restrictive counterparts, only *que* and *o qual* are possible. Table 3 shows the distribution of relativizers with respect to their syntactic function, i.e. subject and object, in headed RCs.

TABLE 3: Distribution of relativizers in subject and object relative clauses

	restrictive		non-restrictive		Σ
	SU	DO	SU	DO	
<i>que</i>	759	308	209	45	1321
<i>o que</i>	3	5	-	-	8
<i>quanto</i>	2	4	-	-	6
<i>quem</i>	-	-	1	-	1
Σ	764	317	210	45	1336

Our results show that indeed, *que* is the default relativizer for subject and object RCs: the majority of these contexts show *que*, and there are only some exceptional cases that show another type of relativizer.

First, note that neither *o que* nor *quanto* are excluded from subject and object positions. As mentioned before, *que* accompanied by the masculine determiner *o* is the usual relativizer for headed RCs in which a quantifier is relativized (cf. footnote 7). For restrictive RCs, we found two instances of subject-*quanto* (21a.) and four instances of object-*quanto*

(21b.). As is expected from this quantifying relativizer, it necessarily has a quantifier as antecedent, namely *tudo*, “everything” or a form of *todo*, “every X”.

- (21) a. Tudo **quanto** ali está, eu acho bem feito.
all how.much there is I think well done
‘All that is there, I think it is well done.’
(ASCRP, Carrapatelo)
- b. Levou as pitas todas **quantas** cá tínhamos.
took the breads all how.much there we.had
‘He took all the bread that we had.’
(ASCRP, Larinho)

Interestingly, our corpus shows one example of *quem* in what seems to be a subject relative clause. This is unexpected because *quem* is said to be excluded in subject and object position for restrictive and non-restrictive RCs. In our example, *quem* occurs with a personal pronoun as antecedent, cf. example (17b.), repeated here as (22):

- (22) Depois sou eu **quem** cozo a louça.
then I.am I who I.cook the dishes
‘Then it’s me who washes the dishes.’
(ASCRP, Graciosa)

At a closer look, however, (22) does not represent a typical restrictive subject relative clause, but rather a wh-cleft sentence, which has been analysed as including a (pseudo)relative clause (cf. Costa & Duarte 2001, who analyse wh-clefts as a small clause configuration, in which the DP “head noun”, *eu* in (22), moves via *scrambling* to an adjunction position left of the “relative clause”; cf. also Brito & Duarte 2003). Interestingly, however, (22) shows person agreement between the antecedent and the relative clause-internal verb: this is not possible in the standard language, however it seems to be an option in vernacular Portuguese (cf. Lobo 2006).⁹

Our fourth research question focussed on the occurrence of complex relativizers in the corpus. Unfortunately, complex relativizers consisting

⁹ In a strict sense, (22) is not a restrictive relative clause as it does not open two sets whose intersection is then the referent. Since the antecedent is a 1st person personal pronoun, it does not need any more reference in order to be identifiable. The subordinated clause adds more of an additional comment to the already established referent, an operation which reminds appositive relative clauses. Therefore, despite possible structural differences, we decided to treat it as an appositive relative clause in Table 3.

of the genitive attribute and the relativizer [*cujas primas*]/ [*as primas do qual*]/ [*as primas de quem*] do not occur at all in our corpus because the genitive/possessive context is absent overall.¹⁰ Therefore, we cannot answer the question whether *que* is indeed excluded from this construction in spoken language.

What our corpus data clearly confirm is the fact that *que* has become the default relativizer in almost all contexts and that *o qual* is virtually absent in the spoken language, although the contexts for realizing this relativizer are clearly present (appositives, prepositional restrictives, etc.).

TABLE 4: Distribution of complex relativizers in comparison to *que*¹¹:

RC		relativizer		<i>o que</i>	<i>o qual</i>	<i>quem</i>	<i>cujo</i>	<i>onde</i> ¹²	<i>quanto</i>	<i>que</i>	Σ
		free									
		+hum	17	-	143	-	-	-	-	-	160
		-hum	184	-	1	-	-	4	-	-	189
SU/DO	RRC	+hum	-	-	-	-	-	-	-	321	321
		-hum	8	-	-	-	-	-	6	727	741
	NRRC	+hum	-	-	1	-	-	-	-	50	51
		-hum	-	1	-	-	-	-	-	186	187
IO/OBL	RRC	+hum	-	-	5	-	-	-	-	7	12
		-hum	-	-	1	-	-	12	-	153	166
	NRRC	+hum	-	-	-	-	-	-	-	3	3
		-hum	-	-	-	-	-	5	-	16	21
GEN	RRC	+hum	-	-	-	-	-	-	-	-	-
		-hum	-	-	-	-	-	-	-	1	1
	NRRC	+hum	-	-	-	-	-	-	-	1	1
		-hum	-	-	-	-	-	-	-	1	1
Σ			209	1	151	-	-	21	6	1466	1854

¹⁰ However, there are instances of RCs with a genitive interpretation that involve a chopping structure – cf. the next section.

¹¹ This table does not include 22 RCs whose antecedent is a whole clause, 12 RCs whose context makes it impossible to decide whether the antecedent is animate or inanimate (usually because the antecedent is a demonstrative or a quantifier and the respective referent is not recorded, cf. (i)), 25 RCs with a presentational or predicational relativizer (cf. (ii)).

- (i) *Há tantos que têm aí!*
there.is so.many rel they.have there
“There are so many that they have there!” (i.e. so many different ones)
(ASCRP, Alcochete)
- (ii) *Isso era o tempo mais triste que pode haver.*
this was the time more sad rel can there.to.be
“This was the saddest time that there could be.”
(ASCRP, Graciosa)

¹² The distribution of *onde* refers only to oblique cases, not to the indirect object.

However, in some sub-contexts, *que* also seems to be avoided. As shown above, in free relative clauses and in headed relative clauses introduced by a preposition, human antecedents are related to the relativizer *quem*, while *o que* and *que* are employed whenever the antecedent is [-human]. This confirms the statement by Kato & Nunes (2009) that *quem* is in fact a [+human] realization of *que* in those contexts.

The relativizer *onde* is only possible in an adverbial position, and can only refer to inanimate referents. In all these cases, however, *onde* can be replaced by *que* when it is selected by a preposition or in a chopping context (see section 4.2.).

Free and headed relative clauses differ with respect to the presence/absence of the determiner *o*. In free relative clauses, the [-human] relativizer is *o que*, in headed relative clauses, we find exclusively *que*. Only relative clauses with a quantifier as antecedent represent an exception to this rule. In combination with a quantifier as antecedent, we find the relativizer *o que* instead of *que* in our data base.

- (23) a. Olhe, se for a explicar **tudo o que** dá o linho.
look.subj if I.was.subj to explain all the rel gives the flax
'Look, if I were to explain everything that the flax gives...'
(ASCRP, Monsanto)
- b. Come o caranguejo, **tudo o que** apanhar.
eats the crab all the rel it.catches.subj
'It eats crabs, everything (i.e. every animal) that it might catch.'
(context: description of a seagull)
(ASCRP, Alvor)

The next paragraph focuses on the properties of resumptive relative clauses and relative clauses with preposition chopping in our corpus.

4.2. Non-canonical relative clauses: resumptive relative clauses and preposition chopping

Table 5 gives an overview over the properties of resumptive relative clauses in the corpus.

TABLE 5: Properties of resumptive relative clauses in the corpus

Σ	Island contexts	type RC			form of relativizer	type of resumptive element			syntactic position of resumptive element				antecedent [+def]	antecedent object of a presentational verb	indicative mood inside the RC
		rest	app	free		clitic	strong	PP	S	O	IO	PP			
48	0	29	18	1	free: o que non-free: que	37	7	4	7	5	31	5	15	0	47

The first observation is that there are no island contexts at all in resumptive constructions and there are no relativizers other than (o) *que*. The usual context is the one exemplified in (24), i.e. a simple structure without any island:

- (24) a. Fui a um senhor que chamam-lhe o Arlindo.
I.went to a gentleman rel they.call-3.sg.dat the Arlindo
'I went there to a gentleman whom they call Arlindo.'
(ASCRP, Outeiro)
- b. E a pedra de cima, que lhe chamamos a mó, é colocada ali em cima daquela segurelha.
and the stone from above rel 3.sg.dat we.call the millstone is placed there on top of.this savory
'And the stone from above that we call the millstone is placed on top of the savory.'

Also the other factors mentioned by Veloso (2013) do not clearly favour resumption, at least not in our data base. At first glance, the in/definiteness of the antecedent seems to be relevant to some extent, given that twice as much relative clauses with resumption have an indefinite antecedent (ex. 24a), as a definite antecedent (ex. 24b). However, given that indefinite contexts are in general much more frequent than antecedents with a definite determiner, the percentage of resumption in each context remains pretty much the same (definite contexts: 749 / resumption 15 = 2%; indefinite contexts: 811 / resumption 32 = 3.9%). Our corpus does also not provide an example with a presentational verb in the matrix clause and resumption in the relative clause, although presentational verbs are quite frequent in relative clauses (cf. section 4.3). Finally, indicative mood in the relative clause might favour the occurrence of resumption, since 47 of 48 of the resumptive contexts show indicative mood. However, subjunctive seems not to be excluded and it has again to be considered that indicative mood is the default mood and subjunctive is very rare in relative clauses: in our data, only 199 out of 1913 relative clauses show subjunctive mood.

In our data base, resumption occurs preferentially in IO-RC (cf. 25a.), although it is also possible with other syntactic positions of the relativizer such as the subject (25b.).

- (25) a. Ajeitei uma garrafa dele a uma pessoa
 I.arranged a bottle of.it to a person
 que lhe o médico receitou essa.
 rel 3.sg.dat the doctor prescribed this
 'I arranged to give a bottle of it to a person to whom the doctor prescribed this.'
 (ASCRP, Monsanto)
- b. Eu tive dois professores que eles não eram contra as praxes.
 I had two professors rel they not were against the customaries
 'I had two professors who were not against the *praxes* (initiation rites for joining a students' society).'
- (ASCRP, Braga)

This tendency is especially striking given the fact that, in the whole corpus data, there is no instance of a canonical IO-RS *without* resumption– while the other syntactic positions are variable. An indirect object relativizer, hence, seems to favour resumption.

Table 6 shows the semantic properties of relative clauses involving chopped and overt prepositions in our corpus. The prepositions were pooled according to their interpretation in order to find out whether chopping is related to the interpretation of the respective relative clause. This expectation is not confirmed. For a more detailed view on the types of prepositions, cf. Table 8.

TABLE 6: Distribution of prepositions in PP-RCs, overt and chopping

type P \ overtness	chopping	overt	Σ
temporal	43 (76.8%)	13 (23.2%)	56 (100%)
indirect	37 (97.4%)	1 (2.6%)	38 (100%)
oblique	17 (51.5%)	16 (48.5%)	33 (100%)
verbal	27 (93.1%)	2 (6.9%)	29 (100%)
local	17 (58.6%)	12 (41.4%)	29 (100%)
genitive	2 (100%)	-	2 (100%)
partitive	2 (100%)	-	2 (100%)
Σ	145 (76.7%)	44 (23.3%)	189 (100%)

The types of prepositions in Table 6 are ordered according to the number of occurrences of their respective interpretation.

The table shows that preposition chopping occurs frequently with indirect objects: in a total of 38 occurrences, 37 instances show preposition chopping. Also verbal prepositions are preferably chopped, with 27 instances or 93.1% showing no overt preposition. Temporal and local prepositions follow with 76.8% and 58.6% of chopping. As for oblique objects, there is no clear tendency either way.

With respect to verbal prepositions, i.e. prepositions which are part of the verb, consider Table 7, which shows a detailed overview of the types of verbal prepositions that underwent chopping:

TABLE 7: Distribution of verb types in verbal chopping RCs

Σ	gostar de	ser/estar habitado a	lembrar-se de	recordar-se de	esquecer-se de	estar ligado a	perguntar por	precisar de	recorrer a	ser curioso por	concordar com
27	12	4	3	1	1	1	1	1	1	1	1

As can be seen, all verbal prepositions that underwent chopping are monosyllabical, with *de* and *a* being the most frequent ones. This is in accordance with Veloso (2013).

Table 8 shows which prepositions are chopped in local, temporal and oblique RCs. Here, the most frequently involved prepositions are *em*, *de* and *com*.

TABLE 8: Distribution of preposition types in local/temporal and oblique chopping RCs and PP-RCs

	chopping				Σ	overt P			Σ	Σ total
	local	temporal	oblique			local	temporal	oblique		
<i>em</i>	14	42	-	56	12	13	3	28	84	
<i>para</i>	2	0	-	2	-	-	-	-	2	
<i>por</i>	1	0	-	1	-	-	-	-	1	
<i>de</i>	-	-	9	9	-	-	-	-	9	
<i>com</i>	-	-	5	5	-	-	13	13	18	
<i>contra</i>	-	-	1	1	-	-	-	-	1	
<i>a</i>	-	-	1	1	-	-	-	-	1	
<i>por</i>	-	-	1	1	-	-	-	-	1	
Σ	17	42	17	76	12	13	16	41	117	

It is difficult to derive conclusions from table 8 because some of the prepositions are not very frequent. The most frequent prepositions are *em* (in) and *com* (with); *em* is chopped in 56 out of 84 clauses (67%), but *com* only in 5 out of 18 occurrences (28%).

As a last point concerning preposition chopping, we considered the distribution of chopping in different types of relative clauses. Table 9 shows that preposition chopping is not only a phenomenon of restrictive relative clauses, but also of non-restrictive and free RCs. Nevertheless, restrictive relatives represent the majority of chopping occurrences.

TABLE 9: Distribution of chopping with respect to the type of RC

Σ	restrictive	non-restrictive	free
145	123 (84.8%)	19 (13.1%)	3 (2.1%)

Interestingly, coming back to Table 6, there are two types of prepositional contexts that only occur in form of a chopped RC: genitive and partitive constructions, which are exemplified in (26a.) and (26b.), respectively:

- (26) a. E havia outra, que está lá em baixo agora a mãe...
 and there.was another rel is there in below now the mother
 ‘And there was another one, whose mother is now down there...’
 (Castro Laboreiro)
- b. Está atravessada pelas duas vias que uma é de Barcelos
 is crossed by.the two roads rel one is of Barcelos
 e a outra é do Porto.
 and the other is of.the Porto
 ‘It is crossed by two roads, of which one comes from Barcelos and the other comes from Oporto.’
 (Braga)

As reported in section 4.1., there are no instances of canonical RCs with a genitive and partitive interpretation, respectively, whose form could be something like ...*outra, cuja mãe está lá...* for (26a.) and ...*duas vias das quais uma é...* for (26b.). It seems to be the case, then, that genitive and partitive RCs are especially prone to be formed with a chopping structure. However, since there are only so few instances in total, it is hard to make a definite statement in this regard.

On the basis of our corpus data, we cannot exactly confirm the asymmetry proposed by Veloso (2013), which could be represented as in (27): the types of prepositions further on the left of the scale are typically more easily chopped than those on the right side of the scale.

- (27) temporal/local > grammatical oblique > indirect > semantic oblique

In our data, as shown in Table 6, indirect objects are the most frequent elements to be chopped, followed by verbal, temporal, local and oblique objects (leaving aside genitive and partitive contexts). Our scale, hence, looks like this:

(28) indirect > verbal > temporal > local > oblique

4.3. Subject-object asymmetries and information focus

Table 10 gives an overview over the distribution of different types of relative clauses according to the syntactic function of the relativizer.

TABLE 10: Distribution of the syntactic function of the relativizer with respect to the types of RCs

relativizer type of RC	SU	DO	IO	pred/pres	advPP	Σ
restrictive	765 (59.8%)	318 (24.8%)	21 (1.6%)	18 (1.4%)	157 (12.3%)	1279 (100%)
non-restrictive	210 (74.7%)	45 (16%)	15 (5.3%)	2 (0.7%)	9 (3.2%)	281 (100%)
free	198 (56.1%)	128 (36.3%)	1 (0.3%)	10 (2.8%)	16 (4.5%)	353 (100%)
Σ	1173 (61.3%)	491 (25.7%)	37 (1.9%)	30 (1.6%)	182 (9.5%)	1913 (100%)

The table shows that we find indeed a subject-object asymmetry in the data. Subject relative clauses are much more frequent than object relative clauses, esp. in restrictive and non-restrictive headed relative clauses. In free relative clauses, the difference is less expressive. As mentioned above (cf. 2.3), we relate the subject-object asymmetry in relative clauses to information structural dependencies: in headed RCs, the head noun preferentially represents an entity that is newly introduced into the discourse and the relativizer refers back to the head noun adding information. Therefore, the relativizer shows a tendency to represent a prototypical topical element (subject) and the antecedent a prototypical new information constituent (object).

If this interpretation is on the right track, we expect to find a strong tendency for RC antecedents to be objects, arguments of presentational verbs and postverbal subjects. Table 11 confirms this expectation.

TABLE 11: Position of the antecedent with respect to the verb¹³

position \ type RC		RRC	NRRC	Σ
postverbal	SU	47	11	58
	O	248	83	331
	presentational	572	86	658
	PP	189	43	232
Σ postverbal		1056 (84.9%)	223 (86.4%)	1279
preverbal	SU	131	29	160
	O	19	1	20
	presentational	1	-	1
	PP	36	4	40
Σ preverbal		187 (15%)	34 (13.2%)	221
Σ total		1243 (100%)	257 (100%)	1500

Although subject antecedents still show a tendency to occur preverbally (which is expected because this is the neutral position for subjects), the overall picture shows a clear tendency for the antecedent of a relative clause to represent new information. The overall number of postverbal antecedents is much higher than the number of preverbal antecedents, in restrictive (1056 postverbal vs. 187 preverbal instances) as well as non-restrictive relative clauses (223 postverbal vs. 34 preverbal instances).

The most frequent type of antecedent are arguments of presentational verbs, constituting 659 out of 1500 instances, or 43.9%, combining pre- and postverbal occurrences. This sentence structure typically serves to introduce a new referent into the discourse.

The fact that the most typical kind of antecedent in our data is a postverbal object or presentational constituent seems to speak in favour of the assumption that relative clauses typically represent a device to introduce a new referent into the discourse and then turn it into a topic by predicating over this referent. The subject-object asymmetry can thus be found on two levels: with respect to the relativizer, subjects are much more frequent, while with respect to the

¹³ 60 sentences could not be included here due to the lack of a matrix verb and, therefore, the impossibility to decide whether the antecedent is pre- or postverbal, e.g.:

(i) *Depois o pessoal que ia conosco, fomos comer todos.*
 then the personnel rel went with.us we.went to.eat all
 "Then the personnel who went with us, all of us went to eat."

head noun, (presentational) objects are more common. The sequence head noun = focus, relativizer = topic has been confirmed by our data as well.

5. Summary and conclusion

The aim of this paper was to investigate the role of different factors determining the variation in European Portuguese relative clauses on the basis of a corpus study in order to verify or falsify generalizations taken from the literature, which were mainly formulated on the basis of introspection or, to a lesser extent, to earlier corpus studies (e.g. Veloso 2007). We focused on three main aspects, viz. 1. the choice of the relativizer in different contexts, 2. the conditions determining the occurrence of non-canonical structures, i.e. resumption and preposition chopping, and 3. subject-object asymmetries and information-structural aspects.

As for the first issue, the choice of the relativizer, we saw that in the spoken language, *que* is indeed the default relativizer in almost all contexts. Nevertheless, it seems to be true that animacy or, more specifically, humanhood is a decisive feature: in free and headed prepositional RCs, there is a strong tendency to use *quem* when the referent is human, and a form of *que* when the referent is inanimate. Exceptions from this tendency are attested but have a certain “slip of the tongue”- character, especially if *quem* refers to an inanimate referent. Another important factor for the choice of the relativizer is its syntactic function: not all forms are possible in all functions. In restrictive subject and object relative clauses, as expected, *que* is the prevailing form. However, our data also showed instances of *o que* and *quanto* whenever the antecedent is a quantifier. Other complex quantifiers are virtually (in the case of *o qual*) or actually (in the case of *cujo*) absent in colloquial European Portuguese.

With respect to the second issue, the conditions determining the occurrence of non-canonical structures, we have seen that resumption and chopping are frequent structures in our data base, which was expected for a corpus of colloquial speech. Yet, not all generalizations made in the literature (Veloso 2007, 2013) could be confirmed: on the one hand, neither island environments nor definiteness of the antecedent nor a presentational verb in the matrix clause nor indicative mood seem to promote resumption. In our data base, a decisive factor was case: all examples of indirect object relative clauses occurring in

the data base showed resumption. Indirect object relative clauses are also a preferential context for preposition chopping. The same seems to be true for verbal prepositions but not necessarily for temporal and local prepositions.

Finally, the third issue concerns information-structural matters. The assumption/expectation that subject relative clauses are more frequent than object relative clauses in spontaneous speech is supported by our data. Furthermore, there is a second subject-object asymmetry on the level of the antecedent, which represents more frequently an argument of a presentational verb, an object or a postverbal subject. This could be interpreted in a way that the head noun typically encodes new information, which the relativizer resumes and about which the rest of the relative clause comments. For free relative clauses, finally, this tendency is not as clear. Since free relatives lack a head noun, at least overtly, it seems quite possible that they differ from their headed counterparts not only syntactically, but also with respect to information-structural aspects.

To sum up, the generalizations and predictions made from earlier works seem to find affirmation from our empirical data, by and large. The only field that is not as clear concerns the non-canonical structures, resumption and preposition chopping. It seems to be the case that the conditions on the occurrence of these structures are yet to be closer examined, possibly by an even greater corpus study. This could also shed some light on further research on the structural side of the phenomena.

REFERENCES

- Agosti, M., Di Buccio, E., Di Nunzio, G. M., Poletto, C. & Rinke, E. 2016. Designing a long lasting linguistic project: the case study of ASIt. In: Calzolari, N.; Choukri, K.; Declerck, T.; Grobelnik, M.; Maegaard, B.; Mariani, J.; Moreno, A.; Odijk, J.; Piperidis, S. (Eds.) *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Paris: ELRA, 4479-4483.
- Alexandre, N. 2000. *A estratégia resumptiva em relativas restritivas do Português Europeu*. Master thesis, Universidade de Lisboa.
- Alexandre, N. & Hagemeyer, T. 2013. Estratégias de relativização de PPs no mundo luso-atlântico: crioulo de base lexical portuguesa e variedades do português. In: Moura, M. D.; Sibaldo, M. A. (Eds.) *Para a história do português brasileiro. Volume*

- III – Sintaxe comparativa entre o português brasileiro e línguas crioulas de base lexical portuguesa*. Maceió: EDUFAL, 49-71.
- Barbosa, P. (coord.) (2011-2014) *Perfil Sociolinguístico da Fala Bracarense*, Braga, Universidade do Minho. URL: <https://sites.google.com/site/projectofalabracaense/>
- Benincà, P. & Poletto, C. 2007. The ASIS Enterprise: A View on the Construction of a Syntactic Atlas for the Northern Italian Dialects. *Nordlyd. Monographic issue on Scandinavian Dialects Syntax*. 34(1). URL: <http://asit.maldura.unipd.it/>
- Bianchi, V. 2004. Resumptive relatives and LF chains. In Cinque, G. (Ed.) *Functional structure in DP and IP: the cartography of syntactic structures*. Oxford: Oxford University Press.
- Brito, A. M. 1995. As orações relativas nas variantes culta e oral em quatro línguas românicas, com especial incidência em português. *Lusorama*.27: 70-81.
- Brito, A. M. & Duarte, I. 2003. Orações relativas e construções aparentadas. In: Mateus, M., Brito, A. M., Duarte, I., Hub Faria, I., Frota, S., Matos, G., Oliveira, F., Vigário, M. & Villalva, A. (Eds.) *Gramática da Língua Portuguesa*. Lisbon: Caminho, 653–694.
- Cinque, G. 2008. Two types of non-restrictive relatives. In: Bonami, O. & Cabredo-Hofherr, P. (Eds.) *Empirical issues in syntax and semantics 7*. Paris: Presses de l'Université Paris-Sorbonne, 99-138.
- Cohen, M. A. 1990. *Syntactic change in Portuguese: relative clauses and the position of the adjective in the noun phrase*. PhD. Dissertation, Universidade Estadual de Campinas.
- Costa, J. & Duarte, I. 2001. Minimizando a estrutura: uma análise unificada das construções de clivagem em português. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa: APL. 627-638.
- Costa, J., Lobo, M. & Silva, C. 2011. Subject-object asymmetries in the acquisition of Portuguese relative clauses: adults vs. children. *Lingua*.121: 1083-1100.
- Duarte, I. 2011. Modo oral e modo escrito, estruturas sintáticas de desenvolvimento tardio e escolarização. In: Valente, A. & Pereira, T. (Eds.) *Língua Portuguesa. Descrição e Ensino*. Rio de Janeiro: Parábola, 15-30.
- Friedmann, N., Belletti, A. & Rizzi, L. 2009. Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua*.119: 67-88.
- Gouvea, A. C. 2003. *Processing syntactic complexity: cross-linguistic differences and ERP evidence*. PhD.Dissertation, University of Maryland.
- Heim, I. & Kratzer, A. 1998. *Semantics in generative grammar*. Malden: Blackwell.
- Kato, M. & Nunes, J. 2009. A uniform raising analysis for standard and nonstandard relative clauses in Brazilian Portuguese. In: J. Nunes (Ed.) *Minimalist essays on Brazilian Portuguese*. Amsterdam/Philadelphia: John Benjamins, 93-120.
- Kleiber, G. 1987. *Relatives restrictives et relatives appositives: une opposition "introuvable"?* Tübingen: Niemeyer.
- Kuno, S. 1976. Subject, theme, and the speaker's empathy – a reexamination of relativization phenomena. In: Li, C. N. (Ed.) *Subject and topic*. New York: Academic Press, 417-444.

- Lambrecht, K. 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge: University Press.
- Lipka, S., Kopp, F. & Pechmann, T. 2000. Referential context effects for subject/object relative-clause ambiguities: The role of the Reading Span. In: Schroeger, E., Mecklinger, A. & Friederici, A. D. (Eds.) *Working on working memory*. Leipzig: Leipzig University Press, 121-138.
- Lobo, M. (2006). Assimetrias em construções de clivagem do português: movimento vs. geração na base. *XXI Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 457-472.
- Martins, A. M. (coord.) (2000-) *CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe*. Lisboa, Centro de Linguística da Universidade de Lisboa. URL: <http://www.clul.ul.pt/en/resources/411-cordial-corpus>
- Móia, T. 2001. Aspectos sintáctico-semânticos das orações relativas com *como* e *quando*. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*. 349-361.
- McCloskey, J. 1990. Resumptive pronouns, A'-binding and levels of representation in Irish. In: Hendrick, R. (Ed.) *The syntax of Modern Celtic languages. Syntax and semantics* 23. San Diego: Academic Press, 199-248.
- Peres, J. & Móia, T. 1995. *Áreas críticas da língua portuguesa*. Lisboa: Caminho.
- Prince, E. 1990. Syntax and discourse: a look at resumptive pronouns. In: Hall, K. (Ed.) *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: University of California, 482-497.
- Rinke, E. (coord.) (2016-) ASCRP: Annotated corpus of the DFG project *Synchronic and diachronic analysis of the syntax of Italian and Portuguese relative clauses*. University of Frankfurt. URL: <http://ims.dei.unipd.it/websites/portuguese-relclauses/search.html>
- Rinke, E. & Aßmann, E. 2017. The syntax of relative clauses in European Portuguese. Extending the Determiner Hypothesis of Relativizers to relative *que*. *Journal of Portuguese Linguistics*. 16(4):1-26.
- Schachter, P. 1973. Focus and relativization. *Language*. 49: 19-46.
- Silva, B. & Lopes, C. 2007. O papel da frequência na gramaticalização no português do Brasil. *Veredas*. 11/1: 80-100.
- Suñer, M. 1998. Resumptive restrictives relatives: a crosslinguistic perspective. *Language*. 74(2): 335-364.
- Tarallo, F. 1985. The filling of the gap: pro-drop rules in Brazilian Portuguese. In: King, L. D. & Maley, C. A. (Eds.) *Selected papers from the XIII Linguistic Symposium*. Amsterdam: Benjamins, 355-375.
- Valente, P. 2008. Produção de frases relativas em alunos do terceiro ciclo do ensino básico e do ensino secundário. MA thesis, Universidade de Lisboa.
- Veloso, R. 2007. Orações relativas em contextos representativos: dados de um corpus oral. In: *Comemorações dos 75 anos do CLUL*. URL: http://www.clul.ulisboa.pt/files/rita_veloso/RVeloso_clul75anos_rev.pdf
- Veloso, R. 2013. Subordinação relativa. In: Paiva Raposo, E., Bacelar do Nascimento, M. F., Coelho da Mota, M. A., Segura, L. & Mendes, A. (Eds.) *Gramática do Português*. Lisboa: Fundação Calouste Gulbenkian, 2061-2136.

