COMPLEMENTIZER DOUBLING
IN EUROPEAN PORTUGUESE

Salvador Mascarenhas*
St Catherine’s College, Oxford

1. Introduction

Since the establishment of the “cartographic” enterprise in generative syntax (Rizzi 1997; Cinque 1999), left periphery phenomena in the Romance languages have been particularly influential in motivating theoretical proposals about the structure of the C-domain. For example, the intricate structure of topicalized and focused phrases in Romance languages was one of the main motivations for Rizzi’s (1997) proposal as well Benincà and Poletto’s (2004) proposed revision.

Romance languages might prima facie appear to be far from an ideal testing ground for cartographic explorations, given that they typically have a relatively small catalog of complementizers. These sorts of functional items are naturally of special significance to the cartographer: complementizers typically populate the heads of the functional projections the theorist is proposing, and thus can be instrumental in motivating novel projections and investigating the boundaries between them. For example, work on the left periphery of, in particular, isolating languages (see, for example, Aboh 2004, 2006) has provided further motivation for the split-C hypothesis, as opposed to, say, Chomsky’s (1995) multiple specifiers proposal. Romance languages, though they may appear to be less exciting than isolating languages in this regard, can display a number of phenomena involving functional heads in the left periphery, most notably complementizer deletion, (what seem to be) doubly filled COMP violations, and complementizer doubling. This article is concerned with the latter.

C-doubling (sometimes referred to, especially in the literature on European Portuguese, as “recomplementation”) is a characteristic of, among others, the Northern Italian dialects Turinese and Ligurian (Paoli 2007), spoken (Castilian) Spanish (Demonte and Soriano 2007), Galician, and certain variants of European Portuguese. The most well-known data are surely those pertaining to the Italian dialects, and it is fair to say that the Iberian data, and the Portuguese data in particular, have been largely overlooked, with some noteworthy exceptions (Uriagereka 1995; Barbosa 2000). This paper aims to contribute to filling that gap.

* This article is a minimally updated version of an unpublished manuscript from 2007, which has been available for download from my webpage since 2008. The changes with respect to the 2007 version are numerous, but almost exclusively concern matters of style. I thank Guglielmo Cinque and Luigi Rizzi for the opportunity to revisit this work, and Guglielmo Cinque for discussion of the earlier manuscript. I am very grateful to Manuela Ambar and João Peres for extensive discussions of this work back in 2007.
As it turns out, C-doubling in European Portuguese (henceforth, EP) has some rather surprising and interesting characteristics that have remained, to the best of my knowledge, unnoticed by the literature, and that seem to set the phenomenon apart from its correlates in other Romance languages. Most notably, in EP there seems to be no syntactic constraint on the number of Cs doubled (provided that appropriate phonologically realized material is between every two instances of C), as can be seen in (1), and doubling is possible with the interrogative complementizer ‘se’ (if / whether) — see (2).

(1) Acho que amanhã que a Ana que vai conseguir acabar o trabalho.
    I think that tomorrow that Ana that will manage to finish the assignment.
    ‘I think tomorrow Ana will manage to finish the assignment.’

(2) Não sei se o João se já chegou.
    not I know if João if already arrived.
    ‘I don’t know if João already arrived.’

I begin by focusing on declarative C-doubling, and noting some of the most relevant differences with respect to Paoli’s (2007) Northern Italian data, so as to offer a comparison with better-known C-doubling data. I then discuss the two properties that single out the EP system from other Romance languages: the possibility of having more than two complementizers in a doubling construction, and the possibility of doubling the interrogative complementizer. While the aim of this article is primarily descriptive, I conclude with some theoretical remarks.

2. C-doubling in European Portuguese

2.1 EP declarative C-doubling

Unlike the case of Turinese and Ligurian, discussed at length by Paoli (2007), C-doubling in EP is insensitive to mood in the embedded verb. In those Northern Italian dialects, doubling is only possible if the embedded verb is in subjunctive mood, and the author argues that the type of matrix verb has no influence whatsoever in legitimizing or barring the construction. The examples in (3)–(5), with an indicative, subjunctive and conditional embedded verb, show that embedded verb mood has no clear influence in the Portuguese construction.

(3) O João disse que a Maria que vai chegar atrasada.
    João said that Maria that will-INDEF arrive late.
    ‘João said that Maria will arrive late.’

1 More recent scholarship (after the time of writing of the original version of this article, in 2007) has steadily contributed toward reversing this trend, with several articles focusing on complementizer doubling in Iberian Romance. See for example Villa-García (2012) for Spanish and Ribeiro and Torres Morais (2012) for Old Portuguese.
(4) Duvido que a Ana que goste de ópera.
   I doubt that Ana that likes-SUBJ opera
   ‘I doubt Ana likes opera.’

(5) Acho que se lhe ligasses que tudo se resolveria.
   I think that if him/her called that everything REFL solve-COND
   ‘I think if you called him/her everything would turn out fine.’

It is not entirely clear to what extent, if at all, the class of matrix verb selecting the CP plays a role in allowing for doubling. It is uncontroversial that matrix epistemic verbs, such as the ones in the three sentences above, are perfectly compatible with doubling, and the same seems to be true of non-epistemic predicates such as ‘esperar’ (to hope), the deontic verb ‘exigir’ (to demand), both subcategorizing subjunctive mood in the embedded verb, or ‘prometer’ (to promise), with indicative:

(6) Espero que a Ana que traga o livro.
   I hope that Ana that bring-SUBJ the book
   ‘I hope Ana brings the book.’

(7) Prometo que a carta que chega amanhã.
   I promise that the letter that arrives-IND tomorrow
   ‘I promise the letter will arrive tomorrow.’

As for factive verbs, the situation appears more complex. ‘Lamentar’ (to regret) seems to reject C-doubling quite strongly (all my informants deemed sentences like (8) quite bad), while sentences with ‘reparar’ (to notice) are perfectly fine.

(8) ?? Lamento que o Filipe que tenha chumbado o exame.
   I regret that Filipe that have-SUBJ failed the exam
   ‘I’m sorry that Filipe failed the exam.’

(9) Reparei a semana passada que os miúdos que chegam sempre atrasados.
   I noticed last week that the kids that arrive-IND always late
   ‘I noticed last week that the kids always come in late.’

Interestingly, a factive verb that carries evaluative content such as ‘lamentar’, requiring a subjunctive embedded verb, rejects C-doubling, while non-evaluative factives such as ‘reparar’, ‘observar’ (to observe) and so on, which call for indicative mood in the embedded verb, accept doubling. Notice that both dimensions, evaluation and factivity, seem to play a role. A verb like ‘esperar’ (to hope) clearly carries evaluative information regarding the embedded proposition and selects subjunctive mood, but it does not pattern with ‘lamentar’, plausibly because it is not factive.

For the time being, I can do no more than to notice this contrast, so I will disregard factive verbs in general for the remainder of this article, and will proceed as if the matrix verb had no
influence on whether C-doubling is allowed. I now turn to the issue of what can occur between the complementizers. This matter the existing literature on EP does consider, though not in sufficient detail, as I hope to show.

In examples (3)–(9) I have shown that DP subjects, adverbs and if-clauses can all occur in the relevant position. We can also observe direct and indirect objects in that same position, with or without a resumptive clitic:

(10) Acho que esse livro que a Ana já (o) leu.
    I think that DET book that Ana already CL-ACC read
    ‘I think that book, Ana already read.’

(11) Disseram-me que ao João que o professor (lhe) deu um dezoito.
    they told me that to João that the professor CL-DAT gave a eighteen
    ‘I heard the professor have João an A.’

Data similar to these prompted Uriagereka (1995) and Barbosa (2000) to argue that topics, left dislocated (LD) and clitic left dislocated (CLLD), can occur in this position. As these authors remark, quantified phrases cannot occur between the two complementizers:

---

2 It is worth remarking, as far as selection by a matrix predicate is concerned, that C-doubling is also fine in embedded sentences selected by predicative nouns and adjectives, such as the following:

(i) Tenho a certeza que a Maria que vai chegar a horas.
    I have certainty that Maria that will arrive at hours
    ‘I’m sure Maria will arrive on time.’

The same sentence with the preposition ‘de’ governing the embedded sentence (that is, “Tenho a certeza de que a Maria que vai chegar a horas”) is extremely odd. It seems very plausible that that is due to the fact that C-doubling is a characteristic of colloquial EP and of conservative EP dialects spoken mainly in rural areas, while the use of the preposition ‘de’ in these constructions rings very formal. The sentence as a whole thus becomes incoherent with both strongly formal and strongly colloquial elements.

One final remark concerning selection is in order. Ambar (1992, 2005) noticed that ‘que’-headed sentences with a subjunctive verb are quite acceptable as (at least apparently) matrix sentences under a strong exclamative / imperative reading:

(ii) Que venham as chuvas!
    that come the rains
    ‘May the rain come!’

As Ambar notes, these kinds of utterances are far from common, and I for one do not use them productively at all. Moreover, I find them very hard to interpret if the verb is not an unaccusative or unergative, and even with these kinds of verbs the sentences become considerably worse with a pre-verbal subject (?? “Que as chuvas venham!”)

In any event, C-doubling seems to be completely out, turning what to most speakers my age think is an odd sentence into an unintelligible one. This point further distances the EP data from the Italian case considered in the previous section. In Turinese and Ligurian, constructions exist that are perfectly analogous to (ii), and C-doubling is indeed allowed in those contexts.
Complementizer doubling in European Portuguese

(12) * Eu acho que ninguém que leu esse livro.
I think that no one that read that book

(13) * Eu acho que muitas pessoas que leram esse livro.
I think that many people that read that book

(14) * Eu acho que alguém que leu esse livro.
I think that someone that read that book

This fact again separates EP C-doubling from Turinese and Ligurian, where quantified phrases, including negative quantifiers such as *no one*, can occur, under certain readings, between the complementizers. In EP, such constructions are sharply ungrammatical.

Uriagereka (1995) goes a step further than Barbosa (2000)³, proposing that only topics can occur between the two complementizers. He argues that the second ‘que’ is the head of a functional projection, which he calls FP, and he stipulates that, in languages such as Portuguese⁴, the head F lacks the focus feature necessary to probe for and attract a focus phrase to its specifier. Thus, under minimalistic assumptions, derivations where a focused constituent is moved to the position between the complementizers are barred.

To the best of my knowledge, the impossibility of quantified expressions to occur between the complementizers is the only reason presented in the literature to conclude that focused constituents cannot appear in that position. This argument seems too weak, although what it purports to show is I believe correct.

In fact, the kinds of quantifiers in the sentences Uriagereka experiments with could not be interpreted as contrastive focus, (see among others, Szabolcsi 1981, and Kiss 1998), though they might be interpreted as informationally focused. The distinction between these two categories of focus has semantic motivations (most notably, that contrastive focus entails exhaustivity, whereas informational focus merely presents new, non-presupposed information) as well as syntactic (contrastive focus displays weak cross over effects, which informational focus does not). Benincà and Poletto (2004) show that contrastive focus must precede informational focus, under a split-C hypothesis, and revise Rizzi’s (1997) proposal concerning the position of focus, rejecting the freely recursive character of Topic phrases. Thus, if Benincà and Poletto’s proposal is right, Uriagereka has only shown that the constituent between complementizers cannot be an informational focus. There might still theoretically be a contrastive focus position that non-quantified DPs could occupy, to the left of a doubling ‘que’.

To discard this possibility, we check for exhaustive interpretation in an element between the two complementizers.

---

³ Barbosa (2000) makes no claims about the syntactic structure behind C-doubling, or the locus of the lower complementizer. Her discussion of C-doubling is directed towards using some of its characteristics as a motivation for her proposal that pre-verbal (non-quantificational) subjects in EP are CLLDed.

⁴ The actual examples Uriagereka (1995) discussed are not contemporary Portuguese, but it is clear from the text that his proposal is meant to apply to EP.
Eu acho que UM LIVRO que a Ana comprou.

I think that a book that Ana bought

'I think A BOOK, Ana bought.'

In a dialogue situation, the hearer cannot reply to (15) with (16). In (16), the hearer tries to deny an exhaustivity implication that is not there, hence the statement is infelicitous. This shows that a contrastive focus interpretation for the element between complementizers is not available.

Não, ela também comprou uma revista.

'No, she also bought a magazine.'

We have established that LDed and CLLDed topics, and not focused elements of any kind, can occur between the complementizers. In terms of Benincà and Poletto (2004), this suggests that C-doubling in EP is intimately associated with what they call the topic field. A natural question arises, concerning hanging topics, what they show to be the outermost category in the topic field: can they occur between the complementizers as well? The answer is yes, as can be seen from the following examples:

Acho que, este livro, que a Ana não gostou dele.

'I think that this book, Ana didn't like.'

The sentence in (17) shows what is clearly a hanging topic (HT): between the complementizers is a DP with a resumptive pronoun governed by the appropriate preposition, given the selection properties of ‘gostar’ (to like). Furthermore, unlike the Italian data presented by Benincà and Poletto (2004), European Portuguese does not accept a HT scoping over the embedded sentence to the left of the topmost ‘que’, as evidenced by (18). This piece of data is particularly interesting, because it suggests that the complementizers ‘que’ and ‘che’ show different characteristics even in simplex (non-doubled) contexts. Given the Italian data, Benincà and Poletto (2004) suggest that ‘che’ be located in a position lower than that of HT, between HT and LD, but they do not elaborate on the proposal. At the very least, the EP data show that such a configuration is by no means universal for Romance declarative complementizers.

To sum up what has been established so far, we have seen that complementizer doubling, modulo certain restrictions on the selection by matrix verbs (the ‘lamentar’ cases I alluded to above), is possible in embedded declarative sentences, under the condition that the material between the complementizers is part of the topic field, in the terms of Benincà and Poletto (2004).
2.2 The puzzles of C-tripling and interrogative C-doubling

Thus far, I have been working under the tacit assumptions that only one topic can occur between the complementizers and that a third complementizer position is not available. The former will prove true, the latter, surprisingly, will not. Consider the following sentences:

(19) \(\text{?* Acho que amanhã(,), a Ana(,), que vai conseguir acabar o trabalho.} \)
    I think that tomorrow Ana that will manage to finish the assignment

(20) Acho que amanhã(?), que a Ana(?), que vai conseguir acabar o trabalho.
    I think that tomorrow that Ana that will manage to finish the assignment

In (19) we have two topical elements between the complementizers, clearly two distinct maximal projections, and, under the assumptions we are working with (namely, that multiple specifiers are not an option), two corresponding T\(^0\) nodes. The sentence is quite bad, according to all my informants, regardless of whether one observes the pauses indicated or not.

Sentence (19) contrasts sharply with (20), where we have what seems to be essentially the same sentence, but with two instances of doubling, one after each topic. Notice that no pauses are required after each topic constituent. In fact, EP speakers prefer little to no pause in the spots indicated with commas, and some informants actually report ungrammaticality judgments with prominent pauses. This indicates that we are dealing with a phenomenon that is syntactic in nature, not a result of performance-related issues (hesitance pauses, fillers, and so on)\(^5\).

\(^5\) João Peres (p.c.) pointed out to me an interesting set of data, for which I am very grateful. Witness the contrasts in (i)-(iv):

(i) Acho que o Pedro e a Ana foram a Roma.
    I think that Pedro and Ana went to Rome
    ‘I think Pedro and Ana went to Rome.’

(ii) Acho que o Pedro e a Ana que foram a Roma.
    I think that Pedro and Ana that went to Rome

(iii) Acho que o Pedro e que a Ana que foram a Roma.
    I think that Pedro and that Ana that went to Rome

(iv) \(*\)Acho que o Pedro e que a Ana foram a Roma.
    I think that Pedro and that Ana went to Rome.

Interestingly, heading the second conjunct in the subject DP with the complementizer forces C-doubling before the IP. Without committing to a any particular analysis, I note that sentences like (iii) seem to require a distributive interpretation of the predicate over the subject, as the following contrast with a collective predicate shows (my own judgments, not checked with other informants):

(v) Acho que o Pedro e a Ana que reuniram ontem.
    I think that Pedro and Ana that met yesterday
    ‘I think Pedro and Ana met yesterday.’

(vi) \(*\)Acho que o Pedro e que a Ana que reuniram ontem.
    I think that Pedro and that Ana that met yesterday
There does not seem to be a syntactic reason to disallow four complementizers, but the judgments become rather degraded (cf. (21)). This is unsurprising, for presumably the same performance restrictions that make (22), a sentence with the same topic constituents as (21) but no C-doubling, quite odd will be at play in (21) as well.

(21) ?? Duvido que ontem que o Pedro que à Ana que lhe tenha telefonado.
    ‘I doubt that yesterday that Pedro that to Ana that her have called’
    ‘I doubt that yesterday Pedro called Ana.’

(22) ? Duvido que ontem o Pedro à Ana lhe tenha telefonado.
    ‘I doubt that yesterday Pedro to Ana her have called’
    ‘I doubt that yesterday Pedro called Ana.’

I now turn to the other oddity in the EP complementizer-doubling phenomenon, interrogative C doubling. In EP, it is possible to double the complementizer of embedded interrogatives, ‘se’ (if / whether), as the following sentences show:

(23) Não sei se o João vai chegar a horas.
    ‘I don’t know if João will arrive on time.’

(24) Não sei se o João se vai chegar a horas.
    ‘I don’t know if João if will arrive at hours’
    ‘I don’t know if João will arrive on time.’

These sentences sound perfectly natural to all my informants. In fact, even speakers who cringed at the sound of a declarative C doubling were more than happy to accept ‘se’-doubling as in (24). Somehow, these constructions are available to a larger number of standard EP speakers than ‘que’-doubling constructions. These facts are remarkable: to the best of my knowledge, interrogative complementizer doubling has not been described in other Romance languages.

As expected, the same restrictions as with ‘que’-doubling apply to the elements between complementizers, namely, only topics can occupy in that position:

(25) * Pergunto-me se à MARIA se o João deu um presente.
    ‘I wonder if to Maria if João gave a present’
    (intended: ‘I wonder if John gave MARY a present.’)
    (no contrastively focused phrases between ‘se’-complementizers)

Constructions such as (vi) are possibly elliptical, for they seem essentially equivalent in meaning to (vii), and are similarly deviant.

(vii) *Acho que o Pedro reuniu ontem e que a Ana reuniu ontem.
    ‘I think that Pedro met yesterday and that Ana met yesterday’
Complementizer doubling in European Portuguese

(26) Não sei se o Filipe se o Pedro gosta dele.
not know if Filipe if Pedro likes of him
‘I don’t know about Filipe whether Pedro likes him.’
(hanging topics are fine below the topmost ‘se’)

Finally, the issue of recursive ‘se’ doubling (‘se’-tripling) is a bit less clear than that of ‘que’. Somehow grammaticality judgments on sentences like (27) are not as strong as the ones concerning ‘que’. No informant thought they were totally out, but most considered them a bit unnatural. In any event, (28) is, as one would expect if the judgments from ‘que’ sentences were to carry over here, uniformly considered quite bad:

(27) ? Não sei se amanhã se o Pedro se consegue entregar o trabalho.
not know if tomorrow if Pedro if manages to hand in the assignment
‘I don’t know if tomorrow Pedro will manage to hand in the assignment.’

(28) ?* Não sei se amanhã o Pedro se consegue entregar o trabalho.
not know if tomorrow Pedro if manages to hand in the assignment

3. Concluding remarks

So far, I have concentrated on giving an accurate description of European Portuguese complementizer doubling phenomena, only sporadically mentioning the theoretical implications the data might have. I recapitulate the main points of what came before:

1. Embedded sentences with overt complementizers (declarative ‘que’ and interrogative ‘se’) have the possibility of displaying more than one instance of their C-head.
2. These (phonologically identical) copies must each follow exactly one topical maximal projection, subject to restrictions on the topic field in the C-domain.
3. Recursion of this process is possibly not bound by syntactic reasons, but by performance related issues.

Or, in a somewhat weaker formulation: The doubling process can apply at least twice.

In connection with the Italian data mentioned briefly in the previous section, it is clear that an analysis along the lines proposed by Paoli for Turinese and Ligurian cannot be applied to EP. Because the lower ‘che’ (che₂) is only licensed in subjunctive contexts, and given that subjunctive mood is by and large morphologically unmarked in those languages, Paoli argues that che₂ contains mood and finiteness features, and that its locus is Fin⁰.

The lower Cs in EP doubling constructions do not seem to instantiate any sort of mood category such as subjunctive (or indicative for that matter), since C-doubling in EP appears to be insensitive to the mood of the embedded verb. Moreover, the position Paoli argues for ‘che’ in these Northern Italian dialects, namely FinP, is much too low to account for the EP data: if a doubling ‘que’ is in FinP, then what could be the reason why focused elements cannot precede
doubling ‘que’s? Finally, Paoli’s proposal cannot account for point 3. above, recursion of the process.\footnote{Roughly the same objections apply to Demonte and Soriano (2007) on Spanish ‘que’ doubling, which does not allow recursion, as it too seems to have a lower locus than that of Portuguese ‘que’.
}

In general, it seems clear that an analysis under the cartographic hypothesis that argues that the identical complementizers delimit the C-domain is bound to fail for EP. If one were to commit to that analysis, one would be forced to either assume that three ‘que’s in a Portuguese triple-C construction occupy three categorially different functional positions, or that the whole domain is recursive. Both options are very undesirable.

On the former, one would have to explain why the presence of a ‘que’ in Fin\textsuperscript{0} allows only one topic in the higher Top phrases, as well as why focused phrases are not allowed to appear between complementizers. Moreover, the triple complementizer case would have to be seen as the instantiation of three complementizers in three different functional positions. This multiplication of diverse functional heads, when there seems to be no reason to assume that the complementizers have different functions, is contrived at least.

The other logical hypothesis, that the whole C-domain is recursive, is far from minimalistic. To stick to the idea that ‘que’s only occur in For and Fin, one would have to say that the left periphery can be doubled, so as to house the three complementizer case. This is extreme, and I know of no independent motivation for it.

It seems that, to account for the EP data, we have to drop the assumption that the complementizers sit in For and Fin, an assumption which, incidentally, can hardly even be formulated coherently for the case of ‘se’-doubling.

This article does not have a complete alternative analysis to present. The data surveyed here are novel and puzzling, apparently suggesting that complementizers such as ‘que’ and ‘se’, in Portuguese at least, can occupy recursive Top\textsuperscript{0} positions, yielding representations such as for sentence (20), repeated here as (30).


