

Questions with definite markers in (Old) Romance, with focus on Old Spanish

Olga Kellert

Georg-August-Universität Göttingen
olga.kellert@phil.uni-goettingen.de



Received: 27-09-2017
Accepted: 16-03-2018

Abstract

We will depart from the observation that Romance languages can be subdivided into two groups with respect to free relatives under question-embedding predicates (Kellert 2017). One group has grammaticalized the definite element (e.g. Pt.¹ *o*, Fl. *i* ‘the’) and *que/che* into one question pronoun (e.g. Pt. *o que* ‘what’ and Fl. *icché* ‘what’); the other group has not (e.g. Spanish and French). We will show that in one group free relatives that are embedded under question predicates resemble more complex nouns (as in Spanish and French), whereas in the other group they are clausal and have the structure of ordinary questions as in Portuguese and Florentine. We will look at the evolution of *lo que* sentences in Old Spanish and demonstrate that they were used as relative clauses under non-question predicates such as *ser* ‘be’ and factive predicates such as ‘know’ with much higher frequency than under genuine question predicates such as *preguntar* ‘to ask’. We will suggest that the interrogative interpretation of *lo que*-relative clauses has its source in the ambiguity of factive predicates. Factive predicates can select both DPs interpreted as definite descriptions and CPs interpreted as interrogatives. *Lo que*-relatives can thus be interpreted as definite descriptions and as interrogatives under factive predicates. As we will argue, this ambiguous interpretation was the precondition for the use of *lo que*-sentences to be used in non-ambiguous question contexts. However, the reanalysis of *lo que*-sentences as questions has not been fully accomplished in Modern Spanish in contrast to Modern Portuguese, as these sentences still show syntactic and semantic differences from ordinary questions.

Keywords: (Old) Romance; free relatives; constituent questions; definite descriptions; grammaticalization; reanalysis; factive predicates; question-embedding predicates.

¹ Fl. stands for Florentine, Sp. for Spanish, Pt. for Portuguese, and Fr. for French. Mo. for Modern and O. for Old and Mi. for Middle languages.

Table of Contents

<p style="text-align: center;">1. Introduction</p> <p>2. Q-SFRs in Modern Romance languages</p> <p style="text-align: center;">3. Diachrony of Q-SFRs</p>	<p style="text-align: center;">4. Analysis of Q-SFRs</p> <p style="text-align: center;">5. Summary and Outlook</p> <p style="text-align: center;">References</p> <p style="text-align: center;">Appendix</p>
---	--

1. Introduction

In some Romance languages, Modern and Old, questions can have the form of relative clauses of a special kind:

- (1) a. Porque entiendan e sepan *lo que* judgaren (Mi.Sp.)
 because understand and know the that judged
 ‘because they understand and know about what they judged.’
 (García Cornejo 2006)
- b. Querer *o que* ella quer (O.Pt.)
 ask the that she asks
 ‘To ask what she asks.’
 (Mattos e Silva 1990)
- c. Je ne sçay *ce que* j'ay voulu dire (Mi.Fr.)
 I not know the that I have wanted say
 ‘I don’t know what I wanted to say.’
 (FRANTEXT)
- d. 'Un lo so *icché c'* era (Fl.)
 not it know the that there was
 ‘I don’t know what it was.’
 (COR ifamev10)
- e. La dimandò, *quello ch'* ella avesse (O.Fl.)
 her asked this which she had
 ‘he asked her what she had.’
 (Benincà 2010:66)

The embedded questions in (1) are introduced by some definite element; e.g. a pronoun or an article, followed by the element *che/que*, which has different functions in Romance. Following Rebuschi (2001), we will call them semi-free relatives (SFRs²). When they are used as questions, for convenience we will label them question semi-free relatives: Q-SFR. This type of question usually contains a gap in the direct object position of the embedded clause, i.e. questions that contain a gap in the adjunct position do not contain any definite element:³

- (2) Sin saber dónde podría estar. (Mi.Sp.)
without knowing where could be
'without knowing where she/he might be.'
(Corpus del Español, CDE)

This possibility is at odds with the nominal category that free relatives are supposed to have and this makes them a very interesting potential mismatch between form and function.⁴ In the paper we will explore how modern Romance languages behave with respect to Q-SFR (section 2), verify whether Old Spanish can tell us something about their evolution (section 3), and propose an analysis of Q-SFRs (section 4).

2. Q-SFRs in Modern Romance languages

The following data demonstrate that Modern Romance languages can be divided into two groups with respect to Q-SFRs (see also Kellert 2017). The first group (Group 1) uses them only in indirect questions, whereas the second group (Group 2) also uses them in direct questions. We will illustrate the first group with European French and European Spanish⁵, and the second group with European Portuguese and Florentine.

- (3) *Group 1 (= indirect questions only)*
a. *(Je sais pas) ce que tu fais. (Fr.)

² The term "SFRs" goes back to Rebuschi (2001), who considers free relatives with an overt definite pronoun to be semi-free. SFRs do not completely satisfy the definition of free relatives (FRs) proposed by Caponigro (2004), because they lack overt wh-morphology.

³ We do find some minor examples with other embedded interrogatives with a gap in an adjunct position: Fr. Je sais *ce* pourquoi il a fait cela (Warrant 1982:305). We will also ignore questions with definite markers in complex noun phrases in this study (cf. also Benincà 2010:57 for O.Fl. examples):

- (i) *Li quex* de voz est Lanceloz? (O.Fr.)
the which of voices is Lanceloz
'Which one of the voices belongs to Lanceloz?'

(Kunstmann 1990:66)

⁴ A compositional problem arises if QSFRs denote definite descriptions due to the definite marker they contain. However, questions are usually not analyzed as definite descriptions (cf. also Ambar & Veloso 2001: Fn.18 for hinting to a similar problem).

⁵ As we will see below, some varieties of Spanish behave more like European Portuguese with respect to Q-SFRs, i.e. they belong to Group 2 and not Group 1 as European Spanish does.

*(I don't know) what you do.'

- b. *(No sé) lo que haces. (Sp.)
*(I know) what you do.

(4) *Group 2 (also direct questions)*

- a. O que é que ela faz? (Pt.)
'What does she do?'
- b. Icchè fai? (Fl.)
'What do you do?'

Our corpus analysis confirms the speaker judgements in (3) and (4). Most occurrences of non-embedded *lo que*-questions in CDE are either elliptical questions or yes/no questions in which *lo que*-sentence has the function of a free relative clause:

- (5) ¿Lo que es malo para Gibraltar es bueno para Ferrol?
the that is bad for Gibraltar is good for Ferrol
'What is good for Gibraltar is bad for Ferrol?'

As to the main clauses of the 2nd group, we found main interrogatives in Portuguese and Florentine in COR (Corpus of spoken Romance languages) (see Kellert & Lauschus 2016 for main interrogatives with *icché* in Florentine).

This distinction correlates with other differences. Q-SFRs in Group 2 can be infinitival in contrast to Group 1:

(6) *Infinitival questions (Group 1)*

- a. *Je sais pas ce que faire. (Fr.)
I know not the that do.

- a' Je sais pas quoi faire. (Fr.)
'I don't know what to do.'

- b. *No sé lo que hacer. (Sp.)
not I.know the that do.'

- b' No sé qué hacer. (Sp.)
'I don't know what to do.'

(7) *Infinitival questions (Group 2)*

- a. Un so icché fare. (Fl.)
'I don't know what to do.'

- b. O que fazer em Amsterdam? (Pt.)
'What is there to do in Amsterdam?'

Group 1 in contrast to Group 2 does not allow multiple *wh*-elements nor *wh*-in situ where the *wh*-element remains in the postverbal position:

- (8) *Multiple wh-elements (Group 1)*
- a. *Je sais pas ce qu'est quoi. (Fr.)
I know not the that is what
- b. *No sé lo que es qué (Sp.)
not know the that is what
- (9) *Multiple wh-elements (Group 2)*
- a. Quem fez o quê a quem? (Pt.)
'Who did what to whom?'
- b. Icché è icché? (Fl.)
'What is what?'
- (10) *wh-in situ (Group 1)*
- a. *faire ce que? vs. faire quoi? (Fr.)
do this that 'do what?'
- b. *hacer lo qué?⁶ vs. hacer qué? (Sp.)
do the that 'do what?'

Group 1 in contrast to Group 2 does not allow ellipsis either:

- (11) *Elliptical (Group 1)*
- a. *Ce que? vs. Quoi? 'what?' (Fr.)
- b. *Lo que? vs. Qué? 'what?' (Sp.)
- (12) *Elliptical (Group 2)*
- a. Icché? 'What?' (Fl.)
- b. O quê? 'What?' (Pt.)

⁶ There is variation with respect to the status of *lo que* in spoken Argentinian (AR) and Uruguay (UY) as these varieties of Spanish allow ellipsis and *wh-in situ*:

- i. ¿Lo qué? (AR)
the what
'What?'
- ii. ¿Una solución a lo qué? (AR)
a solution to the what
'A solution to what?'
- iii. ¿Sabes lo qué? (UY)
know the what
'You know what?'
- iv. ¿Lo qué? le pregunté. (UY)
the what him asked
'What? I asked him.'

There is also variation with respect to SFRs in Aostan French (*français valdôtain*) where the pronoun *ce* and the complementizer are used as one compound *sen-che* 'what' in embedded and matrix questions (see Diemoz 2007, Poletto & Pollock 2009):

- i. Sen-che fi-yen?
ce-that do we
'What shall we do?'

(COR pfamd131)

Nor does Group 1 allow a second complementizer of the *wh*-element in contrast to Group 2:

- (13) *Second complementizer (Group 1)*
 Je ne sais pas ce que (*que) tu fais. (Fr.)
 ‘I don’t know what you do.’
- (14) *Second complementizer (Group 2)*
 O que que é isso? ‘What is it?’ (Pt.)

Q-SFRs in Group 2 can be embedded under prepositions. This property is not present in French that belongs to Group 1.

- (15) *Preposition + SFRs (Group 2)*⁷
- a. co’icché le fanno? (Fl.)
 ‘With what do they make them?’
- b. sabes do que é que aquelas pessoas vivem (Pt.)
 know of that is that these people live
 ‘Do you know how the people live’
- (COR pfammn18)

- (16) *Preposition + SFRs (Group 1)*
- a. Je ne sais pas de quoi tu parles. (Fr.)
 ‘I don’t know what you are talking about.’
- b. *Je ne sais pas de ce que tu parles. (Fr.)
 ‘I don’t know what you are talking about.’

Spanish, however, allows prepositions in embedded SFRs (see also Contreras 1999 and section 3.2 for corpus data):

- (17) a. No sé de qué hablas. (Sp.)

⁷ Fronting of prepositions represents an important argument for a movement analysis of the XP the preposition embeds (see e.g. in i.). The PP *de Jean* ‘about Jean’ must have moved from the embedded clause to the surface position (see Kellert 2015 and the references there):

i. C’est [de Jean]_i que je parle t_j.
 ‘It’s about Jean that I talk’

The fact that prepositions are not allowed in SFRs in French in (16b) might suggest that the preposition *de* in (16b) cannot move with the demonstrative pronoun *ce* to the initial position of the embedded clause and the sentence is thus ungrammatical. As *lo que*-sentences are grammatical with preposition fronting in (17b), we must assume that *P + lo* must have moved from its argument position to its surface position.

not know of what talk
'I don't know what you are talking about.'

b. No sé de lo que hablas. (Sp.)
not know of the that talk
'I don't know what you are talking about.'

Despite the differences mentioned above, both groups allow SFRs to be used in non-question contexts:

(18) *Non-question contexts (Group 1 + 2)*

a. Hanno sfasciato tutto icché c'era (Fl.)
have destroyed all the-that there was
'They destroyed everything there was.'

(COR ifamev10)

b. c' est tout ce que j'ai voulu dire. (Fr.)
the is all thid that I have wanted say
'This is all I wanted to say.'

Moreover, in both groups SFRs can be coordinated with embedded questions with an interrogative pronoun:

(19) *Coordination*

a. Como no sabía de qué era ni lo
because not knew of what was nor the
que podía salir por allí (Sp.)
that could sort.out through there
'Because I don't know of what it consisted nor what could sort out of it.'

(CDE)

b. Je ne sais pas qui j'aime ni ce que j'aime. (Fr.)
I neg.cl. know not who I love nor the that I love
'I don't know who I like nor what I like.'

Both groups allow the definite element + *que/che* to refer to an argument inside the embedded sentence and thus to undergo a non local relation with the embedded verb (e.g. with Fl. *pesare* 'to weigh', Sp. *enseñar* 'to teach' and Fr. *faire* 'to do' or):

(20) *Non local relation with the embedded verb*

a. Icché vuo' che pesi? (Fl.)
what want-you that weigh-you
'It shouldn't weight so much.'

b. ya sé lo que quiero que se enseñe (Sp.)
ya know the that want that imp.pro teach
'I know what I want to be taught.'

(CDE)

- c. Je sais ce que tu veux qu'elle fasse. (Fr.)
 I know the that you want that she does
 'I know what you want her to do.'

We can conclude from the data presented above that the definite marker and the complementizer *que/che* have been lexicalized as interrogative pronouns or as *wh*-elements in Group 2, thus behaving as ordinary *wh*-elements that undergo movement and can be used *in situ*, in infinitival questions, under prepositions, etc.:

- (21) *Group 2: O que/ Fl. icché/ ... 'what' [+wh]*⁸ (Pt.)

Similar Q-SFRs do not contain a *wh*-feature in Group 1 and therefore cannot be positioned *in situ*, co-occur with other *wh*-items, etc.:

- (22) *Group 1: ce que / Sp. lo que 'the that' [- wh]* (Fr.)

Given the possibility of coordination in both groups and the principle that only constituents of equal syntactic status can be coordinated, we must conclude that Q-SFRs have the same syntactic status as ordinary questions (usually analyzed as CPs and not DPs) in both groups. This conclusion is crucial because it excludes an analysis according to which SFRs are analyzed as DPs and not CPs (see Section 4 for an analysis). In French and Spanish Q-SFRs lack a *wh*-feature, which is present in Group 2. However, it seems that there is variation in some varieties of (Latin American) Spanish with respect to the *wh*-feature of *lo que* and of *ce que* in Aostan French (see fn. 6). We will study the variation systematically in future research. In this article we are merely concerned with the variety of Spanish and French in which the element *lo* and *ce* do not have any *wh*-feature.

In order to understand the DP-like property of Q-SFRs in Group 1 and why Q-SFRs are missing a *wh*-feature there, we will study the diachrony of Q-SFRs. The diachronic analysis will shed some light on the synchrony of these sentences and in what respect they are different from *wh*-interrogatives. The next section deals with the diachrony of Q-SFRs.

3. Diachrony of Q-SFRs

3.1. Research questions

The use of Q-SFRs in Old Romance languages has been observed in the literature (cf. Mattos e Silva 1990 for Old Portuguese, Kunstmann 1990 for Old French, Ridruejo 1987 and García Cornejo 2006 for Old Spanish, Benincà 2010 for Old Florentine).

However, some questions still remain open:

- a) Whether their use as questions is derivative from their use as relative clauses, and thus if they appear with non-question predicates more and earlier than with question predicates; and

⁸ They cannot function as *wh*-determiners though they can as *wh*-pronouns: e.g. Fl. *che figliola?* 'What girl?' **Ichché figliola?* 'What girl?' or Pt. (**o*) *que mulher?* 'What woman?'

b) Whether Old Romance languages correspond to Group 1 or Group 2 or the two groups coexisted early on.

We decided to restrict our research to Old Spanish, because Old Florentine does not contain *icché* nor *il che* (instead *quello che/ciò che*, see Benincà 2010 and 1e in section 1). Later Renaissance texts need to be scrutinized for the occurrence of Fl. *il che* manually because of a gap in the corpus. Q-SFRs in Old Portuguese and Old French will be analyzed in future research.

In order to answer the questions in a) and b), we will look at the distribution of predicates that embedded *lo que*-SFRs and *qué*-interrogatives in Old Spanish and investigate whether these two sentence types shared the same syntactic and semantic properties diachronically.

3.2. Corpus Methodology and Statistics

Our goal is to investigate whether *lo que*-SFRs and *qué*-interrogatives show different frequencies with respect to the matrix verb and with respect to the morphosyntactic context in which these two sentence types appear (e.g. morphosyntactic features of the embedded verb, preposition embedding, etc.). We tested the following Null Hypotheses:

1. Null Hypothesis 1: *lo que*-SFRs are equally frequent under different types of predicates in the diachronic corpus CDE (e.g. +/- *preguntar*).

2. Null Hypothesis 2: *lo que*-SFRs and *qué*-interrogatives are equally frequent under every predicate.

3. Null Hypothesis 3: *lo que*-SFRs and *qué*-interrogatives are equally frequent with certain tense and mood features on the embedded verb.

Based on this goal, we decided to use Corpus del Español (CDE), which is annotated for part of speech and for lemma. “This corpus contains more than 100 million words in more than 20,000 Spanish texts from the 1200s to the 1900s. The 20 million words of text from the 1900s is balanced between the genres of spoken, fiction, newspaper, and academic. This corpus thus allows you to compare across genres and time periods.” (Davies, website of CDE).⁹

We searched for different strings depending on which hypothesis we wanted to investigate:

⁹ There is another historical corpus CORDE. This corpus is not annotated in any way and thus does not allow us to search for syntactic configurations. It is only useful for searching for exact phrases. Its main advantage though is that CORDE offers the possibility to limit the search to specific countries and registers across periods and varieties. CORDE contains 240 Million words from 950-1974 and does not contain any oral data, only written texts made up of books and press. It does not give relative frequency which must be calculated manually. Due to the lack of annotation and relative frequency in CORDE, we decided to use CDE. Unfortunately, CDE does not contain any information about the text genre until 18th c. We will leave the investigation of specific textual typology with respect to the use of *lo que*-SFRs for future research. It has been pointed out by Octavio de Toledo & Rodríguez Molina 2017 that CDE contains errors in the dates of texts, especially in medieval texts. We have tried to delete examples and data from erroneously dated texts. It should be noted though that the dating problem does not present a big problem for our generalizations about *lo que*- and *qué*-sentences since our focus goes beyond making general statements about some period in particular, but rather on investigating linguistic factors such as predicate class.

- Null Hypothesis 1: *lo que*-SFRs are equally frequent under different types of predicates in the diachronic corpus CDE (e.g. +/- *preguntar*)
 - Searched string (| stands for *or*, ! synonyme for *no* ‘not’):
 - [lemma: saber|!no saber|preguntar|ser|ver|decir + lo que + finite Verb]
- Null Hypothesis 2: *lo que*-SFRs and *qué* constructions are equally frequent under every predicate
 - Searched string: [lemma: saber|!no saber|preguntar|ser|ver| decir + qué + finite Verb]
- Null Hypothesis 3: *lo que*-SFRs and *qué* constructions are equally frequent with embedded verbs with certain tense and mood features.
 - [lemma: saber|preguntar + lo que|qué+ Verb in Past Perfect= preterito perfecto]
 - [lemma: saber|preguntar + lo que|qué + Verb in subjunctive]

We screened the results to see whether they match the string we were looking for. Afterwards, the frequencies of the syntactic strings have been tested statistically by the *chi square*- test. This test compares a set of observed frequencies O with a set of expected frequencies E (Fry 2011):

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

If the difference between observed and expected frequencies is large, we can reject the null hypothesis of independence.

$$H_0 : \chi^2 = 0 \quad H_1 : \chi^2 > 0$$

As our tests are mainly restricted to two observed frequencies, i.e. the frequency of *lo que*-SFRs and the frequency of *qué*-interrogatives, the χ^2 value must always be more than 3.841 in case the p-value= 0.05 (significant).

df	p=0.1	p=0.05	p=0.025	p=0.01	p=0.001
1	2.706	3.841	5.024	6.635	10.828

Concerning the frequencies of the results, we presented both absolute and relative frequencies. The latter information is especially necessary because the subcorpora used to represent every period are not equal in CDE and the relative frequency allows us to compare the data between periods (see Rojo 2011).

3.3. Q-SFRs vs. *qué*-interrogatives

In order to answer the question of whether the frequency of Q-SFRs is the same as the frequency of *qué*-interrogatives, we conducted a quantitative corpus analysis.

Our aim was to investigate for each century from 12th to 19th centuries whether there is a difference between Q-SFRs and *qué*-interrogatives depending on the matrix verb under which these types are embedded. Our results from the corpus study from 12th to 19th centuries in CDE show that *lo que* SFRs were used more often under non-question predicates (e.g. copula *ser* ‘be’) than under genuine question predicates (e.g. *preguntar* ‘ask’).

(13) Lemma: Ser + lo que + V_{fin} = 12428 occ. in total from 12th. to 19th.c.

Lemma: Preguntar + lo que + V_{fin} = 148 occ. in total from 12th to 19th c.

The difference between predicates is statistically significant (X-squared = 11991, df = 1, p-value < 2.2e-16), so we must assume the two predicates embedding *lo que* are not equally frequent in the population for which our corpus is representative (see Fry 2011 for other linguistic examples).

We did not observe any change across time either, i.e. *ser* has always shown higher frequency in contrast to *preguntar* as the following two specific forms *es lo que* + V_{fin} and *es qué* + V_{fin} show:

Table 1: *es lo que* + V_{fin} vs. *es qué* + V_{fin}

		All	Per Mil	12	13	14	15	16	17	18	19
<i>lo que</i>	Per mil		95.86	52	38.96	80.50	102.79	120.75	85.77	84.62	120.76
	All	9586		352	104	657	1751	1491	842	1633	2756
<i>qué</i>	Per mil		0.22				0.29	0.32	0.10	0.36	0.13
	All	22		2			5	4	1	7	3

We also tested whether there is a difference between *lo que* and *qué* embedded under the question predicate *preguntar*. The following table shows a higher frequency of *qué*-interrogatives than *lo que*-SFRs under this predicate:

Table 2: Lemma: *preguntar* ‘ask’ *lo que* + V_{fin} vs. *qué* + V_{fin}

Type	Total	12	13	14	15	16	17	18	19
<i>Lo que</i>	93	0	0	1	34	22	6	25	5
<i>qué</i>	288	0	0	1	93	37	13	33	111

The difference is significant as the test has shown:

- (14) data: c(93, 288)
X-squared = 99.803, df = 1, p-value < 2.2e-16

The following tables (Tables 3-6) representing frequencies between *lo que*-SFRs and *qué*-interrogatives show statistically significant frequencies between these types (see the statistical test below the tables). *Lo que*-SFRs tend to appear more often under factive verbs (‘know’) and verbs of cognition (‘say’, ‘see’) than *qué*-interrogatives (see Tables 3-5). The latter type tends to appear more often under the non-factive verb ‘not know’ (see table 6 and the statistical test below).

Table 3: *saber lo que* vs. *saber qué*

		All	Per Mil	12	13	14	15	16	17	18	19
<i>lo que</i>	Per mil		11.25	11.47	5.62	6.86	17.55	18.22	11.21	14.82	4.25
	All	1165		77	15	56	299	225	110	286	97
<i>qué</i>	Per mil		9.64		0.37	0.61	16.73	13.69	6.21	10.05	10.91
	All	964		0	1	5	285	169	61	194	249

data: c(1165, 964). X-squared = 18.977, df = 1, p-value < 1.323e-05

Table 4: *ver lo que vs. ver qué*

		All	Per Mil	12	13	14	15	16	17	18	19
lo que	Per mil		12.84	1.49	3.00	3.31	31.35	25.02	7.95	10.16	5.35
	All	1284		10	8	27	534	309	78	196	122
qué	Per mil		7.31	0.15		0.49	12.50	8.83	3.06	5.23	13.71
	All	771		1	0	4	213	109	30	101	313

data: c(1284, 771): X-squared = 128.06, df = 1, p-value < 2.2e-16

Table 5: *decir qué vs. decir lo que*

		All	Per Mil	12	13	14	15	16	17	18	19
lo que	Per mil		6.88				13.85	11.90	7.95	7.57	3.55
	All	688		0	0	0	236	147	78	146	81
qué	Per mil		0.91				1.12	0.89	0.20	0.93	1.80
	All	91		0	0	0	19	11	2	18	41

data: c(91, 688). X-squared = 457.52, df = 1, p-value < 2.2e-16

Table 6: *no sé lo que vs. no sé qué*

		All	Per Mil	12	13	14	15	16	17	18	19
lo que	Per mil		5.31				5.75	7.77	5.50	10.68	3.37
	All	531		0	0	0	98	96	54	206	77
qué	Per mil		32.98	0.45		0.12	42.15	57.01	40.95	38.19	32.12
	All	3298		3		1	718	704	402	737	733

data: c(531, 3298). X-squared = 1999.6, df = 1, p-value < 2.2e-16

Note also that *lo que*-sentences did not become more frequent over time under the question predicate *preguntar* ‘ask’ nor under the non-factive predicate *no saber* ‘not know’, as there is no continuous rise in frequency in time (see table 2 and 6).

3.4. Evaluation of the results in section 3.3

Our data in 3.3 has shown that SFRs did not occur in prototypical interrogative contexts as frequently as they occurred in non-interrogative contexts such as the complement of the copula, verbs of saying, verbs of perception and factive verbs.

Our data do not confirm the hypothesis present in the literature according to which the use of *lo que*-SFRs expanded from 16th century to genuine interrogative contexts (Girón 1988:177, Stulic-Etchevers 2007: 279). Stulic-Etchevers (2007) did a diachronic analysis of *lo que* relatives based on a restricted number of occurrences in the corpus CORDE and CDE and came to the conclusion that *lo que* relatives were first used with verbs of saying, some psych-verbs like ‘see’, ‘hear’, and factive predicates like ‘know’, but not with real question predicates like ‘ask’ or negated factive predicates like ‘not know’ before the 15th century and that only after 16th century their use expanded to interrogative contexts (cf. also Girón 1988:177). According to her data analysis, embedded questions usually appeared with wh-

pronouns such as Sp. *qué* before the 16th century. After the 16th century, *lo que*-relatives started to compete with *que*-interrogatives. One plausible explanation that she advocates for why *qué* and *lo que* were used under different predicates ('know' vs. 'ask') before the 16th century is that *qué* is related to the speaker's uncertainty about *qué*'s reference, whereas *lo que* refers to speaker's certainty (cf. also Girón 1988:76, García Cornejo 2006:87). According to these authors *lo que* has lost this semantic property over time.

We observed in section 3.3 that until 15th c. both sentence types show very low frequencies in interrogative contexts and that the frequencies of the two sentence types depend on the verb predicate (Tables 1-6).

From our data, we cannot conclude that there was a significant change in the use of *lo que* throughout the periods, i.e. *lo que*-SFRs did not replace the use of *qué*-interrogatives under question embedding predicates.

Our data disconfirm the mentioned hypothesis in the literature because the differences between the predicate classes remain stable across periods, i.e. genuine question predicates always have shown lower preference for *lo que*-SFRs than for embedded questions with *qué*. We did not observe that *lo que* SFRs have changed significantly with respect to the predicate class nor that they expanded their use under question embedding predicates and replaced *qué*-interrogatives in any way.

Stulic-Etchevers 2007 shows on the basis of her data that the verb 'say' has an equal frequency in both sentence types. This hypothesis is not true according to our results as *lo que*-sentences are more frequent than *qué*-sentences with this verb. According to Stulic-Etchevers 2007, the verb 'know' is more frequent with *qué* than *lo que*. This is not confirmed with our data either. *Negation* + 'know' is more frequent with *qué*, whereas the factive verb *know* without negation is more frequent with *lo que*. According to Stulic-Etchevers 2007 question predicates show lower frequency with *lo que* than *qué*. This is confirmed with our data.

The question arises as to why *lo que* is more frequent with verbs of perception *ver* 'see', verbs of saying *decir* 'say' and factive verbs *saber* 'know' but not with question embedding predicates and non-factive verbs. One possible explanation is that the former class of verbs is known for their ambiguous status between question-embedding and DP-embedding predicates (see Kiparsky & Kiparsky 1970, Egré 2008). It is thus not surprising that these verbs can take *lo que*-SFRs and *qué*-interrogatives as their complements. But the data also shows that these ambiguous predicates more often take *lo que*-SFRs than *qué*-interrogatives. One possible explanation is that *lo que*-SFRs are semantically ambiguous between an interrogative and a definite description interpretation.

Another observation is that the negated factive verb *no saber* 'not know' does not take *lo que*-SFRs as often as the factive verb *saber* 'know'. According to Stulic-Etchevers 2007, negated factive verbs are similar to question-embedding predicates in that they embed interrogatives (see Stulic-Etchevers 2007). This is why these predicates more often take *qué*-interrogatives than *lo que*-SFRs.

The next section shows further syntactic differences between *lo que*-SFRs and *qué*-interrogatives and suggests that their syntactic status must be different due to differences in distribution.

3.5. Further differences between *lo que* and *qué*

We checked whether embedded sentences with *lo que/qué* had the same syntactic distribution. The following table shows that only the interrogative pronoun *qué* occurs with an infinitive verb with only 2 exceptions in the *lo que*-sentence:

Table 7: occ. embedded *lo que*-SFRs and *qué*-interrogatives with infinite verbs

1	No saber lo que Verb _{infinitive}	0	0	0	2	0	0	0	0
2	No saber qué Verb _{infinitive}	0	0	0	105	24	23	138	139

This observation reinforces the syntactic difference between *lo que*-SFRs and *qué* interrogatives and shows that *lo que* never became similar to *qué*-interrogatives across time.

Another difference is that only *qué* can modify a noun (e.g. *qué NP* vs. **lo que NP*), because *lo* is a pronoun and not a determiner, whereas *qué* can act as a determiner and as a pronoun:

- (15) Saber qué productos tiene
know what products has
'to know what products he has.'
- (16) Saber lo que (*productos) tiene
know the that products has
'to know what he has.'

These observations suggest that *lo que* does not have and never has had the same syntactic status as the interrogative pronoun *qué*.

Another property that confirms our conclusion is the fact that *lo que/qué* behave differently with respect to prepositions (P) that are selected by embedded verbs (see also Stulic-Etchevers 2007 and Fn 7 for the importance of prepositions in syntactic analysis of SFRs).

We observed that *P + lo que* are significantly less frequent than *P + qué* under the question embedding predicate *preguntar*. This observation is expected given that *lo que*-SFRs are less frequent under *preguntar* in general than *qué*-interrogatives (see 3.3):

Table 8: Freq *Preguntar por/de lo que* vs. *preguntar por/de qué*

	All	12	13	14	15	16	17	18	19
Prep <i>lo que</i>	2	0	0	0	1	0	0	0	1
Prep <i>qué</i>	59	0	0	0	19	9	2	8	21

What is not expected though is that *lo que*-SFRs are less frequent under prepositions than *qué*-interrogatives with factive verbs although *lo que*-SFRs are more frequent under this predicate in general (see 3.3, table 4):

- (17) saber + P + *lo que* (19 occ.) vs. saber + P + *qué* (1326 occ.)
data: c(19, 1326). X-squared = 1270.1, df = 1, p-value < 2.2e-16

The findings with preposition distribution suggest that *lo que*-SFRs were generally less frequent with prepositions than *qué*-interrogatives. We checked

whether this difference still holds for Modern Spanish of 20th c. It does, as the following difference shows:

- (18) Distribution in Modern Spanish, 20th c.
[Saber + P + *lo que* + V_{fin}] **1538** vs. [Saber+P+ *qué* +V_{fin}] **9726**
X-squared = 5952, df = 1, p-value < 2.2e-16

The diachronic and synchronic difference in frequency suggests that *P + lo (que)* and *P + qué* cannot be analyzed uniformly in Spanish. One possibility to capture this difference is to assume that *P + lo (que)* does not build a constituent in Old and Modern Spanish (cf. Medeiros Junior 2014 for this conclusion on O. Pt.). However, this assumption is too strong given that *P + lo que* is not ungrammatical in Old and Modern Spanish. It's just not very frequent. Another way to capture this difference is to assume that *lo que*-SFRs have a different feature make up than *qué*-interrogatives which accounts for the difference in frequency (see section 4 on such analysis).

In the next section we will look at possible differences in the interpretation of *lo que* and *qué* interrogatives, especially trying to see whether we find evidence for the hypothesis according to which *lo que* has lost its referential features from 15th /16th centuries and has been reanalyzed as an indefinite pronoun from this time on, which was a necessary condition for its occurrence under question embedding predicates (see Girón 1988:76, García Cornejo 2006:87, Stulic-Etchevers 2007).

3.6. Interpretation of SFRs in the diachrony

If *lo que* had lost its semantic value from the 15th century on and had gained the semantic status of an interrogative pronoun *qué*, we would not see any difference in frequency between *lo que*-SFRs and embedded questions contrary to fact (see 3.3). However, it is also true that *lo que*-SFRs appeared under question predicates together with *qué*-complements (see 3.3). This observation suggests that we should investigate the semantics of *lo que*-SFRs under question predicates and see whether their semantic status is the same as the semantics of *qué*-sentences. Concerning the interpretation, it is still not clear whether there is a semantic difference between the two embedded sentences in Modern Spanish. One possible hypothesis is that *lo que* is always associated with the certainty that there is some value for the variable which is associated with the argument of the embedded verb, whereas *qué* is not (see Keniston 1937: 150-151). According to this hypothesis, *lo que*-SFRs have an existential presupposition, i.e. they presuppose the existence of something to which the definite element refers:

- (19) Le preguntaron qué sabe de eso
her asked what knows of it
'They asked her what she knows about it'
(it might be the case that she/he does not know anything)
- (20) Le preguntaron lo que sabe de eso
her asked the that knows of it
'They asked her what she knows about it'
(# it might be the case that she/he does not know anything)

According to Alarcos Llorach 1984 and Contreras 1999 there is no semantic difference in interpretation in Modern Spanish, i.e. *lo que*-SFRs are interpreted as embedded interrogatives similar to *qué*-interrogatives. We analyzed our data in order to see whether *lo que*-interrogatives have existential presupposition. The following data confirm that *lo que*-SFRs have existential presupposition, because all examples under genuine question predicates refer to some factual event or state expressed by present or past tense morphology of the embedded verb and the agent or the speaker asks a question about this event:

- (21) 14th
 E aquí non preguntamos lo que sabe cada vno por su
 And here not ask what knows everyone for his
 persona. Mas lo que le pertenesce saber por Razon desta arte.
 person. But what him belongs know for reason of this art.
 ‘And here, we do not ask what everyone knows themselves, but what
 belongs to each one for the reason of this art’.
 [De inventione; Retórica. Cicerón; Alfonso de Cartagena tr.]

- (22) 15th
 a. y luego salieron los dos de casa, y así, le
 and then left the two from house, and so, him
 preguntó lo que deseaba,
 asked what desired,
 ‘And then, the two of them left the house, thus, he/she asked him/her what
 he/she desired’.
 (*Don Quijote de la Mancha*, Cervantes Saavedra, Miguel de)

- b. con voz turbada preguntó lo que había sido.
 with voice confused asked what had been.
 ‘With a confused voice, he/she asked what had been’.
 (*Guzmán de Alfarache*, Alemán, Mateo)

- (23) 16th
 Y viendo que naipes daban, preguntó lo que jugaban,
 And seeing that cards gave, asked what played,
 y el horoba dijo: «Besos.»
 and the humped said: ‘Kisses.’.
 ‘Seeing that they gave each other cards, he/she asked what they were
 playing. And the humped said: ‘Kisses.’.
 (*Cómo ha de ser el Privado*, Quevedo, Francisco de)

However, note that *qué*-interrogatives can also have existential presupposition as shown in the following examples that also refer to some events in the actual world:

- (24) 14th
 llegóse a él y preguntóle qué hacía allí y qué miraba
 came to him and asked him what made there and what looked
 en aquel pozo.
 in that well.
 ‘He came to him and asked him what he did there and what he looked for
 in that well’.
 (*Vida de Sanct Isidoro*, Martínez de Toledo, Alfonso)

- (25) 16th
 Y llegándose a él le preguntó qué hacía. Y él
 And coming (se) to him him ask what make. And he
 respondió que era soldado y pobre.
 responded that was soldier and poor.
 ‘And coming to him he asked him what he was doing. And he answered
 that he was a soldier and poor’.
 (*Sucesos y prodigios de amor*, Pérez de Montalbán, Juan)

Our analysis has shown that *lo que*-interrogatives have an existential presupposition. Keniston’s 1937 hypothesis is thus confirmed, but Contreras 1999 hypothesis is not disconfirmed either because *qué*-interrogatives also can have this interpretation and thus both types may have a similar interpretation, as suggested by Contreras 1999.

The following discussion of the data will show that both Keniston 1937 and Contreras 1999 are right to some extent if we take tense and modality into account. It will be shown that it is tense and modality that disambiguates the +/-existential presupposition of a *lo que*-sentence.

Given the hypothesis that *lo que*-sentences refer to events in the actual world, we expected to find more *lo que*-sentences in episodic contexts than *qué*-sentences under Keniston’s hypothesis. Episodic contexts are usually realized by tense and mood that refer to some actual or past event (e.g. present tense, past tense). Whereas modal contexts expressed by modal verbs, subjunctive, etc. usually refer to speaker’s attitude or epistemic state.

We observed expected differences in the distribution of past tense such as Sp. *Pretérito perfecto*, that refers to a completed event in the past (see González 2003), between *lo que*-SFRs and the *qué*-interrogatives, i.e. *lo que*-SFRs occurred more often than *qué*-interrogatives with respect to this tense form that usually trigger episodic interpretations of the embedded clause:

Table 9: Freq. *saber* + *qué* +pret.perf. vs. *saber*+ *lo que* +pret.perf.

	All	12	13	14	15	16	17	18	19
Lo que	152	7	1	13	19	17	15	46	34
qué	96	0	0	1	17	13	2	14	49

data: c(96, 152) X-squared = 12.645, df = 1, p-value = 0.0003765

The higher frequency of episodic interpretations of *lo que*-sentences confirms Keniston’s hypothesis according to which *lo que*-sentences tend to refer more often to factual events than *qué*-interrogatives.

Whereas, embedded *qué*-interrogatives occurred much more often with subjunctive than *lo que*-SFRs:

Table 10: Saber lo que + subj. /saber qué + subj.

	All	12	13	14	15	16	17	18	19
Lo que	24	0	1	0	3	12	3	5	0
qué	168	0	0	3	75	58	6	14	12

data: $c(24, 168)$, X-squared = 108, df = 1, p-value < 2.2e-16

Subjunctive is associated with non-veridicality (i.e. the proposition that contains the subjunctive does not presuppose the truth, see Quer 1998, among others). Embedded questions with subjunctive do not have existential presupposition (see also Homer 2008, among others). That is why *qué*-sentences occur generally more often with subjunctive than *lo que*-sentences which usually have an existential presupposition due to their definite element. However, there are also *lo-que*-SFRs that co-occur with subjunctive even though the frequency is lower than the frequency of *qué*-interrogatives:

- (26) No hay quien mejor sepa lo que sea menester para
 No have anyone better know what is necessary for
 aquel día (15th c.)
 that day
 ‘There is no one better that knows what is necessary for that day’.
 (*Libro de la oración y meditación*, Luis de Granada, (O.P.))
- (27) Las botas. ¿Quién sabe lo que puedan durar un par de
 The boots. Who know what can last a pair of
 ellas, no mojándolas, ni manchándolas..? (18th c.)
 them, not wetting them, not dirtying them...?
 ‘The boots. Who knows how long can last a pair of them, neither getting
 them wet, nor getting them dirty...?’
 (*El buey suelto*, Pereda, José María de)

Under subjunctives, *lo que*-SFRs do not have an existential presupposition. They can have a Free Choice Interpretation (i.e. every proposition of the embedded sentence is a possible option):

- (28) Haga lo que quieras.
 do the that want
 ‘Do whatever you want!’
- (29) [...] cuando con ellos puede hacer lo que quiere
 when with them can do the that want
 ‘when he can do with them whatever he wants’
 (CDE, Huarte de San Juan, Juan. (16th..))

The latter observation shows that modality also plays a role in the interpretation of *lo que*-sentences, although their usual or default interpretation is the existential presupposition. This observation strongly suggests that the existential presupposition cannot be lexically encoded by the definite pronoun *lo* itself but must follow from the interaction with tense and modality. In this respect *lo* does not behave as a strong definite pronoun that lexically encodes the

existential presupposition (see Chierchia 1998, among others). Rather, it acts as a weak definite pronoun (see Carlson & Sussman 2005), because it is sensitive to the interaction with tense and modality.

This suggests that Keniston's 1937 hypothesis is on the right track, i.e. *lo que*-SFRs tend to refer more often to factual events than *qué*-interrogatives. However, we have shown that there are contexts where both types can have the same meaning. Both can refer to factual events and have existential presupposition in episodic contexts or they can have a Free Choice Interpretation in non-episodic or modal contexts.

3.7. Synchronic observations

Our research in CDE shows that a similar pattern to the one attested for Old Spanish can be found in Modern European Spanish of 20th c., i.e. factive predicates and perception verbs embed *lo que* more often than they embed *qué*-interrogatives, whereas the question predicate *preguntar* embeds *qué* more often than *lo que*.

Table 11: synchronic observations

Lemma	<i>lo que</i>	<i>qué</i>
<i>preguntar</i> 'ask'	543	3533
<i>saber</i> 'know'	18760	13860
<i>ver</i> 'see'	9941	8882
<i>no saber</i> 'not know'	5691	10948
<i>ver</i> +P	19.455	30.395
<i>saber</i> + P	2188	7889

3.8. Overall summary

We have observed that there is a clear difference in the frequency between predicates that embed *lo que*-SFRs and *qué*-interrogatives already noted in the literature (see Stulic-Etchevers 2007). Moreover, the distribution of the two types is different with respect to finite verbs, prepositions, tense and modality features. We have shown that these differences are maintained until Modern Spanish. We have also observed differences in semantics between the two types. *Lo que*-SFRs occur less frequently in modal contexts that usually do not trigger existential presupposition of the question. This observation goes hand in hand with the hypothesis that *lo que*-SFRs tend to refer to factual events.

4. Analysis of Q-SFRs

Before proposing an analysis of Q-SFRs, we would like to introduce the analyses of SFRs proposed in the literature in order to see whether they apply to Q-SFRs.

4.1. State of the art

Two classic analyses have been proposed for SFRs in the literature: the Head analysis and the Comp analysis (see Medeiros Junior 2014). The Head analysis assumes that the definite pronoun *lo* in *lo que*-relatives is the head of the relative clause (see also Bresnan & Grimshaw 1978, Rivero 1984, Suñer 1999, Munaro 2000, among others). According to this analysis, the pronoun *lo* first moves to Spec,CP and then to the DP:

(30) verbal predicate [DP lo [CP t_{lo} que [TP Verb t_{lo}]]]

According to the Comp analysis (cf. Medeiros Junior 2014), the complementizer *que* is a wh-element that moves to Spec,CP (against the usual analysis which locates it in C°, see Radford 2004) and the definite pronoun represents the head of the relative clause in D° (see also Groos & van Riemsdijk 1981, Rivero 1984 for Comp analysis):

(31) [DP [D° o [CP que_j C° [TPt_j]]]] Classic Portuguese

Both approaches analyze SFRs as DPs and not as CPs. However, the DP analysis is problematic for Q-SFRs and it remains unclear how DPs should be interpreted under question predicates. Caponigro (2004) gives a unitary interpretation for free relatives built with wh-morphemes (e.g. *I know/eat what you eat*). According to him, wh-morphemes neither denote existential quantifiers (see Hamblin 1973, Karttunen 1977) nor definite descriptions, but have the function of a set restrictor, restricting the set to +/- human individuals. It is the sentence embedding predicate that gives the relative or interrogative clause the existential or definite description interpretation. This analysis captures the formal similarities of free relatives and interrogatives. However, Caponigro (2004) is only concerned with FRs that have a wh-morphology and not with SFRs. It remains open how his analysis applies to SFRs.

4.2. Q-SFRs in Group 2

The languages belonging to Group 2 like Portuguese and Florentine have an interrogative syntax, and the interrogative interpretation does not depend on the matrix predicate but on formal features like [+wh] and [+Q], which mark the sentence as interrogative:¹⁰

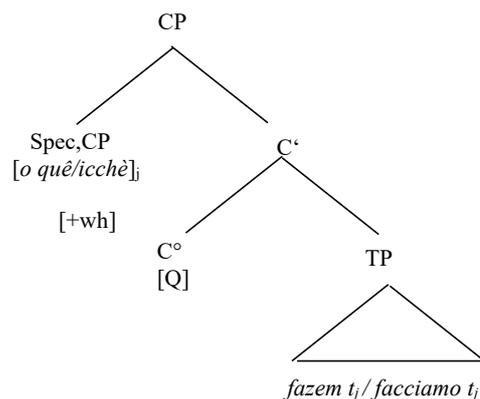


Figure 1 Q-SFRs in Portuguese and Florentine

We assume that the question semantics is derived compositionally by combining the wh-element which denotes an existential quantifier and the

¹⁰ We leave it open whether the wh-constituent [o que/icchè] should be further decomposed as proposed by Ambar and Veloso (2001) for Modern Portuguese.

complementizer phrase which marks interrogative force (see Appendix for compositional derivation of wh-questions):

- (32) O quê fazem/ Fl. icchè facciamo? ‘What shall we do?’ (Pt.)
 $\lambda p \exists x [p = \text{we shall do } x]$ ‘for which x is it the case that we shall do x?’

Since question predicates select for CPs and not DPs, prepositions can only be placed before the wh-element in indirect questions and not inbetween the definite marker and the complementizer *que/che*. This possibility is only possible in Group 2 (Pt. and Fl.), because only these languages have grammaticalized the definite marker and the complementizer into a wh-morpheme:

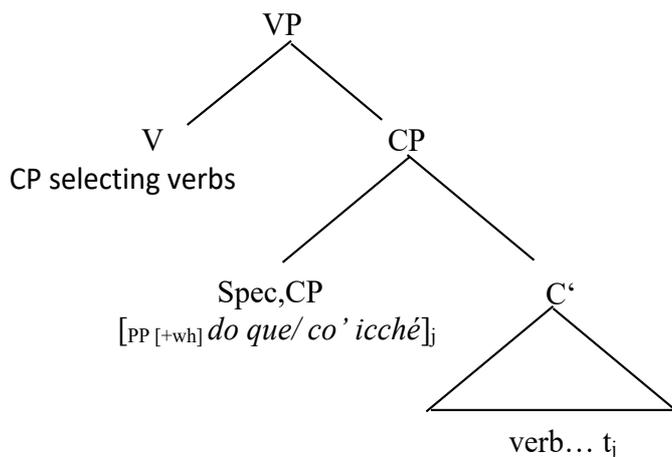


Figure 2 SFRs in Portuguese and Florentine

Q-SFRs under question predicates in Group 2 do not show any special interrogative syntax in comparison to embedded questions in other languages such as English (see Radford 2004 for an analysis of English wh-questions).

4.3. Q-SFRs in Group 1

To account for the difference between Q-SFRs in Group 1 and Group 2, we propose to analyze Q-SFRs in Group 1 as DP-like CPs. In order to show this point, we first illustrate the structure and the interpretation of SFRs under non-question predicates that represent type-shifted CPs (see figure 3).

The following structure of the Spanish example shows that the pronoun *lo* is the argument of the embedded verb *dicho* ‘said’ and it acts as relative operator which moves to Spec,CP (see Arregi 2000 for arguments of the Head-Raising Analysis of Relative Clauses in Spanish and the analysis of definite articles as relative operators in Spanish). As the whole CP is interpreted as a definite description and as an argument of the matrix copula verb *fue* ‘was’, the CP is typeshifted to DP (see Partee 1987 for typeshifting):

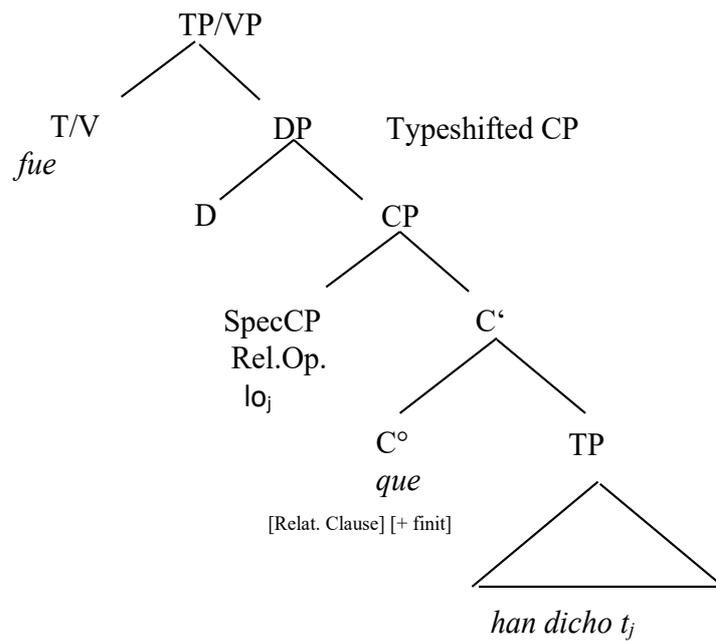


Figure 3 SFRs under ser-predicate in Modern Spanish

We analyze these DPs as definite descriptions, i.e. the copula verb *fue* takes a 3rd person subject pronoun (not realized in a null subject language such as Old Spanish and the *lo que*-sentence as its arguments. The CP is interpreted as a definite description and is typeshifted to DP (see Kellert 2015):

(33) BE (it)(*ip* they said *p*) ‘it is that they said the proposition *p*’

The definite description analysis of *lo que*-SFRs explains some observations about them, namely their tendency to occur under perception verbs, factive verbs (see section 3.3, tables 2-5), because all these verbs refer to factual events.

Our analysis of free relatives as typeshifted CPs to DPs is consistent with some syntactic analyses of free relatives as CPs that become DPs (see Ott 2011).¹¹

The challenge is how to analyze *lo que*-SFRs under question-embedding predicates.

We assume that Q-SFRs in Modern French and Spanish are CPs that share structural similarities with DPs (see figure 3) in that they have a DP-like structure of embedded questions, because they neither contain an interrogative complementizer (as for e.g. *est-ce que* in Modern French) nor a *wh*-pronoun (e.g. fr. *quoi*, sp. *qué* ‘what’) (see figure 4). We assume that the pronoun *lo* acts as a relative operator in embedded interrogatives (see Rivero 1984, Plann 1982, Suñer 1999):

¹¹ Ott (2011) assumes that free relatives become DPs after they have been derived as CPs as they do not have any interpretable features relevant for interpretation. As a consequence, they lose their CP label after they have been derived as CPs and act as DP-arguments for matrix verbs.

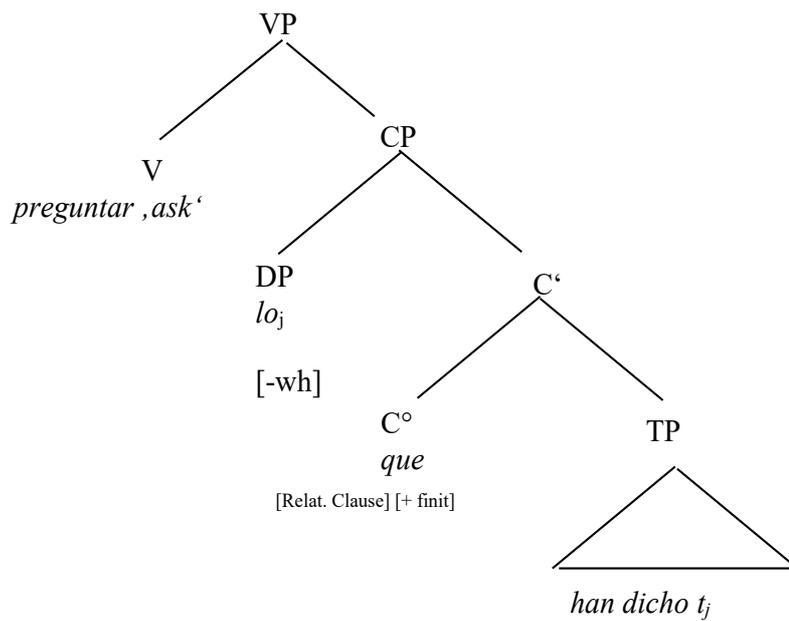


Figure 4 Q-SFRs in Modern Spanish

The CP-structure of Q-SFRs accounts for the coordination of Q-SFRs with *wh*-questions that are usually represented as CPs in generative grammar (see Radford 2004). The coordination of Q-SFRs and embedded *wh*-interrogatives is unproblematic because Q-SFRs are CPs and can thus be coordinated with embedded *wh*-interrogatives (see e.g. (19), repeated here in (44)):

- (34) Como no sabía [CP1 [Spec,CP de qué] era] & ni [CP2 SpecCP lo C que] (Sp.)
 because not knew of what was nor the that
 podía salir por allí]
 could sort out of it
 ‘Because I don’t know of what it consisted nor what could sort out of it.’
 (CDE)¹²

Moreover, the analysis in figure 4 can also account for prepositions preceding the pronoun *lo* which build a prepositional argument with the pronoun *lo* of the embedded verb:

¹² As an anonymous reviewer has pointed out, the coordination of Q-SFRs and embedded interrogatives could be interpreted as a semantic and not syntactic requirement, because question predicates can also embed DPs in Spanish although they might be interpreted as concealed questions on semantic level (i.e. he asked me what time it is).

i. Nos preguntó [la hora], [lo que habíamos dicho] y [dónde estábamos]
 us asked the time, the that have said and where were
 ‘He asked us what time it was, what we have said and where we have been.’

Indeed, our semantic analysis shows that Q-SFRs and embedded *qué*-interrogatives denote questions on the semantic level as shown in (46). Our analysis can thus capture the possibility of coordination in (44) under both assumptions, i.e. coordination is a semantic or syntactic requirement.

- (35) No sé [CP [PP de lo]_j que hablas t_j]
 not know of the that talk
 ‘I don’t know what you are talking about.’

The relative clause type of *lo que*-SFRs and the feature [-wh] of *lo* explains the impossibility of wh-in situ, multiple whs, and wh-modification.

With respect to the semantics of *lo que*-SFRs, we assume the following question meaning of Q-SFRs:

- (36) ASK ($\lambda p \exists x p = \text{they said } x$) ‘ask for which *x* is it the case that they said *x*?’

The question meaning of *lo que*-SFRs is very similar to wh-based interrogatives (see Appendix for compositional semantics of wh-questions based on Karttunen 1977). The only difference to wh-based interrogatives is that the interrogative meaning of *lo que*-SFRs receives its interrogative meaning from the question embedding predicate in order to ensure that *lo-que* SFRs never have the question meaning on their own and accounting this way to their use under non-interrogative matrix predicates (see figure 3).

However, we leave it open how the interrogative meaning of *lo que*-SFRs is derived by the composition of its parts and whether the existential quantifier, necessary to build the question semantics in (36), is expressed by the definite element *lo* under an interrogative matrix verb:

- (37) $L_{\text{Ounder interrogative verbs}} = \lambda P_{\langle e,t \rangle} \exists x[\text{thing}(w)(x) \ \& \ P(x)]$

We also leave it open for future research how to formalize the interpretations of *lo que*-SFRs in modal contexts (see section 3.6), namely that *lo que*-SFRs trigger Free Choice Interpretation in non-episodic contexts and a uniqueness presupposition of a definite description in episodic contexts.

4.4. Differences between Spanish and French

We noted a difference with respect to preposition placement in the initial position of the embedded clause in (16) and (17):

- (38) *Preposition + SFRs (Group 1)*
 a. Je ne sais pas de quoi tu parles. (Fr.)
 ‘I don’t know what you are talking about.’
 b. *Je ne sais pas de ce que tu parles. (Fr.)
 ‘I don’t know what you are talking about.’

Spanish, however, allows prepositions in embedded SFRs (see also Contreras 1999 and section 3.2):

- (39) a. No sé de qué hablas. (Sp.)
 not know of what talk
 ‘I don’t know what you are talking about.’

- b. No sé de lo que hablas. (Sp.)
not know of the that talk
'I don't know what you are talking about.'

We believe that the difference between French and Spanish is due to the fact that *ce* is a clitic in French (represented as $D^\circ ce$) in (50). The sentence in (38)b. represented here in (50) is ungrammatical because the preposition *de* in French always selects a strong pronoun (see Cardinaletti & Starke 1999 for strong pronouns) or a noun, whereas in (50) this requirement is not fulfilled:

- (40) * Je ne sais pas [_{SpecCP} [_{PP} P° de DP $D^\circ ce$] C° que tu parles].
I not know not of the that you talk

If the preposition embeds a strong wh-pronoun *quoi* 'what', the sentence is grammatical:

- (41) Je ne sais pas [_{SpecCP} [_{PP} P° de DP *quoi*] C° tu parles].
I not know not of what you talk
'I don't know what you are talking about.'

5. Summary and Outlook

We have identified two major groups of Romance languages with respect to semi-free relatives under question predicates (Q-SFRs). One group has grammaticalized the definite element and the element *que/che* into one compound with a [+wh] feature (Portuguese, Florentine) which explains different constraints not existent in the other group (French and Spanish) (see section 2).

We then looked at the evolution of Q-SFRs and tested some hypotheses from the literature according to which Q-SFRs started being used under a specific class of predicates (verbs of saying and perception verbs) and have generalized later over real question-embedding predicates (section 3). The spread of SFRs over question-embedding predicates presupposed the bleaching of the semantic feature [+referential] (section 3). Our data show that *lo que* sentences were used as relative clauses under non-question predicates such as *ser* 'be' and factive predicates such as *saber* 'know' always with much higher frequency than under genuine question predicates such as *preguntar* 'ask' in Old Spanish. There was no switch in frequency at a specific time. We have suggested that the interrogative interpretation of *lo que*-relative clauses has its source in the ambiguity of factive predicates. Factive predicates can select both DPs interpreted as definite descriptions and CPs interpreted as interrogatives. *Lo que*-relatives can thus be interpreted as definite descriptions and as interrogatives under factive predicates.

Moreover, we observed that prepositions and subjunctive were used more frequently with *qué* than with *lo que*-SFRs (3.5 and 3.6).

Finally, we suggested in section 4 that Q-SFRs are CPs with [+/-wh], [+/-Q] features. The typology of Romance languages results from the specification of these features, i.e. one group is more of a DP-type [-wh], [-Q] (Group 1) and the other group is more of a CP-type [+wh], [+Q] (Group 2).

In future, we would like to replicate our study on Old French, Old Portuguese and Renaissance Florentine. We will also investigate the variation in

Modern Spanish with respect to lexicalizing *lo que* as a wh-pronoun as well as other varieties of French.

References

- Alarcos Llorach, Emilio. 1984 [1970]. "Español que". In *ibid.*, *Estudios de gramática funcional del español*, 260-274. Madrid: Gredos.
- Ambar, Manuela & Veloso, Rita. 2001. "On the nature of wh-phrases-word order and wh-in situ Evidence from Portuguese, French, Hungarian and Tetum". In Y. D'Hulst, J. Rooryck, and J. Schroten (eds.), *Romance Languages and Linguistic Theory: Selected Papers from 'Going Romance' 1999, Leiden, 9-11 December*, 1-37. Amsterdam: John Benjamins.
- Arregi, Karlos. 2000. "Definite Articles as Relative Operators." Ms., Massachusetts Institute of Technology.
- Benincà, Paola. 2010. "Headless relatives in some Old Italian varieties". In R. D'Alessandro, A. Ledgeway, and I. Roberts (eds.), *Syntactic variation. The dialects of Italy*, 55-70. Cambridge: Cambridge University Press.
- Bresnan, Joan. & Grimshaw, Jane. 1978. "The Syntax of English Free Relatives". *Linguistic Inquiry* 9: 331-391.
- Caponigro, Ivano. 2004. "The Semantic Contribution of Wh-words And Type Shifts: Evidence from Free Relatives Crosslinguistically". *Proceedings of SALT* 14: 38-55.
- Cardinaletti, Anna & Starke, Michal. 1999. "The Typology of Structural Deficiency: On the three Grammatical Classes". In H. van Riemsdijk (ed.), *Clitics in the Languages of Europe*, 145-234. Berlin: De Gruyter.
- Carlson, Greg & Sussman, Rachel. 2005. "Seemingly indefinite definites". In S. Kepsar and M. Reis (eds.), *Linguistic Evidence*, 71-85. Berlin: De Gruyter.
- Chierchia, Gennaro. 1998. "Reference to kinds across languages". *Natural Language Semantics* 6: 339-405.
- Contreras, Heles. 1999. "Relaciones entre las construcciones interrogativas, exclamativas y relativas". In I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española vol.2*, 1931-1963, Madrid: Espasa Calpe.
- Egré, Paul. 2008. "Question-embedding and factivity". *Grazer Philosophische Studien* 77(1): 85-125.
- Fry, John. 2011. "CHI-SQUARE TEST. Linguistics 497: Corpus Linguistics, Spring 2011." Ms., Boise State University.
- García Cornejo, Rosalía. 2006. *Morfología y sintaxis de "que" en la Edad Media*. Sevilla: Universidad de Sevilla.
- Girón, Alconchel, J.L. 1988. *Las oraciones interrogativas indirectas en el español medieval*. Madrid: Gredos.
- González, P. 2003. Aspects on Aspect. Ph.D. thesis, University of Utrecht.
- Diemoz, Federica. 2007. *Morphologie et syntaxe des pronoms personnels sujets dans les parlers francoprovençaux de la Vallée d'Aoste*. Tübingen: Francke.
- Hamblin, Charles Leonard. 1973. "Questions in Montague English". *Foundations of Language* 10: 41-53.
- Homer, Vincent. 2008. "Disruption of NPI Licensing: The Case of Presuppositions". *Proceedings of SALT* 18: 429-446.
- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1: 3-44.

- Kellert, Olga. 2015. *Interrogative und Exklamative. Syntax und Semantik von multiplen wh-Elementen im Französischen und Italienischen*. Berlin: De Gruyter.
- Kellert, Olga & Lauschus, Sebastian. 2016. "The particle o in some Tuscan dialects: Fiorentino, Pisano and Crespinese". *Italian Journal of Linguistics* 28(2): 3-39.
- Kellert, Olga. 2017. "Interrogatives". In A. Dufter and E. Stark (eds.), *Morphosyntax and Syntax*, 569-603. Berlin: De Gruyter.
- Kiparsky, Paul & Kiparsky, Carol. 1970. "Fact". In M. Bierwisch and K. Heidolph (eds.), *Progress in Linguistics*, 143-173. The Hague: Mouton.
- Kunstmann, Pierre. 1990. *Le relatif-interrogatif en ancien français*. Geneva: Droz.
- Medeiros Júnior, Paulo. 2014. "from [o_que] to [o que]: free relatives in a diachronic view (approaching Brazilian Portuguese data)." Ms., Universidade Estadual de Campinas.
- Mattos e Silva, Rosa Virgínia. 1990. *Estruturas trecentistas: elementos para uma gramática do português arcaico*. Lisbon: Imprensa Nacional - Casa da Moeda.
- Munaro, Nicola 2000. "Free relatives as defective wh-elements. Evidence from the North-Western Italian dialects". *University of Venice Working Papers in Linguistics* 10(1): 89-120.
- Octavio de Toledo, Álvaro & Rodríguez Molina, Javier. 2017. "La imprescindible distinción entre texto y testimonio: el CORDE y los criterios de fiabilidad lingüística". *Scriptum digital* 6: 5-68.
- Ott, Dennis. 2011. "A note on free relative clauses in the theory of phases". *Linguistic Inquiry* 42(1): 183-192.
- Partee, Barbara. 1987. "Noun Phrase Interpretation and Type-shifting Principles". In: J. Groenendijk, D. de Jongh, & M. Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, 115-143. Dordrecht: Foris.
- Plann, Susan. 1982. "Indirect questions in Spanish". *Linguistic Inquiry* 13: 293-312.
- Poletto, Cecilia. & Pollock, Jean-Yves. 2009. "Another look at wh- questions in Romance: A look at Mendrisiotto". In D. Torck and W. L. Wetzels (eds.), *Romance Languages and Linguistic Theory 2006: Selected papers from 'Going Romance', Amsterdam, 7-9 December 2006*, 199-258. Amsterdam: John Benjamins.
- Quer, Josep. 1998. *Mood at the Interface*. The Hague: Holland Academic Graphics.
- Radford, Andrew. 2004. *Minimalist Syntax: Exploring the structure of English*. Cambridge: Cambridge University Press.
- Rebuschi, Georges. 2001. "Semi-Free Relative Clauses and the DP Hypothesis: Basque Evidence and Theoretical Consequences". In A. Grosu (ed.), *Papers from the Sixteenth Annual Conference and from the Research Workshop of the Israel Science Foundation [on] The Syntax and Semantics of Relative Constructions*, 55-64. Tel Aviv: Tel Aviv University.
- Riduejo, Emilio. 1987. "Sobre las oraciones interrogativas indirectas deliberativas en español medieval". In G. Lüdi (ed.), "*Romania Ingenuosa*": *Festschrift für Prof. Gerold Hilty zum 60. Geburtstag*, 365-383. Bern: Lang.
- Rivero, Maria Luisa. 1984 "Diachronic Syntax and Learnability: Free relatives in thirteenth century Spanish". *Journal of Linguistics* 20: 81-129.

- Rojo, Guillermo. 2011. "Frecuencia de inventario y frecuencia de uso". *Revista Española de Lingüística* 41(1): 5-43.
- Stulic-Etchevers, Ana. 2007. La grammaire de loke et siendo (ke) eb judéo-espagnol des balkans. Ph.D. thesis, Bordeaux: Université Michel de Montaigne–Bordeaux 3.
- Suñer, Margarita. 1999. "La subordinación sustantiva: la interrogación indirecta". In I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española*, 2149-2195. Madrid: Espasa Calpe.

Corpora

- COR = Cresti, E. / Moneglia, M. (eds.). .2005. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: John Benjamins.
- Davies, Mark. 2002. Corpus del Español (100 million words, 1200-1900). Available online at:
<https://www.corpusdelespanol.org>. (Last access 20 February 2018)
- CORDE: Corpus Diacrónico del Español. Available online at:
<http://corpus.rae.es/cordenet.html>. (Last access 20 February 2018)

Appendix

Compositional Analysis of wh-questions

Let us assume that the interrogative complementizer C° of genuine or usual questions has a *wh*-feature and a question feature ($[+wh, +Q]$) and it denotes an identity function of propositions:

$$(1) \quad \|C^\circ_{[+wh, +Q]}\| = \lambda p_{\langle s, t \rangle} \lambda q_{\langle s, t \rangle}. q=p$$

In order to interpret the *wh*-pronoun in Spec, CP and C' which denotes a set of propositions $\langle \langle s, t \rangle, t \rangle$ we assume *wh*-Quantifying-in, Q-in, (Karttunen 1977):

(2) If α is a daughter of CP with a *wh*-phrase β of type $\langle \langle e, t \rangle, t \rangle$ and an interrogative C' γ of type $\langle \langle s, t \rangle, t \rangle$ with an index i , then for every world w and assignment g , assume $\|\alpha\|^{w, g} = \lambda p \|\beta\|^{w, g} (\lambda x \|\gamma\|^{w, g} [x/1] (p))$.

(3) *Iché facciamo?* 'What should we do?'

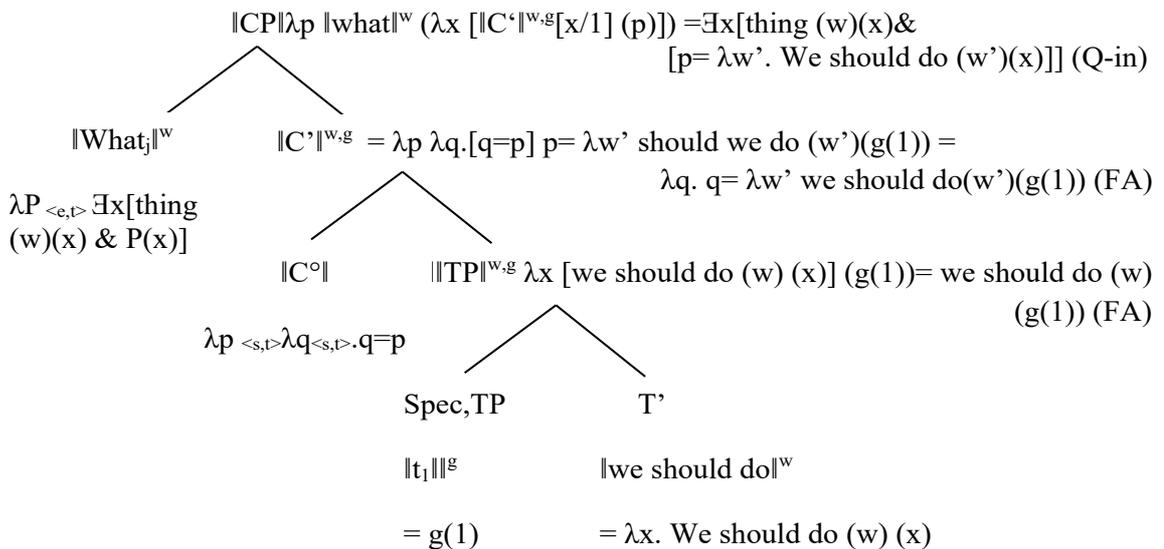


Figure 5 Compositional analysis of *wh*-questions