

# French subject island?

## Empirical studies of *dont* and *de qui*\*

Anne Abeillé & Elodie Winckel

LLF, Université de Paris &  
Humboldt-Universität zu Berlin and LLF, Université de Paris

### Abstract

*Dont* has been claimed to be an exception to the “subject island” constraint (Tellier, 1991; Sportiche and Bellier, 1989; Heck, 2009) and to contrast with true relative pronouns such as *de qui*. We provide corpus data from a literary corpus (FRANTEXT) which show that relativizing out of the subject is possible with *dont* and *de qui* in French relative clauses, and is even the most frequent use of both relative clauses. We show that it is not a recent innovation by comparing subcorpora from the beginning of the 20th century and from the beginning of the 21st century. We also show, with an acceptability judgment task, that extraction out of the subject with *de qui* is well accepted. Why has this possibility been overlooked? We suggest that it may be because *de qui* relatives in general are less frequent than *dont* relatives (about 60 times less in our corpus). Turning to *de qui* interrogatives, we show that extraction out of the subject is not attested, and propose an explanation of the contrast with relative clauses. We conclude that in this respect, French does not seem to differ from other Romance languages.

## Introduction

### *Dont* and *de qui*

There are several possibilities in French in order to extract a complement introduced by *de*, among which *dont* (‘of which’) and *de qui* (‘of who’) (but also *de quel*, *duquel*, etc.). The use of these two expressions obeys syntactic and semantic constraints: *dont* is restricted to relative

---

<sup>0</sup>We thank the anonymous reviewers for their comments and suggestions, as well as Bob Borsley, Rui Chaves, Danièle Godard, Ted Gibson, Jean-Pierre Koenig, Barbara Hemforth, Guido Mensching, Stefan Müller, Céline Pozniak, and Dominique Sportiche. We also thank the audiences of the GAC 2018, AFLS 2019, Quasy 2019 conferences, the 2019 LSA winter workshop on islands and the Berlin 2018 workshop on long-distance dependencies, at the Paris 2019 Linglunch and at the 2019 international HPSG conference, where parts of it were presented.

This work is supported by the German Research Foundation (DFG) as part of the research program “Long-distance Dependencies in French: Comparative Analyses (HPSG and the Minimalist Program)”, G. Mensching, Göttingen [ME 1252/14-1] / S. Müller, Berlin [MU 2822/9-1]. It was also supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris - ANR-18-IDEX-0001.

The data discussed in this article are available on <http://www.llf.cnrs.fr/fr/ressources>

clauses, whereas *de qui* can also be used in interrogatives ; *de qui* is restricted to animates, whereas *dont* is not:<sup>1</sup>

- (1) a. un ami **dont** / de qui je **tiens** ce récit  
 ‘a friend whom I got this story from’  
 b. les lieux **dont** / \*de qui on **se souvient**  
 ‘the places we remembers’

Furthermore, *dont* cannot introduce a free relative (2a), cannot be in situ and cannot be the complement of a preposition (2b): it is therefore not used in pied-piping.

- (2) a. Je me souviens de qui / \***dont** on m’a **dit** du bien.  
 ‘I remember who(ever) I hear good things about.’  
 b. cet homme **loin** de qui / \***dont** je suis  
 ‘the man who I am far from’

To account for these differences, it has been proposed that *dont* is a complementizer (Godard, 1988; Tellier, 1990) while *qui* is a *wh*-pronoun (Tellier, 1991; Le Goffic, 2007).

## Syntactic constraints on subject extraction?

The distinction between *dont* and *de qui* has played an important role in the discussion about the “subject island” constraint in generative grammar (Chomsky, 1973; Chomsky et al., 1977). If extraction out of the subject is banned cross-linguistically, why do certain languages allow it? In Italian for example, Rizzi (1982) has shown that extraction out of the subject is allowed (3a), and has related this to the pro-drop parameter (see also Stepanov 2007). From this perspective, French is a challenging case, since it is not a pro-drop language, contrary to other Romance languages. Godard (1988) has shown that *dont* allows extraction out of the subject (3b), and this can be related to the fact that it is a complementizer and not a true relative pronoun (see also Sportiche, 1981).

- (3) a. Questo autore, di cui so che **il primo libro** è stato pubblicato recentemente  
 (Rizzi, 1982, 61)  
 ‘this author, of whom I know that the first book has been published recently’  
 b. C’est un philosophe **dont un portrait** se trouve au Louvre (Godard, 1988, 47)  
 ‘It is a philosopher of whom a portrait is in Le Louvre’

Another analysis is proposed by Heck (2009). He considers that relativization of the complement of the subject is not an extraction : *dont* is a specifier of the subject phrase and the whole DP is pied-piped to the edge of the relative clause, as shown in example (4a). French *dont* thus lives a double life, since it permits ‘true’ extraction of complements of verbs and of object nouns (4b).

- (4) a. la fille [**dont le frère**]<sub>DP</sub> t’a rencontré (Heck, 2009, 101)  
 ‘the girl whose brother met you’

<sup>1</sup>In all examples we put in bold the item of which a dependent is relativized. We apply this convention also in examples cited from other authors.

- b. La fille dont tu as rencontré [**le frère**  $\_\_$ ]<sub>DP</sub>  
 ‘the girl whose brother you met’

According to Heck’s analysis, no material should intervene between *dont* and the subject when *dont* is a specifier (4a): no long distance dependency (5a) and no subject inversion (5b) is supposed to be possible when relativizing a complement of the subject. However, such constraints have not been tested empirically.

- (5) a. ?? un homme dont je refuse que **le fils** vous fréquente (Tellier, 1991)  
 ‘a man of whom I refuse that the son dates you’  
 b. \* Colin, dont choque **la coiffure blonde peroxydée** (Heck, 2009)  
 ‘Colin, whose bleached blond hair is shocking’

Turning to true relative pronouns (analysed with *wh*-movement in the generative tradition), Tellier (1990) claims that French differs from Italian, and that extraction out of the subject is not possible with *de qui* (6a, 6c), which only allows extraction out of an object (6b). A similar contrast is proposed in Sportiche and Bellier (1989).

- (6) a. ?\* C’est un linguiste de qui **les parents** ont déménagé à Chartres. (Tellier, 1991, 89)  
 ‘this is a linguist of whom the parents have moved to Chartres’  
 b. C’est un linguiste de qui vous avez rencontré **les parents**. (Tellier, 1991, 90)  
 ‘this is a linguist of who you have met the parents’  
 c. le diplomate dont / ?\*de qui **la secrétaire** t’ a téléphoné (Tellier, 1990, 307)  
 ‘the diplomat of who the secretary called you’

Again, this contrast has not been tested empirically. Following (Chomsky, 2008), the subject island constraint only applies to ‘true’ subjects, (7a) i.e. when the verb builds a *v*\*P (a phase). The subject of a transitive verb is in (Spec,*v*\*P) and no element from the subject can move to the edge of the phase, thus blocking extraction. Under this theory, it is possible to extract out of the subject of a passive (7b) or an unaccusative verb, because their subject is in the domain of *v* in the deep structure (see however Legate (2003) who argues that unaccusatives and passives *v*Ps are also phasal). Similar approaches (Polinsky et al. (2013), Uriagereka (2012) among others) assume that extraction out of the subject of transitive verbs is dispreferred cross-linguistically. Applied to French, this theory predicts a difference between (6c) and (7c).

- (7) a. \* Of which car did **the driver** cause a scandal ? (Chomsky, 2008, 147)  
 b. Of which car were **the hoods** damaged by the explosion ? (Ross, 1967, 242)  
 c. l’homme dont/ de qui **la soeur** est venue (Sportiche, 2011):(30b)  
 ‘the man whose sister came’

Using Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994), Sag and Godard (1994) and Godard and Sag (1996) posit no subject constraint for French. They provide good examples with *dont* (8) and predict long distance extraction such as (5a) to be possible, but do not discuss transitive verbs nor *de qui* relatives.

- (8) le Dr X dont la maison de Le Corbusier peut être visitée  
'Dr X, whose house by Le Corbusier can be visited'

Sportiche (2011) argues that *dont* is a weak pronoun, and that there are cases where both *dont* and *de qui* are possible (7c); however, he does not discuss the subject island constraint.

Many syntacticians thus assume subtle contrasts between good and bad cases with *dont* (5) on one hand, between good and bad cases with *de qui* (6) on the other hand. However, these contrasts rely on linguists' intuitions and have not been tested rigorously.

## Aim of this paper

Given the lack of convergent data in the theoretical literature, empirical studies are needed in order to choose between a theory that claims that French differs from other Romance languages in obeying a 'subject island constraint' and a theory that claims that it does not. The questions to be addressed are the following : Is extraction out of the subject allowed in French with *de qui* as well as with *dont* ? Is it restricted to unaccusative verbs and passives? Must *dont* be adjacent to the subject in this case? The aim of this paper is to provide new empirical data on the current use of *de qui*, and to compare it with *dont*. We first review Abeillé et al. (2016), which provides two corpus studies on the present use of *dont*. We then present two new corpus studies, using Frantext which is a large database of literary texts, from well-known authors, which are well edited. We compare the present day use (2000-2013) of *de qui* with that of the last century (1900-1913). We also ran a new experiment on *de qui*, using controlled acceptability judgements (Gibson and Fedorenko, 2013; Sprouse and Almeida, 2017) and compare it with an experiment on *dont* reported in Abeillé et al. (2020).

## 1 Some corpus results on *dont*

Abeillé et al. (2016) present the distribution of *dont* in a newspaper corpus, the *French Treebank* (FTB; Abeillé et al., 2019), and in a spoken corpus, the *Corpus Français Parlé Parisien des années 2000* (CFPP2000; Branca-Rosoff et al., 2012) of contemporary French. An overview of their results is given in Table 1. <sup>2</sup> They find examples where *dont* can relativize the complement of the subject noun, and surprisingly, this usage is the most frequent one in the French Treebank overall (9a) and the most frequent one when relativizing a complement of a noun in the spoken corpus (9b). They propose that the lower percentage of *dont* for subjects in the spoken corpus is due to the lower proportion of nominal subjects (clitic subjects are more frequent in spoken French, and no extraction is possible out of them).

- (9) a. [...] des coopératives dont **les membres** exploitaient sous forme privée des lopins de terre. (FTB-flmf7ab1co-440)  
'cooperatives whose members exploited privately some pieces of land'
- b. [...] des gens dont **les parents** avaient des grosses fortunes (CFPP2000)  
'people of whom the parents had great wealth'

Relativizing the complement of	Nb in FTB	%	Nb in CFPP2000	%
Verb	73	19%	67	63%
Noun				
- Subject	217	57%	19	18 %
- Object	48	13%	9	8%
- Predicate	27	7%	1	1%
Adjective	6	2%	3	3%
Adjunct	10	3%	7	7%
Total	384	100%	107	100%

Table 1: *dont* relatives in FTB and CFPP2000

Looking more closely at the examples of relativization out of the subject, they found a significant proportion of transitive verbs (35% of all extractions out of subject in the FTB, and 33% in the CFPP2000).

In order to compare *dont* and *de qui* in contemporary French, we have searched for *de qui* relatives in the same two corpora used by Abeillé et al. (2016). There is only 1 occurrence of *de qui* in the FTB and 11 in CFPP2000. Leaving aside interrogative uses and dysfluencies, there are only 3 relatives in CFPP2000: 1 relativizing the complement of adjective, 1 the complement of the object noun, and 1 with pied-piping. This is why we turn to a larger corpus.

## 2 *De qui* relatives in Frantext (2000-2013)

We use the Frantext online database ([www.frantext.fr](http://www.frantext.fr)), which comprises well written and well edited texts from various authors. Searching first the 2000-2013 subcorpus (222 texts, 13.2 million words), we found 449 occurrences of *de qui*, among which 201 relatives, and we annotated them manually. Their distribution is presented in Table 2.<sup>3</sup>

Type of clauses	Nb	%
relative clauses	201	44.77
direct questions	78	17.37
indirect questions	51	11.36
others	119	26.50
Total	449	

Table 2: *De qui* in Frantext 2000-2013

We annotated the *de qui* relatives for syntactic functions : *de qui* can relativize the complement of a verb (10a), of a noun (10b), of an adjective (10c) or of a preposition (10d).

- (10) a. [...] les puissants, de qui il avait cru pouvoir se **séparer** [...] (Pense à demain, Garat, 2010)  
‘the powerful people, from who he thought he could part’
- b. [...] la personne de qui elle porte **le nom** [...] (Pense à demain, Garat, 2010)  
‘the person whose name she bears’

<sup>2</sup>The authors have excluded gapless and verbless relative clauses from their results.

<sup>3</sup>The label ‘others’ includes *qui* free relatives (*le geste craintif de qui cherche secours* ‘the fearfull gesture of who is seeking help’), and free choice uses (*qui que ce soit* ‘whoever’).

- c. [...] un groupe de mères de qui l' inquiétude la tenait **éloignée** [...] (L'enfant des ténèbres, Garat, 2010)  
 'mothers, from who her concern made her stay away'
- d. [...] Jean-Claude Passeron, **autour** de qui gravitaient [...] quelques membres de la même promotion [...] (Bardadrac, Genette, 2006)  
 'Jean-Claude Passeron, around which some schoolmates were gravitating'

When *de qui* relativizes the complement of a noun, we also annotate the function of the noun. The distribution is reported in Table 3.<sup>4</sup>

Relativizing the complement of	Nb	%
Verb	32	15.92
Noun		
- Subject	54	26.87
- Object	31	15.42
- Predicate	8	3.98
- Cplt of prep	37	18.41
Adjective	2	1.00
Preposition	30	14.93
others	7	3.48
Total	201	

Table 3: *De qui* relatives in Frantext 2000-2013

Strikingly, relativizing the complement of the subject is the most frequent use of *de qui* relative clauses: 54 cases are relativization out of the subject, i.e. 26.87%, compared to 15.92 % for the complement of a verb and 15.42 % for the complement of an object noun. This is even more striking if we exclude subjects out of which it is impossible to extract (clitics and proper nouns). Among 85 candidate subjects, 54 have their complement relativized, i.e. 62.4%. This is similar to what Author1 et al. (2016) have observed for *dont* in newspaper texts.

- (11) a. [...] les ogres de qui **la danse barbare** [...] vous confisque l' enfance. (Pense à demain, Garat, 2010)  
 'the ogres whose barbaric dance takes your childhood away from you'
- b. [...] cette enfant de qui l'**aïeule** était assise près de lui [...] (L'enfant des ténèbres, Garat, 2008)  
 'this girl of who the grandmother sat next to him'
- c. [...] son mari excentrique, de qui **la collection orientale** dormait au musée de l'Homme. (Pense à demain, Garat, 2010)  
 'her excentric husband, of who the oriental collection slept in the Musée de l'Homme'

Contrary to Sportiche and Bellier (1989), Tellier (1991) or Heck (2009), we found 54 examples of relativizing out of the subject, most of them (23 occurrences) with a transitive verb (11a), some with an unaccusative verb (11b), all with a preverbal subject. Table 4 shows that this is

<sup>4</sup>The label 'others' includes relative clauses with several gaps (also called parasitic gaps).

not restricted to unaccusative verbs and passives, contrary to Chomsky (2008) and Uriagereka (2012). The transitive verbs alone represent 42.59% of the verbs among the relativization out of a subject. They represent 51.74% of the verbs in the relative clauses with *de qui* overall, and a Fisher’s Test fails to show a significant difference between the two percentages ( $p = 0.1516$ ).

Type of verb	Nb	%
passive	3	5.56
unaccusative	2	3.70
mediopassive	4	7.41
transitive	23	42.59
unergative	11	20.37
copula (être)	11	20.37

Table 4: Type of verbs with *de qui* relativization out of subjects

### 3 *Dont* relatives in Frantext (2000-2013)

In order to compare *de qui* with *dont*, we ran a second study on the same subcorpus (Frantext 2000-2013). Since *dont* has more than 13 000 occurrences, we randomly selected a subset of 500 occurrences, among which we selected the ones with an animate antecedent, in order to have a meaningful comparison with *de qui*. 143 relative clauses had an animate antecedent, we annotated the function of the relativized element, and the function of the noun when a complement of noun is relativized. Compare the distribution in Table 5 with Table 3.<sup>5</sup>

Relativizing the complement of	Nb	%
Verb	25	17.48
Noun		
- Subject	61	42.66
- Object	31	21.68
- Predicate	4	2.80
Adjective	3	2.10
others	19	13.29
Total	143	

Table 5: *dont* in a subset of Frantext 2000-2013 (with an animate antecedent)

Here again, relativizing the complement of the subject is possible (12) and it is the most frequent usage of *dont*, as in Author1 et al. (2016). Among 73 subjects out of which it is possible to extract (with a common noun), 58 have their complement extracted (79.50%).

- (12) [...] la belle Antillaise dont l’effigie orna [...] les boîtes de Banania. (La solitude de la fleur blanche, Roux, 2009)

‘the nice Caribbean girl, whose picture decorated the packages of the Banania brand’

Like *de qui* relative clauses, relativizing the complement of a subject with *dont* is not restricted to passive or unaccusative verbs, as can be seen in Table 6. The transitive verbs alone represent 24.59% of the verbs among the relativizations out of subjects. They represent

<sup>5</sup>The category ‘others’ include relative clauses without a gap or with several gaps.

37.76% of the verbs in the relative clauses with *dont* overall, and a Fisher’s Test shows that the difference is significant ( $p < 0.001$ ). However, it would be difficult to claim that transitive verbs are ruled out, since their frequency is non-marginal (this is confirmed by a Binomial Test).

Type of verb	Nb	%
passive	8	13.11
unaccusative	6	9.84
mediopassive	2	3.28
transitive	15	24.59
unergative	16	26.23
copula (être)	14	22.95

Table 6: Type of verbs for relativizing out of subjects with *dont*

Heck (2009)’s hypothesis predicts that no material should stand between *dont* and the subject. No extraction out of a postverbal subject or out of the subject of an embedded clause should be possible. But we find 4 such postverbal subjects in the corpus (13a-b); and one long-distance dependency (13c).

- (13) a. [...] les morts, dont se dissout dans l’air **la présence** [...] (Dans la main du diable, Garat, 2006)  
‘the dead ones, of who vanishes in the air the presence’
- b. [...] Elmosnino lui- même, dont nous charment **la voix, la présence, le mouvement, la malice**. (Voix off, Podalydes, 2008)  
‘Elmosnino himself, of who delight us the presence, the gesture, the maliciousness’
- c. [...] madame Segond-Weber, la grande tragédienne, dont il aime rappeler que **les répliques** tombaient de sa bouche « comme des fûts de colonne ». (Voix off, Podalydes, 2008)  
‘Madame Segond-Weber, the great tragedian, of who he enjoys recalling that the lines fall out of her lips « like fluted columns »’

In order to compare *de qui* and *dont* relative clauses, we exclude *de qui* relative clauses with pied-piping. If we leave aside relativizing the complement of a preposition (*autour de qui* ‘around whom’) (14.93%) and the complement of a noun which is introduced by a preposition (*avec le fils de qui* ‘with the son of whom’) (18.41%) in Table 3, the proportion of relativizing the complement of the subject with *de qui* is 40.30%, which is very close to *dont* (42.66%). This similarity between *de qui* and *dont* (Tables 3 and 5) is confirmed by a statistical test: A Fisher’s Test fails to find a significant difference between the amount of relativization of the complement of the subject in the *dont* relative clauses and in the *de qui* relative clauses ( $p = 0.7156$ ).

Relativizing the complement of the subject is thus possible with *de qui* as well as with *dont*. It is the most frequent use of *de qui* relatives (40.30 % if we leave aside piedpiping uses), as it is the most frequent use of *dont* relatives in our corpus (42.66 %). Furthermore, it is not limited to unaccusative verbs, contrary to Chomsky (2008). In this case, transitive verbs are even more frequent with *de qui* than with *dont* (a test shows the difference is marginally significant



$p < 0.05$ ). With *dont*, it is possible with postverbal subjects, contrary to Heck (2009). Thus, there does not seem to be syntactic restrictions on relativizing the complement of the subject in present day French.

Why has it been overlooked? It is true that *dont* relatives are much more frequent than *de qui* relatives overall, and that the use of *de qui* in pied-piping (67 examples) slightly outnumbers the relativization out of the subject (54 examples). Maybe this is why French linguists have overlooked this possibility with *de qui*. Or maybe it is because it might be a new use of *de qui*, since these French linguists wrote on *dont* and *de qui* at the end of the 20th century. This is why we consider a previous time period in Frantext.<sup>6</sup>

## 4 *De qui* relatives in Frantext (1900-1913)

Since linguists of the 20th century (Tellier, 1990, 1991) have considered *de qui* as unacceptable when relativizing the complement of a subject, and whereas our analysis in the previous section showed that this use of *de qui* is relatively common in contemporary French, we consider here the possibility that it has arisen in the language only recently. This is why we ran another corpus study on Frantext, targeting the period 1900-1913 (179 texts, 7.8 million words). The query *de qui* in Frantext for 1900-1913 gave us 271 occurrences, which we annotated manually. Their distribution is presented in Table 7: 171 occurrences of *de qui* are relative clauses, i.e. 63.10% (vs. 44.77% of the occurrences of *de qui* in Frantext 2000-2013). A Fisher's test show that there is a significant difference between the two periods ( $p < 0.001$ ), so that there is a significantly higher proportion of relative clauses for 1900-1913 than for 2000-2013<sup>7</sup>.

Type of clauses	Nb	%
relative clauses	171	63.10
free relative clauses	3	1.11
direct questions	47	17.34
indirect questions	23	8.49
cleft sentences	5	1.85
others	22	8.12
Total	271	100

Table 7: *De qui* in Frantext 1900-1913

As in contemporary French, *de qui* can relativize any kind of complement: the complement of a verb (14a), of a noun (14b), of an adjective (14c) or of a preposition (14d). The distribution can be seen in Table 8<sup>8</sup>.

- (14) a. [...] celui de qui **décou**le toute consolation réelle [...] (La porte étroite, Gide, 1909)  
 'the one from who all real comfort results'

<sup>6</sup>We observe that all our examples of *de qui* relativization out of the subject are from books by the same author: Anne-Marie Garat. Overall, this author uses *de qui* very much: 147 *de qui* relatives (73.87% of our corpus results) are from Garat. We will see in the next section that other authors also use *de qui* relative clauses for the complement of the subject.

<sup>7</sup>As in Table 2, the label 'others' is for *qui* (and not *de qui*) free relatives and free choice uses.

<sup>8</sup>The label 'others' includes *qui* (and not *de qui*) free relatives, as well as free choice uses

Relativizing the complement of	Nb	%
Verb	30	17.54
Noun		
- Subject	38	22.22
- Object	16	9.36
- Predicate	4	2.34
- Cplt of prep	42	24.56
Adjective	1	0.58
Preposition	37	21.64
others	3	1.75
Total	171	100

Table 8: *De qui* relatives in Frantext 1900-1913

- b. [...] Berthe Hochard, de qui elle a redressé **la serviette** [...] (La Maternelle, Frapié, 1904)  
‘Berthe Hochard, whose towel she readjusted’
- c. [...] du ”monsieur” de qui son père semblait **inférieur**. (Mes Cahiers, Barrès, 1902)  
‘the sir to whom his father seemed to be inferior’
- d. [...] ceux **à l’encontre** de qui il s’exerce [...] (Du côté de chez Swann, Proust, 1913)  
‘those against who it wields’

Among 171 relative clauses, 38 relativize the complement of the subject (15), i.e. 22.22%. This percentage is a bit lower than in Frantext 2000-2013, but a Fisher’s Test fails to show a significant difference between the two periods ( $p = 0.3355$ ). However, if we only take into account subjects out of which it is possible to extract (excluding proper names and clitics), the difference with Frantext 2000-2013 is bigger. Among 95 such subjects in the relative clauses, 38 show a relativization of their complement, i.e. 40% (vs. 62.4% for Frantext 2000-2013). We thus conclude that extraction out of the subject has increased with *de qui* over time, and a Fisher’s Test shows a significant difference between the two periods ( $p = 0.003$ ).

- (15) a. [...] l’homme aux yeux durs, de qui **les grands pas** traversaient les pièces.  
(L’Enfant d’Austerlitz, Adam, 1902)  
‘the man with the hard eyes, whose big steps went through the rooms’
- b. [...] un légiste inattaquable, mais de qui **les victimes** tout de même intéressent [...] (Mes Cahiers : t. 9 : 1911-1912, Barrès, 1912)  
‘an unassailable jurist, but of who the victims however arise interest’
- c. [...] toi de qui **l’âme** est si merveilleusement jumelle de la mienne [...] (Le journal d’une femme de chambre, Mirbeau, 1900)  
‘oh my friend, you of who the soul is so wonderfully twin with mine’

There is no difference in the type of verbs involved in those relativizations out of the subject between 1900-1913 and 2000-2013 (compare Table 9 to Table 4). Transitive verbs represent 36.84% among the relativizations of the complement of the subject, a bit more than in Frantext

2000-2013 (Table 10), but this difference is not significant (a Fisher’s Test fails with  $p = 0.6678$ ). Transitive verbs represent 49.12% in all *de qui* relative clauses in Frantext 1900-1913: this is not significantly different from their amount when relativizing out of the subject (a Fisher’s Test fails with  $p = 0.1398$ ).

Type of verb	Nb	%
passive	2	5.26
unaccusative	1	2.63
mediopassive	4	10.53
transitive	14	36.84
unergative	9	23.68
copula (être)	8	21.05

Table 9: Type of verbs among the relativizations out of subjects

transitive verbs	1900-1913	2000-2013
Relativization out of the subject	36.84%	42.59%
All <i>de qui</i> relatives	49.12%	51.74%

Table 10: Transitive verbs in both periods for *de qui*

We also find 3 examples of relativizing out of the inverted subject, such as (16).

- (16) [...] ce vieil impotent, [...] de qui tremblait doucement **la grosse lèvres blême**.  
(L’Enfant d’Austerlitz, Adam, 1902)  
‘this old paralysed man, of who quivers softly the big pallid lip’

Overall, we find a high amount of relativizing out of a subject with *de qui*, which does not seem to be an innovation, even if it has increased over time. We cannot find any restriction on the type of verbs involved in these relative clauses (contra Chomsky, 2008), as in Frantext (2000-2013).

## 5 *Dont* relatives in Frantext (1900-1913)

We now compare *de qui* with *dont* for this second time period. Frantext 1900-1913 has close to 10 000 occurrences of *dont*. We randomly selected a subset of 1300 occurrences, and excluded relatives with an inanimate antecedent. This left us with 179 relative clauses. Their distribution is presented in Table 11.

56.42% of these relative clauses show an extraction out of a subject NP (17). This is more than in Frantext 2000-2013 (42.66%) and a Fisher’s Test shows that this difference is significant ( $p = 0.0184$ ). Furthermore, among 118 subjects out of which it is possible to extract, 101 are extracted (85.59%). This is more than in Frantext 2000-2013 (79.50%), but the difference is not significant (a Fisher’s test fails with  $p = 0.3199$ ).

- (17) [...] [l]es Bienheureux dont **la langueur** avait le pouvoir de guérir toutes les langueurs [...]. (Exégèse des lieux communs, Bloy, 1902)  
‘the blessed ones, whose languor had the power to cure any languor’

Relativizing the complement of	Nb	%
Verb	19	10.61
Noun		
- Subject	101	56.42
- Object	38	21.23
- Predicate	4	2.23
Adjective	6	3.35
others	11	0.061
Total	179	100

Table 11: *Dont* relatives in Frantext 1900-1913

We also find one case of subject inversion when relativizing out of the subject (18, which contradicts the adjacency constraint proposed by Heck (2009)).

- (18) [...] des monstres cabrés, dont se reconnaissent déjà **les silhouettes**. (Pierre Loti, L’Inde (sans les Anglais), 1903)

‘the prancing monsters, of who can be recognized the shapes

As in Frantext 2000-2013, different types of verbs are attested in the 101 relativizations out of a subject (Table 12). Transitive verbs alone represent 33.66% of the verbs, vs. 45.81% of the verbs in all relative clauses with *dont*. A Fisher’s test shows that the difference is significant ( $p < 0.001$ ). As in Frantext 2000-2013, relativizing with *dont* out of the subject tends to have less transitive verbs than other kind of relativizations.

Type of verb	Nb	%
passive	9	8.91
unaccusative	5	4.95
mediopassive	9	8.91
transitive	34	24.59
unergative	21	20.79
copula (être)	23	22.77

Table 12: Type of verbs for relativizing out of the subject

Table 13 summarizes the distribution of transitive verbs for both periods. There are more transitive verbs when relativizing out of the subject in Frantext 1900-1913 than in Frantext 2000-2013, but the difference is not significant (a Fisher’s Test fails with  $p = 0.2897$ ). As far as transitive verbs are concerned, we can therefore see a stable tendency: *de qui* and *dont* relativizing the complement of the subject have significantly less transitive verbs than the rest of the *de qui* and *dont* relative clauses, but still have a high and non-marginal amount of transitive verbs. In our corpus, this tendency does not evolve across time.

transitive verbs	1900-1913	2000-2013
among the relativizations out of subjects	33.66%	24.59%
among all <i>dont</i>	45.81%	37.76%

Table 13: Transitive verbs in both periods for *dont*

The distributions of *de qui* and of *dont* relative clauses in the time period 1900-1913 are

very similar, even if relativizing out of a subject with *de qui* (41.30% of the relative clauses, excluding piedpiping) is less frequent than with *dont* (56,42%) (the difference is significant, as confirmed by a Fisher’s Test, with  $p = 0.0211$ ). For the 2000-2013 period, the difference between both kinds of relative clauses for relativizing out of subject was not significant: 40.30% for *de qui* (excluding piedpiping) and 42.66 % for *dont*.

Overall, our corpus studies show a high amount of relativizing out of a subject with *de qui*, in a proportion roughly similar to relativizing out of a subject with *dont* (contra Sportiche and Bellier, 1989; Tellier, 1990, 1991). Comparing our two periods (1900-1913 and 2000-2013), this proportion is increasing for *de qui* and decreasing for *dont* (with an animate antecedent). We cannot find any restriction on the type of verbs involved in these relatives (contra Chomsky, 2008; Uriagereka, 2012): when relativizing out of the subject, transitive verbs were even more frequent with *de qui* than with *dont* for both periods. We have also found examples with a postnominal subject, both for *dont* and *de qui*, which contradict the adjacency constraint proposed by (Heck, 2009) for *dont*. Overall, our data contradict the syntactic constraints proposed by Sportiche and Bellier (1989), Tellier (1990, 1991) and Heck (2009).<sup>9</sup>

## 6 Experimental results

Using literary texts from Frantext, which are very well written and edited, we found that *de qui* relatives corresponding to the complement of the subject are well attested in both 20th and 21st century French, even though they are less numerous than with *dont*. One possible reason for the discrepancy between our results and those from the linguistic literature (Tellier, Sportiche) is that we have relied on corpus data while theoretical linguists generally rely on acceptability data. To see whether there really is a difference between corpus data and acceptability data in this domain, we ran a controlled acceptability experiment (Hemforth, 2013; Sprouse and Almeida, 2017) on *de qui* relative clauses. We first report the experiment on *dont* reported in Abeillé et al. (2018) in order to draw a meaningful comparison.

### 6.1 An experiment on *dont*

Abeillé et al. (2016) present an acceptability judgement task, with a follow-up experiment in Abeillé et al. (2018); Abeillé et al. (2020), which we summarize here. We tested 24 items in the following 2 (subject/object) x 3 conditions : with *dont* for the complement of the noun, with *que* instead of *dont* (ungrammatical control), with clause coordination (grammatical control).

(19) a. *dont* + subject:

Un célèbre pâtissier a mis au point une recette dont la simplicité  
a famous pastry.chef has invented a recipe of.which the simplicity

---

<sup>9</sup>Looking for other factors that may govern the choice between *dont* and *de qui* relative clauses with an animate antecedent, we annotated several properties: relative clause restrictiveness (appositive/restrictive), antecedent number, antecedent definiteness, verb type (transitive, unaccusative, unergative...) and subject inversion. We ran several generalized linear regression models (Sakamoto et al., 1986). The variable to be explained was always the relative word. Only one factor appeared significant in Frantext (2000-2013): an appositive relative is more likely to have *de qui* than *dont*, and another factor was marginally significant: a relative with a transitive verb is more likely to have *de qui* than *dont*. In Frantext 1900-1913, however, these factors were not significant.

- ravit les apprentis depuis des générations.  
delights the apprentices since several generations
- b. dont + object:  
Un célèbre pâtissier a mis au point une recette dont les apprentis  
a famous pastry.chef has invented a recipe of.which the apprentices  
aiment **la simplicité** depuis des générations.  
like the simplicity since several generations
- c. coord + subject:  
Un célèbre pâtissier a mis au point une recette, et sa simplicité  
a famous pastry.chef has invented a recipe and its simplicity  
ravit les apprentis depuis des générations.  
delights the apprentices since several generations
- d. coord + object:  
Un célèbre pâtissier a mis au point une recette, et les apprentis  
a famous pastry.chef has invented a recipe and the apprentices  
aiment sa simplicité depuis des générations.  
like its simplicity since several generations
- e. que + subject:  
Un célèbre pâtissier a mis au point une recette que la simplicité  
a famous pastry.chef has invented a recipe that the simplicity  
ravit les apprentis depuis des générations.  
delights the apprentices since several generations
- f. que + object:  
Un célèbre pâtissier a mis au point une recette que les apprentis  
a famous pastry.chef has invented a recipe that the apprentices  
aiment la simplicité depuis des générations.  
like the simplicity since several generations

We chose transitive verbs that come in reverse subject/object pairs (*ravir/aimer* ‘delight/-like’), and nouns denoting qualities (*importance, beauté, simplicité* ‘importance, beauty, simplicity’). We tested 24 experimental items randomly mixed with 24 distractors. We ran an online acceptability rating task (using a 1–10 scale similarly to the French school system) on Ibex (<http://spellout.net/ibexfarm/>), using the RISC platform (<https://expesciences.risc.cnrs.fr/>) and social medias to recruit 48 participants.

The results show a subject advantage for *dont* relative clauses, such that extractions out of the subject (19a) were rated marginally better than extractions out of the object (19b). Both the *dont* relative clauses and the coordination variants (19c,d) were rated significantly higher than the ungrammatical controls with *que*.

## 6.2 An experiment on *de qui*

Using a design similar to that of Abeillé et al. (2018); Abeillé et al. (2020), we ran an acceptability judgement task, comparing extracted variants with clausal coordination (no extraction) and ungrammatical control (missing preposition *de*) conditions, both for subject and object (2x3 design). We chose relational human nouns (*compagne* ‘partner’, *cousin* ‘cousin’) for both subjects and objects, to avoid animacy mismatch, and reversible transitive verbs (*fréquenter* ‘frequent’, *connaître* ‘know’) so we had the same verb in subject and object conditions.

- (20) a. de qui + subject:

- J' ai un voisin de qui **la compagne** connaît ma cousine.  
 'I have a neighbour of whom the partner knows my cousin'
- b. de qui + object:  
 J' ai un voisin de qui ma cousine connaît **la compagne**.  
 'I have a neighbour of whom my cousin knows the partner.'
- c. coord + subject:  
 J' ai un voisin, et la compagne de ce voisin connaît ma cousine.  
 'I have a neighbour, and the partner of this neighbour knows my cousin.'
- d. coord + object:  
 J' ai un voisin, et ma cousine connaît la compagne de ce voisin.  
 'I have a neighbour, and my cousin knows the partner of this neighbour.'
- e. qui + subject:  
 J' ai un voisin qui la compagne connaît ma cousine.  
 'I have a neighbour who the partner knows my cousin'
- f. qui + object:  
 J' ai un voisin qui ma cousine connaît la compagne.  
 'I have a neighbour who my cousin knows the partner'

We had 3 practice items, 24 target items and 24 distractors. The items and the distractors were randomized in 6 lists, using a Latin square design. Each participant saw each item in exactly one condition. Participants were asked to rate the sentences from 0 (not natural) to 10 (perfectly natural). We conducted the study on internet (Ibex, Drummond 2010) with 28 participants, 23 women and 5 men, aged 18 to 75, who volunteered on the RISC platform (<https://expesciences.risc.cnrs.fr/>) and over social medias.

The results of the experiment are reported in Figure 1, the bars indicating 95% confidence intervals. Running maximal linear mixed models (Baayen et al., 2008)<sup>10</sup> with crossed random slopes for subjects and items (Barr et al., 2013), we first compared the extraction out of subject (20a) with extraction out of object (20b) and found no significant difference. We fit two 2x2 analyses, comparing extraction first with grammatical controls, and second with ungrammatical controls. Again, we found no main effect of grammatical function (subject, object), but in each case a effect of extraction type, such that the *de qui* relatives were rated lower than grammatical controls (20c,d) ( $p < .0001$ ) but higher than ungrammatical controls (20e,f) ( $p < .0001$ ). In both models, there was no interaction effect. This confirms what we found in our corpus studies: there is no penalty for relativizing the complement of the subject noun with *de qui*, contrary to Sportiche and Bellier (1989); Tellier (1990, 1991).

### 6.3 Discussion

Our experiment shows that there is no subject penalty for *de qui* relative clauses, and confirms the acceptability of the *de qui* relative clauses found in Frantext corpora (sections 2 and 4). If we compare it with Abeillé et al. (2018)'s experiment on *dont* reported in section 6.1, we notice that there is no subject advantage with *de qui*, while there was one with *dont*. This may suggest that the putative contrast in grammaticality proposed by Tellier (1991)

<sup>10</sup>We used the lme4 package (?) and the lmerTest package (Kuznetsova et al., 2017) for the statistical language R (R Core Team, 2018) to build linear mixed models. Figures 1 and 2 are made with the package ggplot2 (Wickham, 2016).

(extraction out of the subject is grammatical with *dont* and ungrammatical with *de qui*) should be revised as a mere preference: extraction out of the subject is preferred (over extraction out of object) with *dont* and not with *de qui* in relative clauses, as far as acceptability judgements are concerned. However, this difference might be due to the experimental materials, and to the animacy match or mismatch between subject and object. In the *dont* experiment, the baseline (coordination) also displays a subject advantage, while in our *de qui* experiment, there is no subject advantage in the coordination baseline. This is clear in the two interaction graphs below (Figure 2): the two lines are exactly parallel in both experiments, showing no interaction between subject/object condition and extraction, bars indicating 95% confidence intervals.

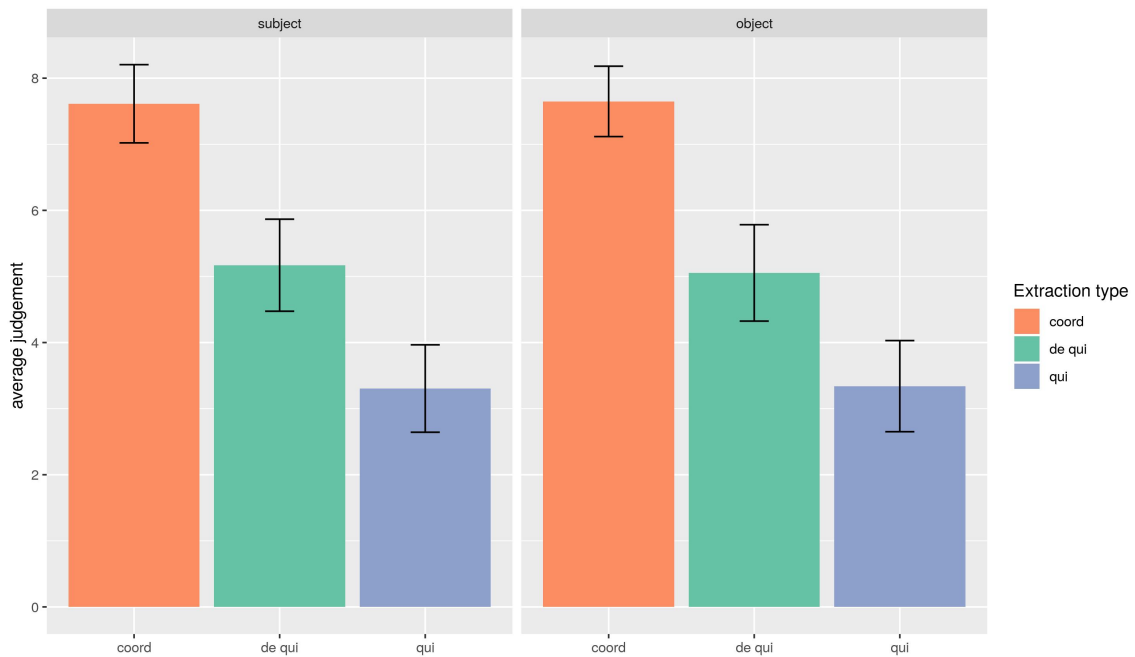


Figure 1: Mean of acceptability judgements for the *de qui* experiment

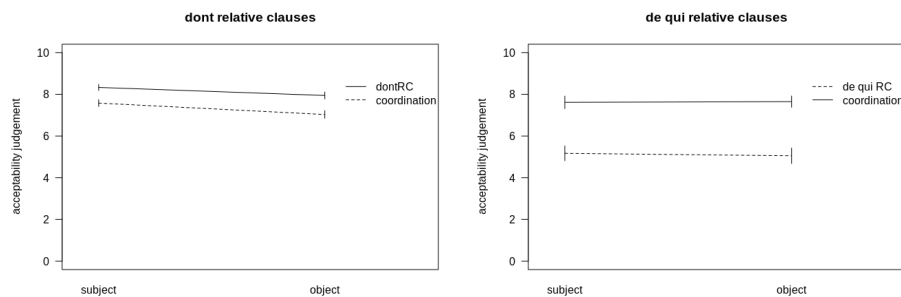


Figure 2: Interaction plots comparing the *dont* and the *de qui* experiments

Another difference between the two experiments is that the *de qui* relatives are below the grammatical coordination baseline, while the *dont* relative clauses were above the coordination baseline (Figure 2). This again may be due to the difference in materials between the two



experiments: the relevant NP in the coordination condition had a possessive determiner in the *dont* experiment (the sentence may therefore be ambiguous); in our *de qui* experiment, we repeated the complement of the noun in the coordination version in order to avoid this possible ambiguity (as can be seen in (20c,d)). The lower acceptability of *de qui*, compared to *dont*, can also be explained by its lower frequency: we had more than 13 000 *dont* relatives vs. 319 *de qui* relatives (including free *qui* relatives) in Frantext 2000-2013; and less than 200 *de qui* relatives (again including free *qui* relatives) in Frantext 1900-1913. The relationship between frequency and acceptability is not a simple issue. As pointed out by Lau et al. (2017) there is no correlation if one considers individual sentences, the frequency of which varies according to lexical frequency and length. Furthermore, unseen sentences (with zero frequency) can be fully acceptable (Featherston, 2005). However, if one considers the frequency of certain constructions, regardless of lexical frequency and sentence length, as we have done in this paper, this frequency shows some correlation with acceptability judgements (Keller, 2000; Lau et al., 2017). The overall difference in frequency between *de qui* and *dont* relative clauses, may explain the difference between our results and those of Abeillé et al. (2020), over all and not specifically in the subject condition. It may also explain the contrasts in judgements reported by Tellier (1990, 1991) between *dont* and *de qui*, if rarer constructions tend to receive lower acceptability judgements.

We do not know why *dont* is much more frequent than *de qui* overall. We hypothesize it may be because *dont* is a complementizer (Godard, 1988; Tellier, 1991), hence simpler to process than a *de qui* PP (Kluender and Kutas, 1993). It may also be because *dont* is a weak pronoun (Sportiche, 2011) and (*de*) *qui* a strong pronoun, given a general preference for weak over strong forms.

## 7 General discussion

Our corpus studies (Frantext 1900-1913 and 2000-2013) did not confirm the syntactic constraints proposed on subject extraction in French by Sportiche and Bellier (1989); Tellier (1990, 1991); Heck (2009). Relativizing out of a nominal subject is possible and frequent, both with *de qui* and with *dont*. In both periods, there are more *de qui* relativizations out of the NP subject than out of the NP object (which is considered acceptable in the generative literature), and these relativizations out of the subject also outnumber the relativization of verbal complements. If one excepts the pied-piping cases, the distribution of *de qui* is similar to the distribution of *dont*, for which relativizing out of the NP subject is the most common usage. Both types of relativization out of the subject are not an innovation in French, even if the proportion of subjects with *de qui* extraction has risen, and they are not limited to unaccusative verbs, contra Chomsky (2008); Uriagereka (2012). On the other hand, attested examples of relativization out of postverbal subjects make Heck’s proposal that *dont* belongs to the subject phrase, implausible.

This does not necessarily mean that extraction out of the subject is not constrained in French. Relative clauses are an extraction type, but there are others to be tested. *De qui* can

indeed also be used in wh-questions, and (Tellier, 1990, 1991) also suggested that questioning out of the subject is not grammatical with *de qui*.

- (21) ?\* De qui est-ce que **la secrétaire** t' a téléphoné? (Tellier, 1990, 307)  
 'Of who did the secretary call you?'

We thus compare *de qui* relatives with *de qui* interrogatives in the same corpus, for the same time periods.

## 7.1 A comparison with *de qui* interrogatives

We found 128 *de qui* questions in Frantext 2000-2013 (Table 2). After taking out 43 verbless questions, 86 *de qui* questions are left, 41 direct and 45 subordinate. We annotate them with the syntactic function of *de qui* and the function of the noun when *de qui* is a complement of noun. Their distribution is strikingly different from that of relative clauses (compare Table 14 with Table 3). We do not find any question out of the subject, while we find questions for the complement of the verb (22a) or of the object (22b).

<i>De qui</i> is a complement of	Nb	%
Verb	53	61.63
Noun		
- Direct Object	2	2.33
- Indirect Object	3	3.49
- Predicate	12	13.95
- Cplt of prep	8	9.30
Adjective	2	2.33
Preposition	1	1.16
others	5	5.81
Total	85	100

Table 14: *De qui* questions in Frantext (2000-2013)

- (22) a. Et d'abord, de qui **tenez**-vous qu'ils sont partis ensemble, là-bas ? (Dans la main du diable, Garat, 2006)  
 'First of all, of who know you that they are gone together there'
- b. Dans la première oeuvre de cette exposition, [...] de qui as-tu utilisé **les voix** ? (La vie possible de Christian Boltanski, Boltanski and Grenier, 2007)  
 'In the first artwork of this exhibition, of who did you use the voices?'

In Frantext 1900-1913, we found 70 *de qui* questions (Table 7). After taking out 17 verbless questions, we are left with 53 *de qui* questions, 34 direct and 19 subordinate. As for 2000-2013, we observe a striking difference between relative clauses and interrogatives (compare Table 15 with Table 8): we do not find any example of questioning out of the subject. This suggests a contrast between relative clauses and wh-questions, even though for 1900-1913, we did not find any questioning out of the object either. Interrogatives with *de qui* strongly favor the complement of the verb (23a): this case represents 71.7% of the *de qui* verbal questions in 1900-1913. We found 7 examples of questioning out the predicative noun (23b) but no example

of questioning out of the object. This specialization does not evolve across time: the difference between the two periods in this respect is not significant (a Fisher’s Test fails with  $p = 0.2704$ ).

<i>De qui</i> is a complement of	Nb	%
Verb	38	71.70
Noun		
- Predicate	7	13.21
- Cplt of prep	7	13.21
Adjective	1	1.89
Total	53	

Table 15: *De qui* questions in Frantext 1900-1913

- (23) a. [...] l’interlocuteur saura bien de qui vous voulez **parler** [...] (À la recherche du temps perdu : Du côté de chez Swann, Proust, 1913)  
‘the interlocutor will know well who you mean’ (lit. ‘of who you want to talk’)
- b. De qui était-il **la proie** ? (Jean-Christophe : L’Adolescent, Rolland, 1905)  
‘Of who was he the prey?’

While it is true that constructions with zero frequency can have various degrees of acceptability (Featherston, 2005), we can say that our corpus data do not contradict the subject penalty suggested by (Tellier, 1990, 1991) for *de qui* questions. Our corpus studies show a striking difference between *de qui* relative clauses on one hand, and interrogatives on the other hand. Relativizing the complement of the subject is frequent, even favored in our corpus, questioning is not.

## 7.2 Are relative clauses special?

Does French obey a syntactic subject island constraint ? Our corpus studies have provided numerous examples of extraction out of the subject with *de qui* relative clauses, and an acceptability judgment task has found no difficulty with relativizing the complement of the subject with *de qui*. However, we did not find any example of extraction out of the subject in wh-questions. The subject island constraint can be viewed as a specific syntactic constraint, or as the result of cumulative independent constraints : Haegeman et al. (2014) consider several non syntactic factors that may ameliorate extraction out of the subject, for example indefiniteness. (Bianchi and Chesi, 2014) also suggest that subjects ofthetic sentences do not occupy the same position as those of categorical sentences, and are easier to extract from. While these proposals may explain why some examples are better than others, they do not address the basic contrast we have found between relative clauses and wh-questions. Why are relative clauses special ?

We first consider the hypothesis that relativization does not involve extraction, as suggested by an anonymous reviewer. The ease of relativizing may mean that such relative clauses do not involve extraction, contrary to wh-questions. It is true that there are cases of gapless *dont* relative clauses with a resumptive pronoun in French (Godard, 1988; Tellier, 1991), but not with *de qui*.

In our corpora, we found no *de qui* relatives with a resumptive pronoun. For 2000-2013, we found 5 such *dont* relative clauses (24), and 3 for 1900-1913. In this case, as noticed by Godard (1988), the resumptive pronoun must be embedded (*\*celui dont<sub>i</sub> il<sub>i</sub> l'accompagne* 'the one of which he accompanies him'), contrary to our cases with extraction out of the subject.

- (24) [...] celui dont<sub>i</sub> elle voit bien qu'il<sub>i</sub> l'accompagne [...] (Devant ma mère, Pachet, 2007)  
 'the one of which she sees well that he accompanies her'

Several authors have proposed a gapless analysis, such that the relative PP would be a hanging topic (Giorgi and Longobardi, 1991; Broekhuis, 2006). This analysis would predict that an anaphoric pronoun or a possessive should be possible. Going back to our corpus examples, it is clear that such a possessive is not grammatical with *dont* (ex. 12) nor with *de qui* (ex.11a), in standard French:<sup>11</sup>

- (26) a. la belle Antillaise, dont l'effigie orna les boîtes de Banania.  
 'the nice Caribbean girl, of which the picture decorated the packages of the Banania brand'
- b. \*la belle Antillaise, dont<sub>i</sub> son<sub>i</sub> effigie orna les boites de Banania  
 'the nice Caribbean girl, of which her picture decorated the packages of the Banania brand'
- c. les ogres de qui **la danse barbare** vous confisque l' enfance  
 'the ogres of whom the barbaric dance takes your childhood away from you'
- d. \*les ogres de qui<sub>i</sub> leur<sub>i</sub> danse barbare vous confisque l'enfance  
 'the ogres of whom their barbaric dance takes your childhood away from you'

Another problem for the gapless hypothesis is that it does not account for connectivity effects: *dont* relativizes a *de*-PP, while *que* relativizes an NP (in standard French). In the experiment reported by Abeillé et al. (accepted) (see section 6.1), *que* was tested as a variant of *dont* and was judged unacceptable, both for the complement of the subject and the complement of the object; so it is difficult to claim that such *dont* relative clauses are gapless. In our experiment (see section 6.2), *qui* was tested as a variant of *de qui* and was judged unacceptable, in both subject and object conditions, which provides evidence for a PP[de] gap (see Haegeman et al. (2014) for more arguments on the same line). We thus conclude that our empirical studies show that extraction out of the subject is possible with *de qui* and *dont* in relative clauses, but not attested with *de qui* in *wh*-questions.

A similar difference among constructions has been observed experimentally by Sprouse et al. (2016) for Italian: in two acceptability judgement tasks, extraction out of subjects was rated higher than extraction out of objects in relative clauses, and lower in *wh*-questions. They

<sup>11</sup>In our experimental items (20a, 20b), such a possessive is not possible either:

- (25) a. J' ai un voisin de qui **la compagne** connaît ma cousine.  
 'I have a neighbour of whom the partner knows my cousin'
- b. \*J'ai un voisin de qui sa compagne connaît ma cousine.  
 'I have a neighbour of whom his partner knows my cousin'

conclude that ‘this [cross construction] variation requires modification to all existing syntactic theories of island effects’.

### 7.3 A discourse-based hypothesis

The lack of extraction out of subjects in *wh*-question but its high frequency (and high acceptability) in relative clauses may be difficult to explain under purely syntactic accounts (Chomsky et al., 1977; Chomsky, 2008). Other approaches to the subject island constraint have proposed that processing (Kluender, 2004) and discourse factors (Erteschik-Shir, 2007) play a role. It may thus reflect a non-syntactic difficulty (Chaves, 2013). Erteschik-Shir (2007) and Goldberg (2013), among others, have proposed that the difficulty of extraction does not come from syntactic configurations but from discourse infelicity. Since extraction makes a constituent more salient (it becomes a Topic or a Focus), they claim that extraction is infelicitous out of backgrounded or non-focal constituents. Since the subject is usually a Topic, extraction out of the subject is not ungrammatical but infelicitous.

Abeillé et al. (2020) have proposed to revise such an explanation in order to account for the difference between constructions, and suggest that it is not felicitous to focus part of a non focal constituent. The extracted element in a *wh*-question is the Focus of the clause, whereas the subject is prototypically a Topic. Focusing (with a *wh*-question) a subpart of a (Topic) subject is thus dispreferred. This type of approach predicts that relative clauses are not constrained in this respect: the relativized element is not focused and the antecedent of the relative clause can have any discourse status in the main clause. Such a theory only constraints focalizing extraction, such as *wh*-question, but not relativizing. It does not predict that it is more frequent than relativizing out of object, as in our corpus studies. This latter fact may come from non syntactic factors that make subject relatives more frequent than object relatives crosslinguistically (see for example Roland et al. 2007) and from a more general accessibility hierarchy of grammatical functions (Keenan and Comrie, 1977).<sup>12</sup>

If valid, this proposal should apply to other languages as well, making relativizing out of the subject easier than questioning out of the subject (see Sprouse et al. 2016 for such a contrast in Italian). It may also explain that some *wh*-questions are better than others, if a subject can be made more focal in a given context.<sup>13</sup>

## 8 Conclusion

Theoretical discussions of long distance dependencies have claimed that it is more difficult to extract out of subjects than out of objects (Ross, 1967; Chomsky, 1986), even though various

---

<sup>12</sup>Psycholinguistic experiments also show that subject relatives are preferred to object relatives, in French (Pozniak and Hemforth, 2015), and crosslinguistically (Holmes and O’Reagan, 1981).

<sup>13</sup>Of course, other factors may play a role as well in other languages. For example in English, it is well known since Ross (1967) that preposition stranding penalize extraction out of the subject (27a), even if more acceptable examples have been proposed by various authors (Kluender, 1998, 2004; Chaves, 2013):

- (27) a. \* Which cars were **the hoods of** damaged by the explosion? (Ross, 1967, fn. 31)  
b. What were **pictures of** seen around the globe? (Kluender, 1998, 268)  
c. Which problem will **a solution to** never be found? (Chaves, 2013, 301)

factors may play a role (Haegeman et al., 2014). French *dont* has been claimed to be an exception in this respect (Sportiche and Bellier, 1989; Tellier, 1990, 1991) and *de qui* to follow the general rule. In a large corpus of literary texts (Frantext 2000-2013), we have found that relativizing out of the subject is possible with *de qui* as well as *dont*. Furthermore, our corpus analysis shows that both for *dont* and *de qui*, relativization of the complement of the subject noun is the most frequent use of these relative clauses (pied-piping excepted for *de qui*). It is possible with intransitive as well as transitive verbs, with preverbal subjects as well as postverbal subjects. We compared with Frantext 1900-1913 and could not identify any meaningful evolution between the two periods. We ran an acceptability judgement task that confirmed that relativizing the complement of the subject with *de qui* is as acceptable as relativizing the complement of the object.

The syntactic difference between the two types of relativizers (a complementizer for *dont* vs. a relative pronoun for *qui*, or a weak vs. a strong pronoun following Sportiche 2011) does not play a role. We thus conclude that French behaves like Italian in this respect (Rizzi, 1982).

However, looking at *de qui* interrogatives, we did not find any extraction out of subject. We conclude that extraction out of the subject is sensitive to the construction type: it is possible in relative clauses and more difficult in interrogatives. This difference among constructions has also been observed by Sprouse et al. (2016) for Italian, and found to be difficult to explain by most current syntactic theories. We suggest it is compatible with theories which view the subject island as a matter of syntax-discourse interface (Erteschik-Shir 2007; Abeillé et al. 2020).

On a more general note, Sprouse et al. (2013) argued that most intuitions reported on English in a major linguistic journal were confirmed by acceptability judgements. However Linzen and Oseki (2018) argue that it is not the case when testing data from other languages (Hebrew and Japanese). This paper suggests, using both corpus data and acceptability judgements, that it might not be the case for French either.

## References

- Abeillé, A., Clément, L., and Liégeois, L. (2019). Un corpus annoté pour le français : le French Treebank. *TAL Traitement Automatique des Langues*, 60(2):19–43.
- Abeillé, A., Hemforth, B., and Winckel, E. (2016). Les relatives en *dont* du français: études empiriques. In Neveu, F., Bergounioux, G., Côté, M.-H., Fournier, J.-M., Hriba, L., and Prévost, S., editors, *5e Congrès Mondial de Linguistique Française*, volume 27 of *SHS Web of Conferences*.
- Abeillé, A., Hemforth, B., Winckel, E., and Gibson, E. (2018). A construction-conflict explanation of the subject-island constraint. Poster at the 31st Annual CUNY Sentence Processing Conference in UC Davis.
- Abeillé, A., Hemforth, B., Winckel, E., and Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition*.
- ATILF, CNRS, and de Lorraine, U. (Version décembre 2016). Base textuelle FRANTEXT.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):309–412.

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.
- Bianchi, V. and Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry*, 45(4):525–569.
- Branca-Rosoff, S., Fleury, S., Lefeuvre, F., and Pires, M. (2012). *Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*.
- Broekhuis, H. (2006). Extraction from subjects: some remarks on Chomsky’s “On phases”. In Broekhuis, H., Corver, N., Huybregts, R., Kleinhenz, U., and Koster, J., editors, *Organizing Grammar*, pages 59–68. De Gruyter, Berlin, Boston.
- Chaves, R. (2013). An expectation-based account of subject islands and parasitism. *Journal of Linguistics*, 49(2):297–344.
- Chomsky, N. (1973). Conditions on transformations. In Anderson, S. and Kiparsky, P., editors, *Festschrift for Morris Halle*, pages 232–285, New York. Winston.
- Chomsky, N. (1986). *Barriers*, volume 13 of *Linguistic inquiry monographs*. MIT Press, Cambridge, MA.
- Chomsky, N. (2008). On phases. In Freidin, R., Michaels, D., Otero, C. P., and Zubizarreta, M. L., editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pages 133–165, Cambridge, MA. MIT Press.
- Chomsky, N., Wasow, T., and Akmajian, A. (1977). On Wh-movement. In Culicover, P. W., editor, *Formal syntax*, pages 71–132. Academic Press, San Francisco, London.
- Drummond, A. (2010). Internet Based EXperiments (IBEX).
- Erteschik-Shir, N. (2007). *Information Structure: The Syntax-Discourse Interface*. Oxford, Oxford University Press edition.
- Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh- constraints in German. *Lingua*, 115:1525–1550.
- Gibson, E. and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28:88–124.
- Giorgi, A. and Longobardi, G. (1991). *The syntax of noun phrases: Configuration, parameters, and empty categories*, volume 57 of *Cambridge studies in linguistics*. Cambridge University Press, Cambridge, UK, and New York.
- Godard, D. (1988). *La syntaxe des relatives en français*. Ed. du Centre national de la Recherche Scientifique, Paris.
- Godard, D. and Sag, I. (1996). Quels compléments de nom peut-on extraire en français? *Langages*, 122:60–79.
- Goldberg, A. (2013). Backgrounded constituents cannot be extracted. In Sprouse, J. and Hornstein, N., editors, *Experimental Syntax and Island Effects*, pages 221–238. Cambridge University Press, Cambridge, UK.
- Haegeman, L., Jiménez-Fernández, L., and Radford, A. (2014). Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review*, 31:73–150.
- Heck, F. (2009). On certain properties of pied-piping. *Linguistic Inquiry*, 40(1):75–111.
- Hemforth, B. (2013). Experimental linguistics. In Mark Aronoff, editor, *Oxford Bibliographies*. Oxford University Press, New York.
- Holmes, V. M. and O’Reagan, J. K. (1981). Eye Fixation Patterns during the Reading of Relative-Clause Sentences. *Journal of Verbal Learning and Verbal Behaviour*, 20:417–430.
- Keenan, E. L. and Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 8(1):63–99.
- Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. PhD thesis, University of Edinburgh.

- Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In Culicover, P. and McNally, L., editors, *The Limits of Syntax*, pages 241–279. Academic Press, New York.
- Kluender, R. (2004). Are subject islands subject to a processing account? In Schmeiser, B., Chand, V., Kelleher, A., and Rodriguez, A., editors, *Proceedings of the WCCFL*, volume 23, pages 101–125, Somerville, MA. Cascadilla Press.
- Kluender, R. and Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8:573–633.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.
- Lau, J., Clark, A., and Lappin, S. (2017). Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, 5(41):1–40.
- Le Goffic, P. (2007). Les mots qu- entre interrogation, indéfinition et subordination : quelques repères. *Lexique*, 18:13–46.
- Legate, J. A. (2003). Some Interface Properties of the Phase. *Linguistic Inquiry*, 34(3):506–516.
- Linzen, T. and Oseki, Y. (2018). The reliability of acceptability judgments across languages. *Glossa*, 3(1):100.
- Polinsky, M., Gallo, C., Graff, P., Kravtchenko, E., Morgan, A., and Sturgeon, A. (2013). Chapter 13. Subject islands are different. In Sprouse, J. and Hornstein, N., editors, *Experimental Syntax and Island Effects*, pages 286–309. Cambridge University Press.
- Pozniak, C. and Hemforth, B. (2015). Processing subject and object relative clauses in French and Mandarin Chinese. Poster at the 28th Annual CUNY Sentence Processing Conference in Los Angeles.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rizzi, L. (1982). *Issues in Italian syntax*, volume v. 11 of *Studies in generative grammar*. Foris Publications, Dordrecht, Holland and Cinnaminson, N.J., U.S.A.
- Roland, D., Dick, F., and Elman, J. (2007). Frequency of basic English grammatical structures. *Journal of memory and language*, 57(3):348–379.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Ph.d thesis, MIT, Cambridge, MA.
- Sag, I. A. and Godard, D. (1994). Extraction of *de*-Phrases from the French NP. In *Proceedings of the North East Linguistic Society 24*, pages 519–541, Amherst, MA. GLSA.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. D. Reidel, Dordrecht.
- Sportiche, D. (1981). Bounding nodes in French. *The Linguistic Review*, 1(2):219–246.
- Sportiche, D. (2011). French Relative Qui. *Linguistic Inquiry*, 42(1):83–124.
- Sportiche, D. and Bellier, P. (1989). Le mouvement syntaxique : contraintes et paramètres. In *Langages*, 24 année, n°95, 1989. *La géométrie du langage. Etat présent de la grammaire générative, sous la direction de Patrick Bellier*. Armand Colin, Paris.
- Sprouse, J. and Almeida, D. (2017). Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences*, 40:e311.
- Sprouse, J., Caponigro, I., Greco, C., and Cecchetto, C. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory*, 34(1):307–344.
- Sprouse, J., Schütze, C. T., and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010. *Lingua*, 134:219–248.
- Stepanov, A. (2007). The end of CED? Minimalism and extraction domains. *Syntax*, 10(1):80–126.
- Tellier, C. (1990). Subjacency and Subject Condition violations in French. *Linguistic Inquiry*, 21(2):306–311.
- Tellier, C. (1991). *Licensing theory and French parasitic gaps*, volume v. 26 of *Studies in natural lan-*



- guage and linguistic theory*. Kluwer Academic Publishers, Dordrecht, The Netherlands and Boston.
- Uriagereka, J. (2012). *Spell-Out and the Minimalist Program*. Oxford Scholarship Online, Oxford.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.