Extraction out of subject phrases in French: experimental evidence

Elodie Winckel
Humboldt-Universität zu Berlin
Inst. für dt Sprache und Linguistik
10117 Berlin, Germany
&
Université de Paris, LLF, CNRS
F-75013 Paris, France
`elodie.winckel@hu-berlin.de`

Anne Abeillé
Université de Paris, LLF, CNRS
F-75013 Paris, France
`anne.abeille@`
`linguist.univ-paris-diderot.fr`

Abstract

Extracting the complement outside a subject has been claimed to be more difficult than out of the object (Chomsky, 1973, a.o). In a series of controlled experiments, we compare extraction out of subject and out of object in French, and we also compare relative clauses and wh-questions. As it turns out, relativizing out of the subject is easier than out of the object, while the opposite pattern holds for wh-questions. We claim that our results cast doubt on a universal syntactic constraint on this type of operation, and call for a discourse based explanation, since the discourse function of relative clauses is different from that of wh-questions.

1 Introduction

Processing (Kluender and Kutas, 1993) and discourse-based (Goldberg, 2013) approaches to islands pose a challenge to syntax-based approaches to the problem. We focus here on the "subject island constraint" (Ross, 1967), which is supposed to ban extraction out of the subject. In Ross (1967)'s original account, only extraction out of sentential subject in English is concerned by the constraint, but followers (Chomsky, 1973; Chomsky, 1986; Chomsky, 2008; Huang, 1982; Boeckx, 2012) extended the constraint to nominal subjects as well. The constraint is also claimed to be universal, despite Rizzi (1982), Godard (1988) and Stepanov (2007). Experimental data is still rare (Kravtchenko et al., 2009; Polinsky et al., 2013; Sprouse et al., 2016; Abeillé et al., 2018).

In a series of experiments, we compare relative clauses and interrogatives in French. Our results show that extraction out of subjects in relative clauses is felicitous (and not restricted to *dont* relative clause, as claimed by Tellier, 1991). They also show that wh-questions differ from relative clauses in this respect, and we provide an explanation which takes into account the different discourse function of the constructions.

All our experiments are controlled online acceptability judgement studies (Gibson and Fedorenko, 2013), in which participants have to rate sentences on a scale from 1 (bad) to 10 (good), with experimental items mixed with other items serving as distractors. The items are randomized and each participant sees each experimental item in only one condition. Participants also have to answer a comprehension question after each item. Results of participant with a low score in comprehension questions are excluded for the statistical analysis. In the following, we report only the significant results (p > .05) of linear mixed-effect models (Bates, 2010). We use the R.I.S.C. website (http://experiences.risc.cnrs.fr/) and social media to gather volunteers for these experiments.

2 Experiment 1: *dont* Relative Clauses

We investigate relative clauses introduced by *dont*, which have been reported as felicitous for relativizing the complement of a subject (Godard, 1988).

(1) C'est un philosophe dont le portrait se trouve au Louvre.

'It's a philosopher of who the portrait stands in the Louvre.'

We compare relativizing the complement of a subject noun with that of an object noun, using closely related verbs (enchanter 'delight'/ aimer 'like'). We compare relative clauses (extraction) with clausal

coordination (no extraction) and ungrammatical control conditions (*que* instead of *dont*), both for subject and object (2x3 design). We have 24 target items and 24 distractors. It is an acceptability judgement study conducted on internet (Ibex) with 48 participants.

Material for Experiment 1

(2) a. subj, PP-ext: Le concessionnaire a une décapotable dont [la couleur _] enchante le footballeur à cause de sa luminosité.

‘The dealer has a sportscar of which [the color _] delights the football player because of its luminance’

b. obj, PP-ext: Le concessionnaire a une décapotable dont le footballeur adore [la couleur _] à cause de sa luminosité.

‘The dealer has a sportscar of which the football player loves [the color _] because of its luminance.’

c. subj, no-ext: Le concessionnaire a une décapotable et sa couleur enchante le footballeur à cause de sa luminosité.

‘The dealer has a sportscar and its color delights the football player because of its luminance’

d. obj, no-ext: Le concessionnaire a une décapotable et le footballeur adore sa couleur à cause de sa luminosité.

‘The dealer has a sportscar and the football player loves its color because of its luminance.’

e. subj, ungram: Le concessionnaire a une décapotable que la couleur enchante le footballeur à cause de sa luminosité.

‘The dealer has a sportscar which [the color -] delights the football player because of its luminance’

f. obj, ungram: Le concessionnaire a une décapotable que le footballeur adore la couleur à cause de sa luminosité.

‘The dealer has a sportscar which the football player loves the color because of its luminance.’

On average, relativization from subject position (2a) is rated higher than relativization from object position (2b). There is however no interaction effect, because clausal coordination (no-extraction) is rated higher in the subject condition (2c) than in the object condition (2d). Ungrammatical controls (2e,f) are rated much lower than the other ones (see Figure 1 with z-scores).

These results confirm the corpus findings of Abeillé et al. (2016) in written and spoken French: *dont* relative clauses are more frequent for relativizing the complement of the subject than of the object. This confirms the theoretical proposals of Godard (1988), Sportiche (1981) : there is no difficulty to extract the complement of a subject in *dont* relative clauses in French.

Figure 1: Results of E1

3   Experiment 2: *de qui* Relative Clauses

Tellier (1991) claimed that *dont* relative clauses are an exception in French, and that extraction out of the subject with a relative pronoun like *qui* (‘who’) is ruled out.
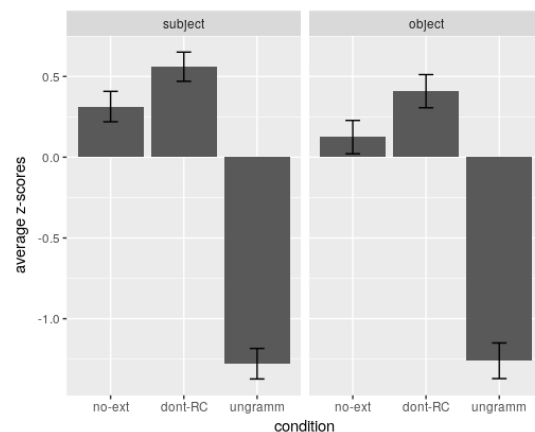
(3)   a. C'est un linguiste dont/*de qui [les parents _] ont déménagé à Chartres. (Tellier, 1991)

       'It is a linguist of whom the parents moved to Chartres'

     b. C'est un linguiste dont/ de qui vous avez rencontré [les parents _]. (Tellier, 1991)

       'It is a linguist of whom you have met the parents'

In a generative perspective, Tellier (1991) proposed that *dont* as a complementizer does not involve syntactic movement, contrary to true relative pronouns like *duquel* or *de qui*. Heck (2009), in the same theoretical school, proposed that *dont* is a specifier which may belong to the subject Noun Phrase, but it is not the case for *de qui*.

We therefore ran a second experiment to investigate relative clauses introduced by *de qui*. We compare extracted variants with clausal coordination (no extraction) and ungrammatical control (missing preposition) conditions, both for subject and object (2x3 design). We have 24 target items and 24 distractors. It is an acceptability judgement study conducted on internet (Ibex) with 28 participants. We choose animate human nouns for both subjects and objects, to avoid animacy mismatch, and reversible transitive verbs (aimer 'like', connaître 'know').

Material for Experiment 2

(4)   a. subj, PP-ext: J'ai un voisin de qui [la compagne _] connaît ma cousine.

       'I have a neighbour of whom [the partner -] knows my cousin'

     b. obj, PP-ext: J'ai un voisin de qui ma cousine connaît [la compagne _].

       'I have a neighbour of whom my cousin knows [the partner _].'

     c. subj, no-ext: 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. obj, no-ext: 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. subj, ungram: J'ai un voisin qui [la compagne _] connait ma cousine.

       'I have a neighbour who the partner knows my cousin'

     f. obj, ungram: J'ai un voisin qui ma cousine connait [la compagne _].

       'I have a neighbour who my cousin knows the partner'

We obtained no significant difference between extraction out of subject (4a) and out of object (4b), and both were rated higher than ungrammatical controls (4c,d). We conclude that Tellier (1991)'s contrast in grammaticality should be revised as a preference difference: extraction out of subject is possible in both *dont* and *de qui* relative clauses, and actually preferred with *dont* relative clause.

Figure 2: Results of E2

4   Experiment 3: interrogatives

In a similar kind of experiments, Sprouse et al. (2016) show that extraction out of NP subject in Italian is felicitous in relative clauses and degraded in wh-questions but they do not explain this difference. Abeillé et al. (2018) find a similar contrast in English between relative clauses (with pied-piping) and wh-questions. This is why we test wh-questions as well.
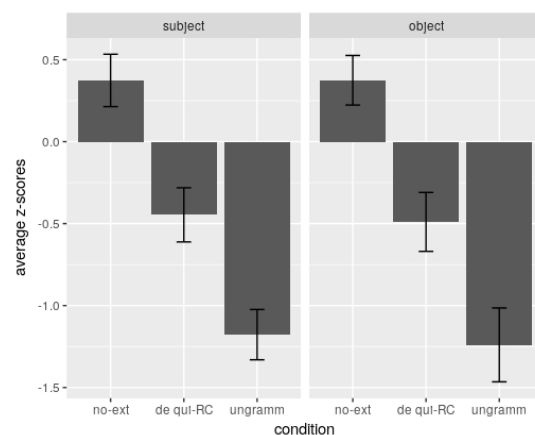
In Experiment 3, we compare questioning the complement of a subject and of the object. We avoid subject-verb inversion and use [*de quel(le)* + N] (litt. 'of which N') as extracted element. We compare wh-questions (extraction condition) with polar questions (no extraction) and ungrammatical control (missing preposition) conditions, both for subject and object (2x3 design). We have 24 target items and 32 distractors. It is an acceptability judgement study conducted on internet (Ibex) with 47 participants.

Material for Experiment 3

(5) a. subj, PP-ext: De quelle décapotable est-ce que [la couleur _] enchante le footballeur à cause de sa luminosité ?

'Of which sportscar does [the color _] delight the football player because of its luminance?'

b. obj, PP-ext: De quelle décapotable est-ce que le footballeur adore [la couleur _] à cause de sa luminosité ?

'Of which sportscar does the football player love [the color _] because of its luminance?'

c. subj, no-ext: Est-ce que la couleur de la décapotable enchante le footballer à cause de sa lumisosité?

'Does the color of the sportscar delight the football player because of its luminance?'

d. obj, no-ext: Est-ce que le footballeur adore la couleur de la décapotable à cause de sa lumisosité?

'Does the football player love the color of the sportscar because of its luminance?'

e. subj, ungram: Quelle décapotable est-ce que la couleur enchante le footballeur à cause de sa luminosité?

'Which sportscar does the color delight the football player because of its luminance'

f. obj, ungram: Quelle décapotable est-ce que le footballer adore la couleur à cause de sa luminosité ?

'Which sportscar does the football player love the color because of its surprising luminance?

We find that, on average, extraction out of subjects (5a) was rated lower than out of objects (5b), and also lower than polar (non-extraction) questions (5c,d), but better than ungrammatical controls (5e,f). On the other hand, polar questions are rated higher in subject than in object condition: there is an interaction between extraction and subject/object condition (see Figure 3 with z-score).

Discussion Contrary to relative clauses, which show no subject penalty both with *dont* and *de qui*, we find a subject penalty with wh-questions. This is in line with corpus data: Abeillé and Winckel (2018) find no examples of questioning the complement of the subject in Frantext (2000-2010).

Figure 3: Results of E3

These results are difficult to explain under syntactic theories which analyse relativization and wh-question via the same movement operation. Why would movement be easy in relative clause and difficult in wh-question if the syntactic structures are similar? They are easier to explain under discourse based theories which view locality constraints as coming from a discourse status clash (Erteschick-Shir, 2007; Goldberg, 2013): under discourse based theories, extraction makes an element more salient and is more felicitous for elements belonging to the "focal domain" of the sentence, i.e. not backgrounded elements (like sentential subjects) nor topical elements (like nominal subjects). Following Abeillé et al. (2018),
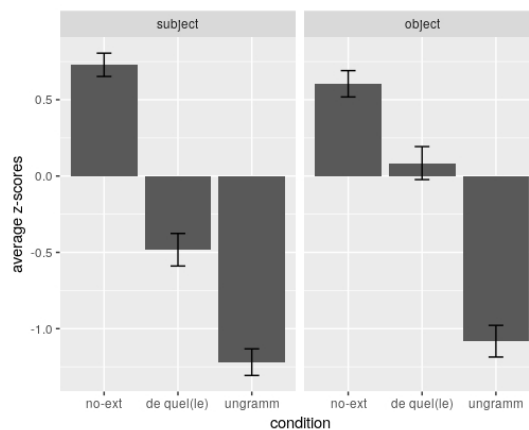
we propose that syntactic theories should be revised in order to take into account the discourse function of the construction. Indeed relativization is a topicalization (Kuno, 1976) whereas wh-extraction in direct interrogatives is a focalization (Jackendoff 1972). We therefore expect different results for both constructions: focusing (in a wh-question) an element which belongs to the focal domain (a complement of the object) is easier than focusing an element which belongs to the sentence topic (a complement of the subject). Of course, such discourse status are flexible, and some subjects may be more focal, or less topical, and thus make it easier for their complement to be questioned, as in the following attested[1] example:

(6) De qui l'anniversaire tombe-t-il le 27 ?

   'Of whom does the birthday take place on the 27th?'

On the other hand, relative clauses are subordinate clauses which add a property to an entity (the antecedent) which can have various discourse status in the main clause (in our experiments, the antecedent noun is a complement). They do not put the relativized element into focus. There is thus no discourse clash and no subject penalty.

5   Experiment 4: *où*-extraction out of sentential subjects

We have shown that no subject penalty arises with relative clauses, and proposed that the difficulty found with wh-questions come from the discourse status of the construction (questioning means putting the questioned element into focus). This proposal is supported by corpus studies (Abeillé and Winckel, 2018) and experiments on nominal subjects of transitive verbs. What about other subjects? While sentential subjects (7a) are supposed to be (Miller, 2001; Goldberg, 2013), and rare (only 24 in the French Treebank), compared to impersonal variants (7b), infinitival subjects have seldom been discussed (Chaves and Dery, 2018).

(7) a. ? Que Paul soit parti m'étonne.

      'That Paul left surprises me.'

   b.   Ca m'étonne que Paul soit parti.

      'It surprises me that Paul left.

Infinitival subjects are also rarer than nominal ones (respectively 99 and 26 000 in the French Treebank; Abeillé et al., 2019) and an impersonal variant is possible to:

(8) a. ? Partir demain m'ennuie.

      'To leave tomorrow bothers me.'

   b.   Ca m'ennuie de partir demain.

      'It bothers me to leave tomorrow.

In Experiment 4, we turn to extraction out of infinitival subjects. We test infinitival subjects with a locative complement and extract this complement using *où* ('where'). We compare extraction out of subjects and out of object, using the impersonal variant. We compare extracted variants with no extraction conditions (coordination), both for subject and object (2x2 design). We have 12 target items and 36 distractors including some ungrammatical ones. It is an acceptability judgement study conducted on internet (Ibex) with 33 participants.

Material for Experiment 4

(9) a. subj, PP-ext: Il y a une guerre civile dans ce pays, où [aller _] est dangereux en ce moment.

      'There is a civil war in this country, where [to go _] is dangerous right now.'

---

[1] https://www.quizz.biz/quizz-1027073.html (Accessed: February 13, 2019)

b. pred, PP-ext: Il y a une guerre civile dans ce pays, où il est dangereux d'[aller _] en ce moment.

'There is a civil war in this country, where it is dangerous [to go _] right now.'

c. subj, no-extr: Il y a une guerre civile dans ce pays, et y aller est dangereux en ce moment.

'There is a civil war in this country, and to go there is dangerous right now.'

d. pred, no-extr: Il y a une guerre civile dans ce pays, et il est dangereux d y aller en ce moment.

'There is a civil war in this country, and it is dangerous to go there right now.'

We find a tendency for extractions out of subjects (9a) to decrease in acceptability compared to the other three conditions (9b,c,d) but no significant difference. Their rating is very high though (average rating of 7.6). They differ significantly from the ungrammatical items (Figure 4 with z-scores). Surprisingly, there is no significant difference between infinitival subject and impersonal constructions, although the latter is more frequent in corpora.

Figure 4: Results of E4

6 Conclusion

Our results show that relativizing out of subject is either rated better or rated in the same way as relativizing out of objects. This is not true for all extraction, because extracting out of a subject in a wh-question shows a clear decline of acceptability compared to extracting out of an object.

We propose that Ambridge and Goldberg (2008)'s claim that subjects are difficult to extract from because they are background information, only applies to questions. We propose that: (a) the discourse status of the extracted element is a focus in wh-question, not in relative clause; (b) the discourse status of the extracted element must match the discourse status of the noun it is extracted from. We thus predict that wh-questions disfavor extraction of an element out of a subject, which is a default clause topic (Lambrecht, 1994), and that relative clauses don't. Ross (1967)'s constraint on extraction out of subject phrases only applies to sentential subjects. Our results show no significant interaction effect for extracting out of infinitival subjects, which received good ratings. We conclude that a syntactic constraint banning extraction out of subjects does not hold in French. French not only allows extraction from nominal subjects, but even prefers such extraction, given the right construction and the right relativizer.
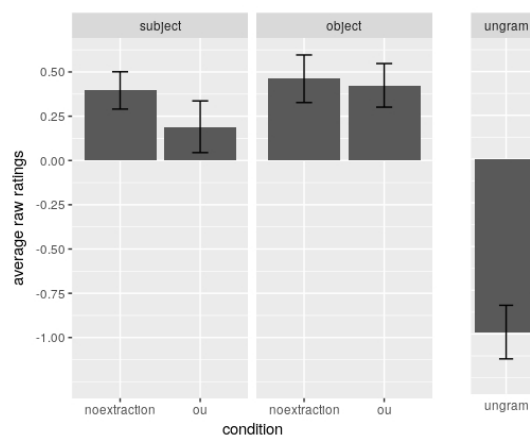
References

Anne Abeillé and Elodie Winckel. 2018. 'Dont' and 'de qui' relatives in written French, November. Talk at the Conference Grammar and Corpora in Université Paris-Diderot.

Anne Abeillé, Barbara Hemforth, and Elodie Winckel. 2016. Les relatives en dont du français: études empiriques. In F. Neveu, G. Bergounioux, M.-H. Côté, J.-M. Fournier, L. Hriba, and S. Prévost, editors, *5e Congrès Mondial de Linguistique Française*, volume 27 of *SHS Web of Conferences*.

Anne Abeillé, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2018. A construction-conflict explanation of the subject-island constraint, March. Poster at the 31st Annual CUNY Sentence Processing Conference in UC Davis.

Anne Abeillé, Lionel Clément, and Loïc Liégeois. 2019. Un corpus annoté pour le français : le French Treebank. *TAL Traitement Automatique des Langues*, 60.

Ben Ambridge and Adele E. Goldberg. 2008. The island status of clausal complements: evidence in favor of an information structure explanation. *Cognitive Linguistics*, 19(3):349–381.

Douglas M. Bates. 2010. *lme4: Mixed-effects modeling with R*. Springer.

Cedric Boeckx. 2012. *Syntactic islands*. Key topics in syntax. Cambridge University Press, Cambridge and New York.

Rui P. Chaves and Jeruen E. Dery. 2018. Frequency effects in subject islands. *Journal of linguistics*, pages 1–47.

Noam Chomsky. 1973. Conditions on transformations. In Steven Anderson and Paul Kiparsky, editors, *A festschrift for Morris Halle*, pages 232–285, New York. Winston.

Noam Chomsky. 1986. *Barriers*, volume 13 of *Linguistic inquiry monographs*. MIT Press, Cambridge, MA.

Noam Chomsky. 2008. On phrases. In Robert Freidin, David Michaels, Carlos P. Otero, and Maria Luisa Zubizarreta, editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pages 133–165, Cambridge, MA. MIT Press.

Nomi Erteschick-Shir. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford, oxford university press edition.

Edward Gibson and Evelina Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28:88–124.

Danièle Godard. 1988. *La syntaxe des relatives en français*. Paris : Ed. du Centre national de la Recherche Scientifique, Paris.

Adele Goldberg. 2013. Backgrounded constituents cannot be extracted. In Jon Sprouse and Norbert Hornstein, editors, *Experimental Syntax and Island Effects*. Cambridge University Press.

Fabian Heck. 2009. On certain properties of pied-piping. *Linguistic Inquiry*, 40(1):75–111.

C.-T. James Huang. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. thesis, MIT.

Robert Kluender and Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8:573–633.

Ekaterina Kravtchenko, Ming Xiang, and Maria Polinsky. 2009. Are all subject islands created equal? Poster at the 22nd Annual CUNY Sentence Processing Conference in UC Davis.

Susumu Kuno. 1976. Subject, theme, and the speaker's empathy - a reexamination of relativization phenomena. In Charles N. Li, editor, *Subject and Topic*, pages 417–444, New York, NY. Academic Press.

Knud Lambrecht. 1994. *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*. Number 71 in Cambridge studies in linguistics. University Press, Cambridge.

Philip Miller. 2001. Discourse constraints on (non)-extraposition from subject in English. *Linguistics*, 39(4):683–701.

Maria Polinsky, Carlos G. Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon. 2013. Chapter 13. Subject islands are different. In Jon Sprouse and Norbert Hornstein, editors, *Experimental Syntax and Island Effects*. Cambridge University Press.

Luigi Rizzi. 1982. *Issues in Italian syntax*, volume v. 11 of *Studies in generative grammar*. Foris Publications, Dordrecht, Holland and Cinnaminson, N.J., U.S.A.

John Robert Ross. 1967. *Constraints on variables in syntax*. Ph.d thesis, MIT, Cambridge, MA.

Dominique Sportiche. 1981. Bounding nodes in French. *The Linguistic Review*, 1(2):219–246.

Jon Sprouse, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory*, 34(1):307–344.

Arthur Stepanov. 2007. The end of CED? Minimalism and extraction domains. *Syntax*, 10(1):80–126.

Christine Tellier. 1991. *Licensing theory and French parasitic gaps*, volume v. 26 of *Studies in natural language and linguistic theory*. Kluwer Academic Publishers, Dordrecht, The Netherlands and Boston.