French subject island constraint?

A discourse-based HPSG account

Elodie Winckel[1,2] and Anne Abeillé[2]
[1] Humboldt-Universität zu Berlin | [2] Université de Paris, LLF, CNRS

We present new experimental results (corpora and experiments) showing that extraction out of subject, compared with extraction out of object, obeys a pragmatic constraint and not a syntactic constraint. We show how such a constraint can be formalized in an HPSG grammar of French which views relative clauses, *wh*-questions and *it*-clefts as different constructions.

Keywords: island constraints, HPSG, French, relative clauses, *it*-clefts, questions

1. Introduction

Locality constraints on extraction constructions ("island constraints") are at the core of most linguistic theories. The main issue is to know whether they are purely syntactic (Ross, 1967; Chomsky, 1995, 1990; Tellier, 1990, 1991; Uriagereka, 1988) or come from an interplay of syntactic, discourse and processing constraints (Erteschik-Shir, 1973; Kluender & Kutas, 1993; Hofmeister & Sag, 2010; Goldberg, 2006). The aim of this paper is two-fold: first we review new empirical evidence showing that while relative clauses allow for extraction out of subjects in French, as in other Romance languages, it is more difficult in *wh*-questions and *it*-clefts; second, Abeillé *et al.* (2020a) suggest a Focus-Background Conflict constraint to handle this cross-construction difference and we propose a HPSG formulation of this constraint.

1.1 The subject island constraint

Since Ross's (1967) seminal work on long distance dependencies, subjects are considered as "strong" islands (Cinque, 1990; Szabolcsi, 2006). Although Ross (1967: 241–255) only constrains sentential subjects, the constraint has been claimed to apply to nominal subjects as well (Chomsky (1973) and following); for

https://doi.org/10.1075/li.00041.win
Lingvisticæ Investigationes 43:1 (2020), pp. 95–128. issn 0378-4169 | e-issn 1569-9927

some authors, it falls under the same constraint as adjuncts (i.e. in Huang's (1982) Condition on Extraction Domains, or Goldberg's (2006) Backgrounded Constituents are Islands constraint) while for others it is a subject specific constraint (Stepanov, 2007). In what follows, we put brackets around the NP a complement of which has been extracted and mark its non-extracted canonical position with an underscore:

(1) a. *Who did John hear [stories about _]?*
b. **Who did [stories about _] terrify John?* (Chomsky, 1973: 249)

The contrast in (1) is seen as indicating a syntactic island. Counterexamples have been found in many languages, especially Romance languages: Italian (Rizzi (1982), see example (2a)); Spanish (Torrego (1985), see example (2b)) and French (Godard (1988), see example (2c)).

(2) a. *questo autore, di cui so che [il primo libro _] è stato*
this author of which know.1SG that the first book AUX been
pubblicato recentemente (Rizzi, 1982: 61, example 30)
published recently
'this author, of which I know that the first book has been published recently'

b. *¿De que autora no sabes [qué traducciones _] han ganado premios*
by what author not know.2SG what translations AUX won awards
internationales? (Chomsky, 1986: 26)[1]
international
'Of which author don't you know which translations have won international awards?'

c. *la jeune femme dont [le portrait_] est à la fondation Barnes*
the young woman of.which the portrait is at the foundation Barnes
'the young lady of which the portrait is at the Barnes foundation' (Godard & Sag, 1996: 63)

If these examples were to be explained by the null subject parameter (as claimed by Uriagereka (2012) following Rizzi (1982)), French – a language without null subjects – should not pattern like the other Romance languages.[2]

Following Tellier (1990, 1991) and Sportiche & Bellier (1989), Heck (2009) argues that apparent extraction out of the subject is limited to *dont* in French, and,

1. This example is attributed to a 1985 manuscript by Esther Torrego.

2. Following Miller & Sag (1997) and Miller & Monachesi (2003), complement clitics and postverbal subject clitics are affixes. French thus does not allow null subjects, except with a suffixed verb: * *Vient ?* 'Is he coming?' *(Paul) vient-il ?* 'Is Paul/he coming?'.

since *dont* is not a pronoun but a complementizer, (2c) is not a case of true extraction. However, Godard provides exemples of long distance dependency with *dont* (3a) and claims that extraction out of subject is not limited to *dont* (3b-c):

(3) a. *J' ai rencontré Paul, dont je crois bien qu' [un ami _] veut*
I AUX met Paul of.which I believe well that a friend wants
venir te voir (Godard, 1988: 109)
come.INF ACC.2.SG see.INF
'I met Paul, of whom I believe that a friend wants to come to see you.'

b. *un homme de qui [la force de travail _] est étonnante*
a man of who the power of labor is astonishing
'a man whose labor power is astonishing' (Godard, 1988: 56)

c. *De qui te semblait-il que [la force de travail _] est étonnante?*
of who DAT.2.SG seemed-3SG that the power of labor is astonishing
'Of who did it seem to you that the labor power is (Godard, 1988: 56)
astonishing?'

Further works on Spanish and Italian have also proposed that some (postverbal) subjects allow for subextraction more easily than others, and propose that subject "island" is the result of cumulative constraints (Jiménez-Fernández, 2009; Bianchi & Chesi, 2014; Haegeman *et al.*, 2014); see below Section 3.

1.2 The subject island in Head-Driven Phrase-Structure Grammar

In Head-Driven Phrase-Structure Grammar (HPSG), Pollard & Sag (1994: 200), following early work in Generalized Phrase-Structure Grammar (GPSG) (Pollard, 1984: 175), have an English-specific Subject Condition, given in (4):

(4) **Subject Condition** [English]:
A lexical head's SUBCAT list may contain a slashed subject only if it also contains another slashed element.

This allows for extraction out of the subject only if the filler is related to another gap in the sentence as in (5b), i.e only in case it is a 'parasitic' gap:

(5) a. **Who did [rivals of _] shoot Castro?*
b. *Who did [rivals of _] shoot _?* (Chaves, 2013: 303)

(6) shows the HPSG description for *shoot* in (5). The first NP subcategorized by the verb is the subject, and, following (4), this NP can only have an non-empty SLASH set if the verb has another slashed element (i.e. the second (empty) NP in (6b)).

(6) a. *shoot* in (5a):
*[SUBCAT ⟨NP [SLASH {NP_i}], NP_j⟩]
b. *shoot* in (5b):
[SUBCAT ⟨NP [SLASH 1], NP_i [SLASH {1 NP_i}⟩]

For the authors, the constraint in (4) is not universal, but restricted to English, with possible variation, since "many [English] speakers" consider extraction out of the subject acceptable (Pollard & Sag, 1994: 183). For instance Sag & Godard (1994) and Godard & Sag (1996) on French do not have any subject condition and freely allow extraction of a PP complement:

(7) *C' est un philosophe dont [un portrait_] se trouve au Louvre.*
it is a philosopher of.which a portrait REFL finds at.DEF.SG Louvre
'It is a philosopher of whom a portrait is in Le Louvre.' (Godard, 1988: 47)

Subsequent work on English extraction in HPSG has abandoned this Subject Condition (Sag, 1997; Bouma *et al.*, 2001; Sag, 2010). More generally, following Kluender (1991), Hofmeister & Sag (2010) and Hofmeister (2011) view many islands constraints as a result of different non-syntactic factors, including processing difficulties. Such factors (e.g. working memory limitations, contradictions to expectations) may add up and cause unacceptability. Moreover, acceptability is gradient, whereas syntactic approaches expect a categorical judgment on islands. Following Kluender (2004), Chaves (2013) considers that the subject penalty on extraction is gradient and due to a combination of pragmatic and processing factors. Indeed, the subject is usually a pronoun; therefore, it is rare, and hence unexpected, to have a complex subject. He cites a series of acceptable examples in English:

(8) a. Which disease will [the cure for _] never be discovered? (Chaves, 2013: 313)
b. Which problem will [a solution to _] never be found? (Chaves, 2013: 301)

2. New empirical evidence for French

Following Abeillé & Winckel (2020) and Abeillé *et al.* (2020a, 2020b), we review some empirical evidence showing that French allows extraction out of subjects, but that not all constructions behave alike.

2.1 Some French corpus data

For French, we showed in Abeillé *et al.* (2016) and Abeillé & Winckel (2020) that *dont* relative clauses are mostly used to relativize the complement of the subject

in contemporary French. We used newspapers articles (French Treebank, Abeillé *et al.* (2019)), literary texts (FRANTEXT after 1900) and spoken data (CFPP 2000, Branca-Rosoff *et al.* (2012)), where extractions out of the subject with *dont* are very frequent:

(9) *[l'] Antillaise dont [l' effigie _] orna les boîtes de Banania.*
the Caribbean.F.SG of.which the effigy decorated the boxes of Banania
'the Caribbean girl whose picture decorated the Banania boxes' (*La solitude de la fleur blanche*, Roux, 2009) [3]

In line with Godard (1988), and contrary to Tellier (1990, 1991), Sportiche & Bellier (1989) and Heck (2009), we found that other relativizers (*de qui, duquel*) are also possible in French (examples from Frantext after 1900):

(10) a. *les ogres de qui [la danse barbare _] vous confisque l' enfance*
the ogres of who the dance barbaric DAT.2.PL seizes the childhood
'the ogres whose barbaric dance takes the childhood away from you' (*Pense à demain*, Garat, 2010)

b. *ce livre, duquel [la reliure _] tache les doigts de moisissure et [les feuilles _] sentent l' amande amère*
this book of.which the binding stains the fingers of mold and the sheets smell the almond bitter
'this book, whose binding stains the fingers with mold and whose sheets smell like bitter almond' (*La première fois*, Garat, 2013)

Extractions out of the subject are indeed the most common use of *dont* and *de qui* in the relative clauses of two Frantext subcorpora (1900–1913 and 2000–2013). Looking at *it*-clefts with *dont* or *de qui*, we found a few examples out of subject:[4]

(11) *C' était lui maintenant, dont [les yeux _] évitaient les yeux de l' autre.*
it was him now of.which the eyes *avoided* the eyes of the other
'Now it was him whose eye avoided the other's eye.' (*Jean-Christophe: Le Buisson ardent*, Romain Rolland, 1911)

Looking at *de qui* and *de quel*, we did not find any extraction out of the subject in *wh*-questions in the same corpora. But some are attested on the internet:

3. from FRANTEXT www.frantext.fr

4. There are two types of *it*-clefts in French (Moreau, 1971; Moreau, 1976): with an NP pivot and a relative clause as in (11), or with an XP pivot and a *que*-clause (*C'est de lui que les yeux _ évitaient les yeux de l'autre*), see below in Section 4.5. We did not search our corpora for the latter kind of clefts.

(12) *De quel pays [la dépense militaire _] dépasse annuellement mille*
of which country the budget military exceeds yearly thousand
milliards de dollars […]?[5]
billions of dollars
'Of which country does the military budget exceeds 100 Bn. dollars?'

We thus conclude that extraction out of the subject is possible in French, and very frequent in *dont* and *de qui* relative clauses.

2.2 Experimental data on French

Several experimental works have compared extraction from subject and from object, starting with Kluender & Kutas (1993). Although most corpus examples involve relative clauses, most of the experiments have tested *wh*-questions and found a subject penalty. Sprouse *et al.* (2016) found a subject penalty in *wh*-questions but not in relative clauses in Italian.

(13) a. *Di chi pensi che [il quadro _] raffiguri la nascita di Venere?*
of who think.2SG that the painting depicts the birth of Venus
'Who do you think the painting of depicts the birth of Venus?'

b. *Di chi pensi che il quadro di Maria raffiguri [la nascita _]?*
of who think.2SG that the painting of Maria depicts the birth
'Who do you think Maria's painting depicts the birth of?'

c. *Ho incontrato il giornalista del quale pensi che [l'*
AUX.1.SG met the journalist of.DEF.MSG whom think.2.SG that the
articolo_] abbia causato il licenziamento del direttore.
article AUX.3.SG caused the firing of.DEF.MSG director
'I met the journalist who you think that the article of caused the firing of the director.'

d. *Ho incontrato il giornalista del quale pensi che l'*
AUX.1.SG met the journalist of.DEF.MSG whom think.2.SG that the
articolo del direttore abbia causato [il licenziamento _].
article of.DEF.MSG director AUX.3.SG caused the firing
'I met the journalist who you think the director's article has caused the firing of.'

In Abeillé *et al.* (2020a), we compared extraction out of subject and object NP in English and French: we found a subject penalty in *wh*-questions ((14a) and (15a)) but not in relative clauses (with pied-piping, (14c) and (15c)).

5. https://www.voltairenet.org/article171526.html, last access 14/03/2020

(14) a. *Of which private practice did [the cleanliness _] impress the clients when they entered the room?*
b. *Of which private practice did the clients appreciate [the cleanliness _] when they entered the room?*
c. *The doctors ran a private practice, of which [the cleanliness _] impressed the clients when they entered the room.*
d. *The doctors ran a private practice, of which the clients appreciated [the cleanliness _] when they entered the room.*

(15) a. *De quel cabinet médical est -ce que [la propreté _] rassure les patients*
of which practice medical is it that the cleanliness reassures the clients
quand ils entrent?
when they enter
b. *De quel cabinet médical est -ce que les patients apprécient [la*
of which practice medical is it that the clients appreciate the
propreté _] quand ils entrent?
cleanliness when they enter
c. *Le docteur a un cabinet dont [la propreté _] rassure les patients*
the doctor has a practice of.which the cleanliness reassures the clients
quand ils entrent.
when they enter
d. *Le docteur a un cabinet dont les patients apprécient [la propreté _]*
the doctor has a practice of.which the clients appreciate the cleanliness
quand ils entrent.
when they enter

We also tested French *c'est*-clefts (16) (Abeillé *et al.'s*, 2020b), with materials similar to Abeillé *et al.* (2020a) experiment on relative clauses (15c). Although French allows for all-focus clefts (Lambrecht, 1994; Destruel, 2012), the narrow focus interpretation was more likely in our materials. We found the same subject penalty as with *wh*-questions: (16a) was rated lower than (16b).

(16) a. *C' est de ce cabinet que [la propreté _] rassure les patients quand ils*
it is of this practice that the cleanliness reassures the clients when they
entrent.
enter
b. *C' est de ce cabinet que les patients apprécient [la propreté _] quand*
it is of this practice that the clients appreciate the cleanliness when
ils entrent.
they enter

These experimental results are in line with the French corpus data. We conclude in Abeillé *et al.* (2020a) that the same difference between constructions holds across languages: a difficulty arises when extracting out of a subject in *wh*-questions, but not in relative clauses.[6]

3. A discourse-based account

In Construction Grammar, many of the constructions usually referred to as "islands" are viewed as the result of discourse infelicity (Ambridge & Goldberg, 2008; Goldberg, 2013): extraction is supposed to make a constituent more salient and is only felicitous if this constituent belongs to the "potential focus domain" (Erteschik-Shir, 1973; Takami, 1992; van Valin, 1995). In this perspective, islands are "backgrounded constituents", and the subject is an island because it is a primary topic, and not part of the potential focus domain. Thus sentences violating island constraints are not ungrammatical but unfelicitous, and can be ameliorated in the right discourse context. While this approach seems to be on the right track, it does not account for the cross-construction differences that were presented in Section 2.

Following suggestions in Functional and Construction Grammar, we propose in Abeillé *et al.* (2020a) a new constraint based on the discourse function of the construction. In a *wh*-question, the questioned element belongs to the focus domain (Jackendoff, 1972), but not in a relative clause. Assuming that the subject is the default topic (or theme) of the utterance (Kuno, 1976; Webelhuth, 2007), questioning the complement of a subject amounts to a discourse status clash (the topical subject has a focused subpart), but relativization does not. Abeillé *et al.* propose the following constraint:

(17) **Focus Background Conflict Constraint (FBC):**
It is unfelicitous to focus part of a backgrounded constituent.

This constraint is supposed to hold crosslinguistically. It predicts that only backgrounded subjects in focalizing constructions show a subject penalty. Since the relative clause, which adds a property to any discourse entity, is not a focalizing construction, no subject penalty is expected (2c, 10a, 14c, 15c). A *wh*-question, on the other hand, puts the questioned element into focus, hence a subject penalty (14a,15a). Another focalizing construction are *it*-clefts, which usually put the pivot element into focus (Lambrecht, 1994; Doetjes *et al.*, 2004; Destruel, 2012), hence

6. We found an English-specific subject penalty with preposition stranding: the preposition stranding counterparts of (13a) and (13c) (*Which private practice did the cleanliness of impress the clients? the doctors ran a private practice, which the cleanliness of impressed the clients*), were found degraded in our experiments. See (7) above however for better examples.

a subject penalty as well (16). But if the subject is not backgrounded, questioning (or clefting) its complement is predicted to become more natural: this was the case in the English question (8a) which can be interpreted as all-focus (Chaves, 2013).[7] As for the French question in (12), it is a rhetorical question in a critical speech about the United States, and therefore the question does not open an alternative set: the extracted element is more topical than focus.

Support from this view can be found in other Romance languages like Italian and Spanish: in these languages subject inversion is preferred when the subject is part of the focal domain. Uriagereka (1988) reports a contrast between pre and postverbal subjects when the complement of the subject is questioned:

(18) a. **¿De qué conferenciantes te perece que [las propuestas _]*
of what speakers DAT.2SG seem.3.SG that their proposals
me van a impresionar?
ACC.1SG go.3PL to impress.INF
'of which speakers does it seem to you that the proposals will impress me?'

b. *¿De qué conferenciantes te perece que me van a*
of what speakers DAT.2SG seem.3.SG that ACC.1SG go.3PL to
impresionar [las propuestas _]? (Gallego & Uriagereka, 2007: 162)
impress.INF their proposals
'Of which speakers does it seem to you that the proposals will impress me?'

A related constraint has been proposed for Italian by Bianchi & Chesi (2014) so that "only the subject of a thetic structure is transparent for extraction". They follow Kuno (1976) and Ladusaw (1994) in assuming that subjects are topics only in categorical sentences, and not in thetic sentences. In their view, only topic (or presupposed) subjects move to a higher position, and become "opaque" (=islands) for extraction (see also Diesing (1992)). They report a contrast so that (19b), with a postverbal subject, was rated higher than (19a), with a preverbal subject, in an acceptability study:

(19) a. *Di quale articolo ritieni che [una revisione _] sarebbe ormai*
of which section think.2SG that a revision be.COND by.now
opportuna?
appropriate
'Of which section do you think that a revision would be appropriate by now?'

7. Abeillé *et al.* (2020) apply a topic test to (8a), and find that it does not pass it:

(i) a. A solution to this problem will never be found.
b. #Speaking of a solution to this problem, it will never be found.

b. *Di quale articolo ritieni che sarebbe ormai opportuna [una*
of which section think.2SG that be.COND by.now appropriate a
revisione _]?
revision

Unlike the FBC constraint (17), which takes into account the discourse function of the construction, they predict the same contrast in relative clauses:

(20) a. *il personaggio di cui è stata pubblicata [un' intervista _]*
the personality of whom AUX.3.SG been published an interview
'the personality of whom an interview has been published'
b. **il personaggio di cui [un' intervista _] è stata pubblicata*
the personality of whom an interview AUX.3.SG been published
(Bianchi & Chesi, 2014: 533)

We conclude that the grammar of French, like that of other Romance languages, may involve a constraint that limit extraction out of subject in some cases, that however this constraint is not syntactic but discourse-based.

As for English, we follow Abeillé *et al.* (2020a) who found the same cross-construction difference with pied-piping (14a, c) and notice an interaction with preposition stranding which penalizes extraction out of subject across constructions (*??Which private practice did the cleanliness of _ impress the clients when they entered the room?*; *??The doctors ran a private practice, which the cleanliness of _ impressed the clients when they entered the room.*). They notice that extraction out of a PP is not the same as extraction out of an NP, and suggest that this extra penalty may be due to processing factors.

4. An HPSG account of French extraction constructions

We briefly review the analysis of extraction in Head-Driven Phrase Structure Grammar (HPSG), adopting the lexicalist traceless analysis of Bouma *et al.* (2001), and the construction-based analysis of Ginzburg & Sag (2000), Sag (1997) (2010) and Abeillé & Godard (2007). We then show how a constraint like the FBC (17) can be integrated into an HPSG grammar of French.

4.1 A lexicalist view on extraction

We follow Bouma *et al.* (2001) and Ginzburg & Sag (2000) who propose a lexicalist analysis of long distance dependencies. In this approach, the value of the attribute SYNSEM, that encodes syntactic and semantic information, is either *non-*

canonical (for elements non realized in their canonical position, i.e. gaps) or *canonical*. Therefore, no empty complements are needed in the syntax. The feature SYNSEM encodes both local (CAT, CONT, CTXT) and NON-LOCAL (SLASH, REL, WH) information and has a value of type *synsem* with different subtypes: the type *pro* is for unexpressed complements (*Paul mange.* 'Paul is eating') and the type *aff* is for pronominal clitics analysed as verbal affixes (*Paul le mange.* 'Paul is eating it') (Miller & Sag (1997); Abeillé & Godard (2002); Bîlbîie (2017); Aguilar-Multner & Crysmann, this volume).

Predicates code their syntactic arguments in an argument structure (ARG-ST) list, and realize them through three valence features: specifier (SPR), subject (SUBJ) and complements (COMPS). The Argument Realization Principle (ARP) ensures the correspondence between the list of expected arguments and the valence features (Bouma *et al.*, 2001; Bîlbîie, 2017).

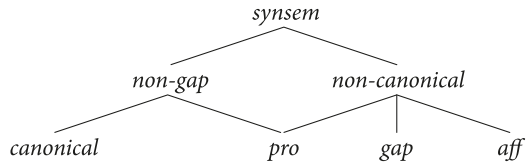

Figure 1. Type hierarchy of *synsem*

(21) **Argument Realization Principle:**

$$\text{word} \rightarrow \begin{bmatrix} \text{SPR} & \boxed{1} \\ \text{SUBJ} & \boxed{2} \\ \text{COMPS} & \boxed{3} \\ \text{ARG-ST} & \boxed{1} \oplus \boxed{2} \oplus \boxed{3} \bigcirc \textit{list\_of\_non-canonical} \end{bmatrix}$$

A head combines locally with its canonical arguments, via the *head-spr-phrase*, the *head-subject-phrase* and the *head-complements-phrase*, while the non-canonical arguments may be morphologically incorporated (Miller & Sag, 1997; Aguila-Multner & Crysmann, this volume) or realized non-locally. When an argument is extracted, it is typed as *gap* (which is a subtype of *non-canonical*) in the ARG-ST list of the predicate, and may be realized non-locally as a filler.[8] A *gap* synsem is a description with a non empty SLASH feature corresponding to its local (syntactic and semantic) features:

(22) $$\textit{gap} \rightarrow \begin{bmatrix} \text{LOC} & \boxed{1} \\ \text{NON-LOC}|\text{SLASH} & \{\boxed{1}\} \end{bmatrix}$$

8. Argument means syntactic argument (member of ARG-ST). Following Bouma *et al.* (2001) and Abeillé & Godard (1997), adjuncts are analysed as complements (members of ARG-ST) when they can be extracted (*Where did you sleep _?*), and only extractable adjuncts can be extracted from (*What did you sleep on _?*).

As a non-local feature, SLASH mediates the dependency between the expected gap argument and the filler. For example in (23), the noun *couleur* has a gap prepositional argument, corresponding to the fronted PP, and a corresponding SLASH feature with a prepositional value.[9]

(23) *De quelle voiture Paul aime-t-il [la couleur _]?*
of which car Paul like.3MSG the color
'Of which car does Paul like the color?'

Following Ginzburg & Sag (2000) and Sag (2010), we assume the lexical constraint in (24), which ensures that a head inherits the non-local features of its arguments:[10]

(24) **Non-Local amalgamation principle:**
A word's non-local feature is by default the set union of the non-local features of its arguments.

This enables the main verb to inherit the SLASH information of any of its arguments: extraction is thus possible out of the subject or out of the object. For example, in (23), the verb inherits the SLASH information of the nominal complement *couleur*. The verb's SLASH feature thus contains a PP[de].

---

9. See Godard (1992), Kolliakou (1999) and Godard & Sag (1996) for a detailed study of which complements of (which) nouns can be extracted in French.

10. This principle marks with a non-empty SLASH set all heads along an extraction path. This accounts for binding domains discovered by Zaenen (1983): Certain languages have different verb forms or complementizers not only when the verb has an extracted argument, but also for all verbs along an extraction path. Stylistic inversion in French may be such a case (see Bonami *et al.* (1999) for another analysis):

(i) *Je me demande à qui Paul parle / parle Paul.*
I REFL ask at who Paul talks/ talks Paul
'I wonder who Paul is talking to.'

(ii) *Je me demande à qui Paul veut parler / veut parler Paul*
I REFL ask at who Paul wants talk.INF / wants talk.INF Paul
(Kayne & Pollock, 1978)

The non-local amalgamation principle is a default constraint because certain heads may block the propagation of SLASH information. This is the case with *tough* adjectives in English (Pollard & Sag, 1994; Ginzburg & Sag, 2000):

(iii) This sonata is easy to play _ on this violin.

The adjective (*easy*) coindexes the gap complement of the infinitive (*play*) with the subject noun (*sonata*) and blocks the propagation of the gap to the rest of the sentence.

(25) a. *couleur* ('color'):

$$\begin{bmatrix} \textsc{arg-st} & \left\langle \text{det , PP} \quad \begin{bmatrix} \textit{gap} \\ \textsc{slash} & \{\boxed{1}\} \end{bmatrix} \right\rangle \\ \textsc{slash} & \{\boxed{1}\} \end{bmatrix}$$

b. *aime-t-il* ('like-3.M.SG'):

$$\begin{bmatrix} \textsc{arg-st} & \langle \text{NP, NP}[\textsc{slash} \ \{\boxed{1}\}] \rangle \\ \textsc{slash} & \{\boxed{1}\} \end{bmatrix}$$

c. *quelle* ('which'):

$$\begin{bmatrix} \textsc{head} & \textit{Det} \\ \textsc{cont} & \boxed{1} \begin{bmatrix} \textit{param} \\ \textsc{index} \ i \end{bmatrix} \\ \textsc{wh} & \{\boxed{1}\} \end{bmatrix}$$

The SLASH feature of the verb is inherited by the VP via the Generalized Head Feature Principle, which ensures that (by default) syntactic and semantic (SYNSEM) features are shared between (headed) phrases and heads:[11]

(26) **Generalized Head Feature Principle:**
A headed-phrase's SYNSEM features are by default the SYNSEM's features of its head-daughter.

A simplified tree-like representation of (23), using the *head-filler-phrase*, which builds a sentence with a fronted filler matching an element of the SLASH set of the Head Daughter, is provided in Figure 2.

(27) *head-filler-phrase:* →

$$\begin{bmatrix} \textsc{slash} & \boxed{1} \\ \textsc{head-dtr} & [\textsc{slash} \ \boxed{1} \cup \boxed{2}] \\ \textsc{filler-dtr} & [\textsc{loc} \ \boxed{2}] \end{bmatrix}$$

In this tree, the object noun (*couleur*) has a missing PP complement (SLASH PP), this information is inherited by the main verb (*aimer*) via the Non-local amalgamation Principle, and shared with the VP node via the Generalized Head feature principle. The top node has an empty SLASH set because it is a head-filler phrase: the filler (*de quelle voiture* 'of which car') matches the SLASH value of the VP node and is removed from the SLASH set at the S node.

11. It is a default principle, which can be overridden by more specific principles: The valence features of the phrase (which are part of SYNSEM) may differ from the valence features of the head. For example, the *head-subject-phrase* states that the SUBJ value of its HEAD-DTR is not empty and must match the SYNSEM value of its SUBJ-DTR, and the SUBJ value of the phrase is empty. The RELS features of a phrase (which are part of SYNSEM) also differ from the RELS of the head; see below.

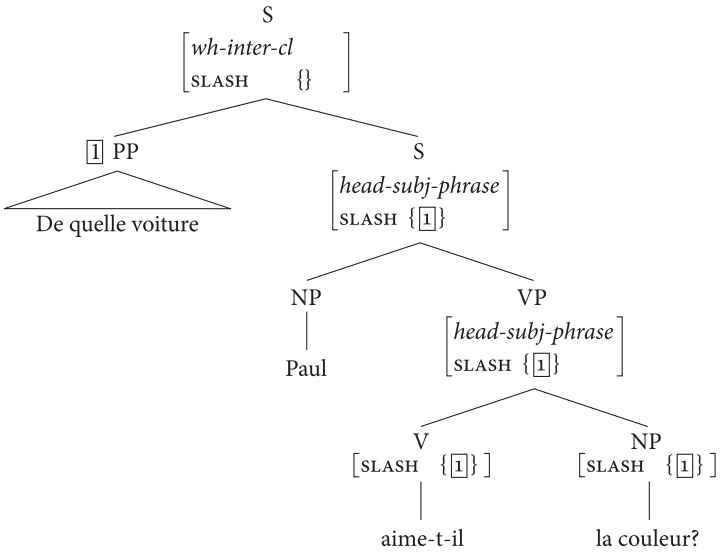

Figure 2. Simplified tree for (23)

4.2 Locality constraints in HPSG

Hofmeister & Sag (2010: 406) suggested that most 'island constraints' are "arbitrary in the sense that they bear no relationship to other constraints, emanate from no general principles of language" and "offer little insight into anything about language or cognition, except islands themselves." Recent work in HPSG assumes that no general locality constraints are needed in the syntax of long distance dependencies and that possible difficulties may arise due to different heterogenous factors (Chaves, 2012; Chaves, 2013; Chaves & Putnam, 2020). For example, the mere definition of coordination phrases, together with a traceless account of extraction, is enough to rule out extraction of a conjunct: a coordination phrase must have at least two conjuncts, thus preventing **Who did you see _ and Mary?* and coordinating conjunctions cannot stand alone, thus preventing **Who did you see Mary and _?* (Sag, 2010: 511; Chaves, 2012: 505–507).[12]

Of course, there may be language specific constraints: for example, prepositions cannot be stranded in French, as in other Romance languages (28a); but some complements may be extracted out of a preposition phrase (28b):

12. As shown by Goldsmith (1985), Lakoff (1986), Kehler (2002), Goldberg (2013) a.o., Ross' original Coordinate Structure constraint was too strong, since extraction is allowed out of both conjuncts (*Which book did John buy _ and Mary read _?*) but also out of a single conjunct (*What did Bob go and buy _?*; *How many beers can you drink _ and stay sober?*).

(28) a. **Qui as-tu parlé à*?*
who have.2.SG spoken to
'Who did you spoke to?'

b. *[…] il a vu de qui […] cette morne et interminable liste avait*
he AUX.3.SG seen of who this bleak and endless list had
le chagrin de faire part [de la soudaine et tragique
the grief to do announce of the sudden and tragic
disparition _] […].
loss
(*La voix des mauvais jours et des chagrins rentrés*, Jean-Luc Benoziglio, 2004)
'He saw who this endless bleak list was very sorry to announce the sudden and tragic loss of.'

This can be captured by a constraint on Romance prepositions that states that their ARG-ST list must be *canonical* or *pro* (see Figure 1), in order to allow for a null complement: *Je joue avec.* 'I play with (it)', *Je suis pour.* 'I am for (it)'.

(29) *prep* → [ARG-ST *list* (*non-gap*)] [Romance]

4.3 A classification of extraction constructions

As shown by Sag (1997, 2010) for English, extraction is a cover term for a variety of constructions, that have in common to allow long distance dependencies between filler and gap, but nonetheless present many syntactic, semantic and pragmatic differences. For example, *wh*-questions can be main or subordinate clauses, whereas relative clauses are always subordinate; *wh*-questions must comprise a *wh*-word (*Who* do you love?), while relative clauses may have a relative word (*the man whom I love*), a complementizer (*the man that I love*) or nothing (*the man I love*). To capture the commonalities as well as the differences, he proposes a cross-classification of extraction constructions. Notice that he does not define specific syntactic constraints on long distance dependencies: he assumes that difficulties may come from processing factors (Hofmeister & Sag, 2010; Hofmeister, 2011).

Wh-questions and relative clauses come in a variety of different subconstructions (Ginzburg & Sag (2000), Sag (1997, 2010) for English), see Figure 3. We distinguish three kinds of *wh*-words: relative words (with a REL feature), exclamative words (which we ignore here) and interrogative words (with a WH feature). In French, *quand* 'when' can only be an interrogative word, while *où* 'where' can serve both in questions and relative clauses. *Quand* thus has a non-empty WH feature, like locative *où*, while relative *où* has a non-empty REL feature (with a temporal or locative meaning). REL and WH are both non-local features, propagated by

the Non-Local Amalgamation Principle (and the Generalized Head Feature Principle).

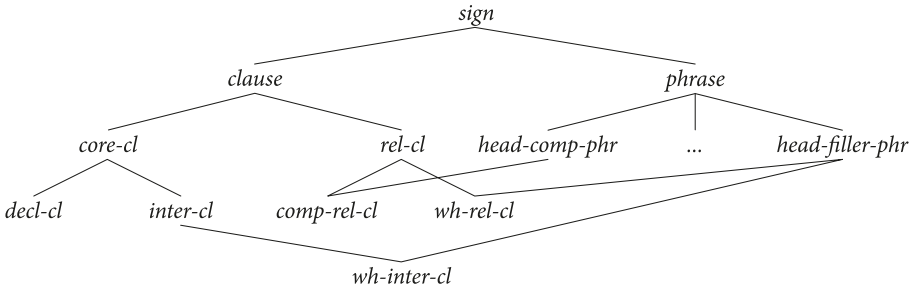

Figure 3. Type hierarchy of clauses

For interrogatives, Ginzburg & Sag (2000) propose that interrogative words have a non-empty WH feature and a STORE value of type *parameter* when they are fronted. Since WH is a NON-LOCAL feature, it is shared between a *wh*-word and the rest of the *wh*-phrase (between *quelle* and *de quelle voiture* for instance). *Wh*-questions are subtypes of *head-filler-phrase* with a *wh*-filler:

(30) *wh-inter-cl* → [FILLER-DTR [WH *non-empty-set* (*param*)]]

We now turn to the classification of relative clauses before formalizing our discourse constraint.

4.4 Cross-classification of French relative clauses

We follow Abeillé & Godard's (2007) classification of relative clauses, and adapt it to our present purposes. We follow Godard's (1988) distinction between relative pronouns on one hand (*où* 'where', (preposition) *qui* 'who', *lequel* 'which') and complementizers (*que, qui* 'that', *dont* 'of which') on the other hand. Contrary to relative pronouns, complementizers cannot occur in infinitival clauses (31) and cannot be pied-piped (32).

(31) a. *un endroit où aller / *que visiter*
a place where go.INF that visit.INF
'a place to go / a place to visit'
b. *un enfant de qui / *dont parler*
a child of who of.which talk.INF
'a child to talk about'

(32) a. *un collègue avec le fils de qui je partirai*
a colleague with the son of who I leave.FUT.1.SG
'a colleague the son of whom I will leave with'
b. **un collègue avec le fils dont je partirai*
a colleague with the son of.which I leave.FUT.1.SG

Notice that subject *qui* is analyzed as a variant of complementizer *que*. Only relative pronouns, or PP comprising relative pronouns, can serve as fillers; complementizers are heads. Abeillé & Godard (2007: 48–49) propose the crossclassification presented in Figure 3. Relative clauses are specific constructions which are headed by a verbal category (a verb of a complementizer) and which modify a Noun or Noun Phrase.[13]

(33)
$$\textit{rel-cl} \rightarrow \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verbal} & \\ \text{MOD} & \left[\text{HEAD } \textit{noun}\right] \end{bmatrix} \\ \text{SLASH} & \{\} \\ \text{REL} & \{\} \end{bmatrix}$$

wh-rel-clauses inherit their properties from *rel-cl* and *head-filler-phrases*. They are constrained in French to have a PP filler, with the non-local feature REL coindexed with the antecedent of the relative clause:

(34)
$$\textit{wh-rel-cl} \rightarrow \begin{bmatrix} \text{HEAD-DTR} & \left[\text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{MOD} \left[\text{INDEX } \boxed{\text{i}}\right] \end{bmatrix}\right] \\ \text{FILLER-DTR} & \begin{bmatrix} \text{HEAD } \textit{prep} \\ \text{REL } \left\{\boxed{\text{i}}\right\} \end{bmatrix} \end{bmatrix}$$

Contrary to English, the filler cannot be an NP and must be a PP:

(35) a. a child [whose mother] I can't reach
b. **un enfant [la mère de qui] je n' arrive pas à joindre*
a child the mother of whom I NEG succeed not to reach
'a child the mother of whom I cannot reach' (Abeillé & Godard, 2007: 49)
c. *un enfant [avec la mère de qui] je peux parler*
a child with the mother of whom I can talk
'a child with the mother of whom I can talk'

An example of a *wh*-relative clause with *de qui* extraction out of the subject is shown in (36a): The noun *danse* has a gap complement, and the corresponding

13. In what follows, we ignore gapless and verbless relatives (see Bîlbîie & Laurens (2010) for a presentation and discussion of verbless relative clauses in French).

SLASH information is inherited by the main verb *plaît* and percolated up the tree until it finds a matching filler (*de qui*). A simplified tree-like representation is shown in Figure 4.

(36) a. *un homme de qui la danse me plaît*
a man of who the dance DAT.1SG pleases

b. *danse*:

$$\begin{bmatrix} \text{SPR} & \langle \boxed{1}\ \text{DET} \rangle \\ \text{COMPS} & \langle\rangle \\ \text{ARG-ST} & \left\langle \boxed{1}, \text{PP} \begin{bmatrix} \textit{gap} \\ \text{SLASH}\ \boxed{2} \end{bmatrix} \right\rangle \\ \text{SLASH} & \{\boxed{2}\} \end{bmatrix}$$

c. *me-plaît*:

$$\begin{bmatrix} \text{SPR} & \langle \boxed{1}\ \text{NP} \rangle \\ \text{COMPS} & \langle\rangle \\ \text{ARG-ST} & \left\langle \boxed{1}\left[\text{SLASH}\ \{\boxed{2}\ \text{PP}\ [\text{de}]\}\right], \text{PP} \begin{bmatrix} \textit{aff} \\ \textit{à} \end{bmatrix} \right\rangle \\ \text{SLASH} & \{\boxed{2}\} \end{bmatrix}$$

Treating the complementizer as a verbal head, *comp-rel-clause* inherit their properties from *rel-cl* and *head-comp-phr*. They are constrained to be finite and coindex the gap (in SLASH) with the antecedent of the relative clause:

(37)

$$\textit{comp-rel-cl} \rightarrow \begin{bmatrix} \text{MARKING} & \textit{dont} \vee \textit{que} \\ \text{HEAD-DTR} & \left[\text{HEAD} \begin{bmatrix} \textit{comp} \\ \text{MOD}\ [\text{INDEX}\ i] \end{bmatrix} \right] \\ \text{COMP-DTR} & \left[\begin{matrix} \text{HEAD}|\text{VFORM} & \textit{finite} \\ \text{NON-LOC}|\text{SLASH} & \{[\text{INDEX}\ i]\} \end{matrix} \right] \end{bmatrix}$$

They are headed by *que* or *dont*. A simplified entry for *dont* is as follows: it takes as complement a finite clause with a gap PP[*de*] and marks the clause with *dont*.

(38)

$$\textit{dont}: \begin{bmatrix} \text{HEAD} & \textit{comp} \\ \text{MARKING} & \textit{dont} \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle \text{S}\ [\text{SLASH}\ \{\text{PP}[\textit{de}]\}] \rangle \end{bmatrix}$$

An example of a *dont* relative clause with subject extraction is illustrated in (39a). The noun *couleur* has a gap PP argument, and a corresponding SLASH information, that is inherited by the verb *plaît* and shared by the *head-subject-phrase*. This

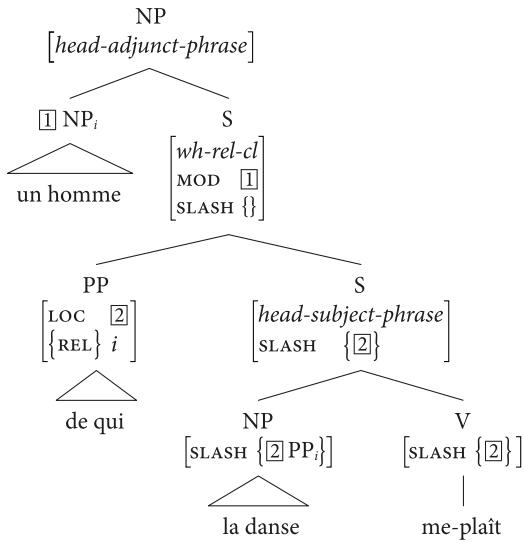

Figure 4. Simplified tree for (36a)

SLASH feature is inherited by *dont*, but not propagated further since the relative clause is constructionally constrained by (33) to be SLASH empty. A simplified tree-like representation is shown in Figure 5.

(39) a. *la voiture dont la couleur me plaît*
the car of.which the color 1.SG pleases
'the car whose color pleases me.'

b. *couleur*: $\begin{bmatrix} \text{SPR} & \langle \boxed{1}\ \text{DET} \rangle \\ \text{COMPS} & \langle\rangle \\ \text{ARG-ST} & \left\langle \boxed{1}, \text{PP} \begin{bmatrix} gap \\ \text{LOC}\ \boxed{2} \end{bmatrix} \right\rangle \\ \text{SLASH} & \{\boxed{2}\} \end{bmatrix}$

c. *me-plaît*: $\begin{bmatrix} \text{SPR} & \langle \boxed{1}\text{NP} \rangle \\ \text{COMPS} & \langle\rangle \\ \text{ARG-ST} & \left\langle \boxed{1}\left[\text{SLASH}\ \{\boxed{2}\ \text{PP}\ [de]\}\right], \text{PP} \begin{bmatrix} aff \\ à \end{bmatrix} \right\rangle \\ \text{SLASH} & \{\boxed{2}\} \end{bmatrix}$

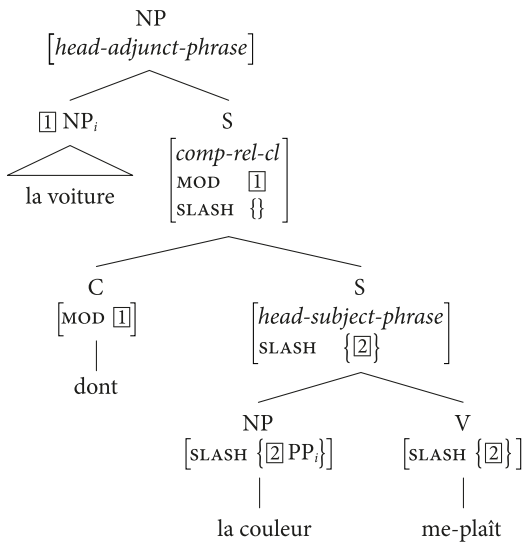

Figure 5. Simplified tree for (39a) (syntactic structure)

4.5 *C'est* clefts

We adapt Pollard & Sag (1994: 260–262) and Sag (2010: 528)'s analysis of *it*-clefts as having a special entry for *be* with two complements: the pivot and the *that*-clause.

For French, we have a special entry for *être* taking (expletive) *ce* as a subject and two complements: the pivot which can be of any category, and the *que* clause with a gap coindexed (or shared) with the pivot.[14, 15]

(40) *être* in *it*-cleft:

$$\left[\begin{array}{l}\textsc{arg-st}\ \left\langle \textsc{np}\ \left[ce\right], \left[\begin{array}{ll}\textsc{loc} & \boxed{1}\\ \textsc{slash} & \boxed{2}\end{array}\right], \textsc{s}\left[\begin{array}{ll}\textsc{marking} & que\\ \textsc{slash} & \{\boxed{1}\}\end{array}\right]\right\rangle \\ \textsc{slash}\ \boxed{2}\end{array}\right]$$

14. As in Abeillé & Godard (2007), MARKING *que* is used for both *que* and *qui*.
Notice that *être* stops SLASH propagation for the *que*-clause, and only inherits the SLASH information of the pivot. This SLASH set will usually be empty but could allow for extraction out of the pivot as in (i):

(i) *un élève dont c' est toujours [le père _] que je vois aux réunions*
a pupil of.which it is always the father that I see at.DEF.PL meetings
'a pupil of which it is always the father that I see by the meetings'

15. For the second type of clefts, with an NP pivot and a relative clause, as in (11), a related entry for *être* is needed: $\left[\text{ARG-ST}\ \left\langle \text{NP}\left[ce\right], \text{NP}_i, \left[\begin{array}{ll}\textit{rel-cl} & \\ \textsc{mod} & i\end{array}\right]\right\rangle\right]$. The information structure of this second type of *it*-clefts has not been studied, and we ignore it in what follows.

Figure 6 gives a simplified tree-like representation of (41).

(41) *C' est de cette voiture que j' aime la couleur*
it is of this car that I like the color
'It is of this car that I like the color.'

5. A discourse-based HPSG account

We sketch how information structure can be encoded within HPSG (see also De Kuthy (2020) and Abeillé *et al.* (2008) for spoken French), before turning to a reformulation of the Focus-Background Conflict constraint (17).

5.1 Encoding information structure in HPSG

Different views of information structure have been proposed in HPSG (e.g. Engdahl & Vallduví (1996) put it inside a context CTXT attribute while De Kuthy (2002) and Bildhauer (2008) put it inside the semantic content CONT). We follow Song (2017) who represents Topic and Focus within Minimal Recursion Semantic (Copestake *et al.*, 2005), which is the semantics implemented in LKB grammars (Copestake, 2001) and in the CoreGram Project (Müller, 2013). He encodes information structure in an Individual CONStraints (ICONS) feature, whose value is a list of *icons, info-str* being a subtype of *icons*.[16] This is part of the content features of the sign. In this approach, the index of a phrase is that of its head daughter, while the elementary relations (RELS) and information structures (ICONS) are amalgamated from DTRS to mother.

(42) *headed-phrase* →

$$\left[\begin{array}{ll}
\textsc{cont} & \left[\begin{array}{ll}
\textsc{index} & \boxed{1} \\
\textsc{rels} & \boxed{2} \oplus \boxed{4} \ldots \oplus \boxed{n} \\
\textsc{icons} & \boxed{3} \oplus \boxed{5} \ldots \oplus \boxed{m}
\end{array}\right] \\
\textsc{head-dtr} & \left[\textsc{cont} \left[\begin{array}{ll}
\textsc{index} & \boxed{1} \\
\textsc{rels} & \boxed{2} \\
\textsc{icons} & \boxed{3}
\end{array}\right]\right] \\
\textsc{non-head-dtrs} & \left\langle \left[\textsc{cont} \left[\begin{array}{ll}
\textsc{rels} & \boxed{4} \\
\textsc{icons} & \boxed{5}
\end{array}\right]\right] \ldots \left[\textsc{cont} \left[\begin{array}{ll}
\textsc{rels} & \boxed{n} \\
\textsc{icons} & \boxed{m}
\end{array}\right]\right] \right\rangle
\end{array}\right]$$

16. For the sake of clarity, we will assume in this paper that the ICONS list only contains object of the type *info-str*. There are however other subtypes of *icons* objects (Song, 2017: 103), but this is irrelevant for our purpose.

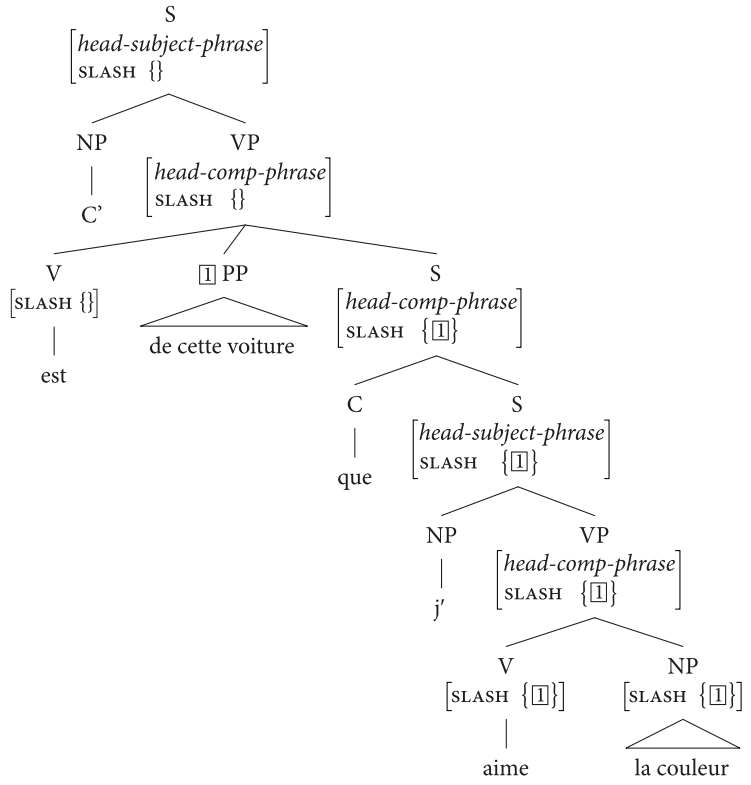

Figure 6. Simplified tree for (41)

The type hierarchy of *info-str* is very fine-grained and ranges from underspecified types like *non-topic* or *non-focus* to various types of topic or focus, like *contrast-focus* or *semantic-focus* (see Figure 7). The focalization in *wh*-questions, that have informational focus, falls therefore under the type *semantic-focus*, and the focalization in clefts under the type *contrast-focus*.

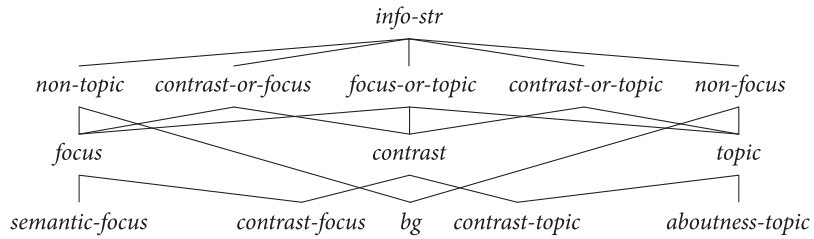

Figure 7. Type hierarchy of *info-str* (Song, 2017: 114)

A structure of type *info-str* describes the kind of information structure relation holding between a TARGET (the element focused, topicalized or back-

grounded)[17] and a CLAUSE (its clausal scope), as shown in (43). Every linguistic unit with a semantic meaning introduces one *info-str* value. There cannot be more than one *info-str* object for the same CLAUSE/TARGET combination.

(43)
$$\begin{bmatrix} \textit{info-str} \\ \text{CLAUSE} \;\; \textit{individual} \\ \text{TARGET} \;\; \textit{individual} \end{bmatrix}$$

Following Webelhuth (2007) and Abeillé *et al* (2008) a.o., we assume that in a categorical clause, the subject is the topic:

(44) *categorical-cl* →
$$\begin{bmatrix} \textit{head-subj-phrase} \\ \text{CONT}|\text{INDEX} \;\boxed{1} \\ \text{SUBJ-DTR} \;\; \begin{bmatrix} \text{CONT} \;\; \begin{bmatrix} \text{INDEX} \;\boxed{2} \\ \text{ICONS} \;\; \textit{list} \oplus \left\langle \begin{bmatrix} \textit{topic} \\ \text{CLAUSE} \;\boxed{1} \\ \text{TARGET} \;\boxed{2} \end{bmatrix} \right\rangle \oplus \textit{list} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 8 illustrates with a simplified tree-like representation of (45) how the information structure is collected at the clausal level for a sentence with a topical subject.

(45) *La couleur de cette voiture me plaît.*
the color of this car DAT.1.SG pleases
'The color of this car pleases me.'

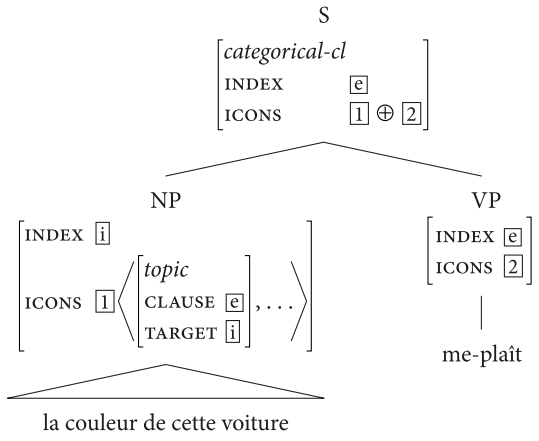

Figure 8. Simplified tree for (45) (syntactic structure)

17. Notice that Song (2017) calls "background" what is neither focus nor topic (similar to Goldberg's (2006, 2013) definition of background). In (17), our use of the term "background" relies on the distinction Focus/Background (Krifka, 2007), and includes topics. Our background corresponds therefore to Song's "non-focus". (17) can be reformulated as "It is unfelicitous to put focus on a non-focus constituent." in order to match Song's terminology.

5.2 Information structure in extractions

In this analysis, fronted *wh*-phrases are focused in interrogatives. We therefore add to the Filler Inclusion Constraint in Ginzburg & Sag (2000: 228) a focus feature on the extracted phrase.[18]

(46) *wh-inter-cl* →

$$\begin{bmatrix} \textsc{head-dtr} & \begin{bmatrix} \textsc{index} & \boxed{1} \end{bmatrix} \\ \textsc{filler-dtr} & \begin{bmatrix} \textsc{index} & \boxed{2} \\ \textsc{icons} & \mathit{list} \oplus \left\langle \begin{bmatrix} \mathit{semantic\text{-}focus} \\ \textsc{target} & \boxed{1} \\ \textsc{clause} & \boxed{2} \end{bmatrix} \right\rangle \oplus \mathit{list} \end{bmatrix} \end{bmatrix}$$

We also add to the lexical entry for the *it*-cleft *être* given in (40) some information about the information structure, as in (48). This is compatible with both uses of *it*-clefts (Lambrecht, 1994; Doetjes *et al.*, 2004): in a narrow focus *it*-cleft like (47a), the subject is further typed *contrastive-focus*, while in a broad focus cleft as in (47b), the S *que*-clause will also be marked *focus*.

(47) a. *A: Qui vient? B: C' est Paul qui vient.*
who comes it is Paul who comes
'Who is coming? Paul is coming.'

b. *A: Qu' est-ce qui se passe? B: C' est le téléphone qui sonne.*
what is-3SG that REFL happens it is the phone that rings
'What is happening? The phone is ringing.'

(48) *être* in *it*-clefts:

$$\begin{bmatrix} \textsc{arg-st} & \left\langle \mathrm{NP}\left[\mathit{ce}\right], \begin{bmatrix} \textsc{index} & \boxed{3} \\ \textsc{icons} & \mathit{list} \oplus \left\langle \begin{bmatrix} \mathit{focus} \\ \textsc{target} & \boxed{3} \\ \textsc{clause} & \boxed{4} \end{bmatrix} \right\rangle \oplus \mathit{list} \end{bmatrix}, \mathrm{S}\begin{bmatrix} \textsc{marking} & \mathit{que} \\ \textsc{cont}\vert\textsc{index} & \boxed{4} \end{bmatrix} \right\rangle \\ \textsc{index} & \boxed{4} \end{bmatrix}$$

5.3 Our proposal

We propose to formalize the Focus-Background Conflict constraint (17) – a non-focus element should not comprise a focus element[19] – as (49).

18. Another possibility would be to define interrogative *wh*-words as having this focus feature. However, in-situ interrogatives seem to have a different information structure than interrogatives with the *wh*-word extracted, as least in French. For this reason, we prefer to add the constraint on focus on the clausal level.

19. This bears some resemblance with Webelhuth's (2007) Preference Principle: A topic should not contain any focused element.

(49) **Focus Background Conflict constraint (formalization):**

$$\left[\begin{array}{l} \textit{word} \\ \textsc{head}\ \ \textit{non-verb} \\ \textsc{arg-st}\ \textit{list} \oplus \left\langle \left[\begin{array}{l}\textsc{index}\ \boxed{3} \\ \textsc{icons}\ \textit{list} \oplus \left\langle \boxed{4}\left[\begin{array}{l}\textsc{target}\ \ \boxed{3} \\ \textsc{clause}\ \ \boxed{2}\end{array}\right]\right\rangle \oplus \textit{list}\end{array}\right]\right\rangle \oplus \textit{list} \\ \textsc{cont}\ \ \left[\begin{array}{l}\textsc{index}\ \boxed{1} \\ \textsc{icons}\ \oplus \textit{list}\left\langle \left[\begin{array}{l}\textit{non-focus} \\ \textsc{target}\ \ \boxed{1} \\ \textsc{clause}\ \ \boxed{2}\end{array}\right]\right\rangle \oplus \textit{list}\end{array}\right] \end{array}\right] \rightarrow \boxed{4}\ \textit{non-focus}$$

The rule states that any element that has a non-focus status with respect to a certain clause does not licence any of the element in its ARG-ST list to have a focus status with respect to the same clause. Because the value of CLAUSE is the index of the finite verb of the current clause, we need to make sure that this does not apply to verbs that have a non-focus status. That's why the constraint only applies to non-verbs. Otherwise, a non-focus verb would constrain the discourse status of its complements.

A non-focus noun, for example a topic subject NP as in Figure 9, can only have non-focus complements, otherwise this would violate (49):

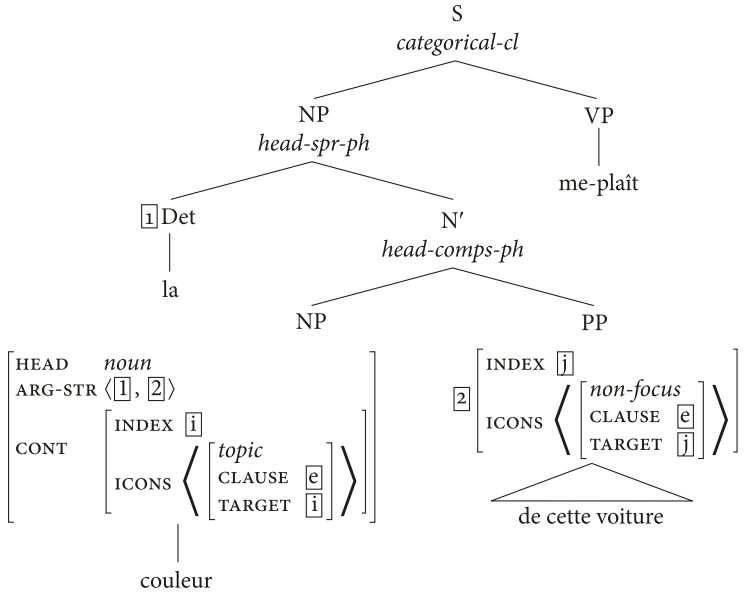

Figure 9. Simplified tree for (45) (information structure)

Therefore, the FBC constraint does not concern extractions alone. Indeed, if *de cette voiture* in Figure 9 had a focus status, the NP would be infelicitous. It may however be contrastive (contrastive topic) and accented.

Examples like (50) are ruled out as well. Indeed, as Figure 10 shows, the subject is the topic of the clause, but its ARG-ST list contains an element that is non-focus (because it is the filler in a *wh-inter-cl*) with respect to the same clause. This violates (49) and is therefore not well formed (we ignore SLASH information in the following trees since it is not relevant).

(50) #*[De quelle voiture]*$_{Focus}$ *[la couleur]*$_{Topic}$ *te plaît-elle?*
of which car the color DAT.2.SG please-3.F.SG
'Of which car does the color please you?'

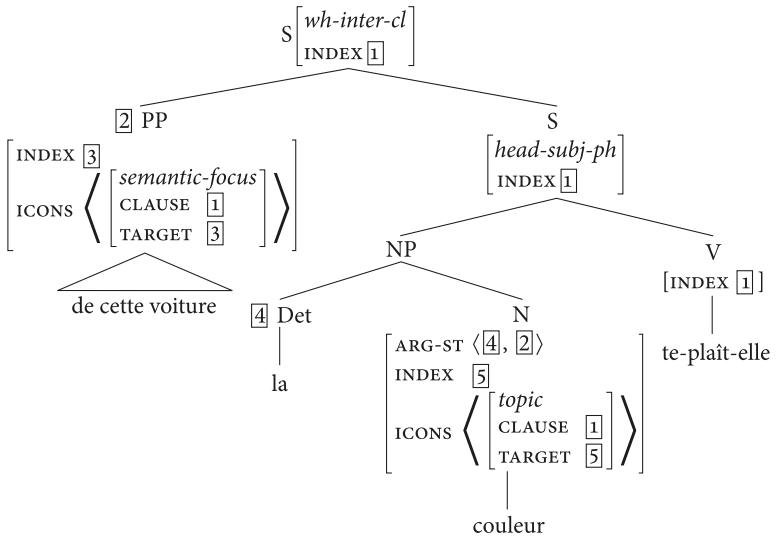

Figure 10. Simplified tree for (50)

In focalizing *it*-cleft sentences also, extraction out of topic NP subjects is ruled out. An infelicitous sentence like (51) would result in the tree in Figure 11 that is not well-formed by virtue of (49).

(51) #*C' est [de cette voiture]*$_{Focus}$ *que [la couleur]*$_{Topic}$ *me plaît*
it is of this car that the color DAT.1.SG pleases
'It is of this car that the color pleases me.'

Notice that (50) and (51) are ruled out with their specific discourse status (extracted element focus + subject topic) but the same sentences with another discourse status for their elements are not ruled out. For example, the constraint (49) is not violated if the subject is a focus or part of a focus, for instance in all-focus clefts. We may consider that the subject is focal in thetic sentences (Kuroda, 1976; Abeillé *et al.*, 2008) and in good examples like (3c), (8a-b) and (12).

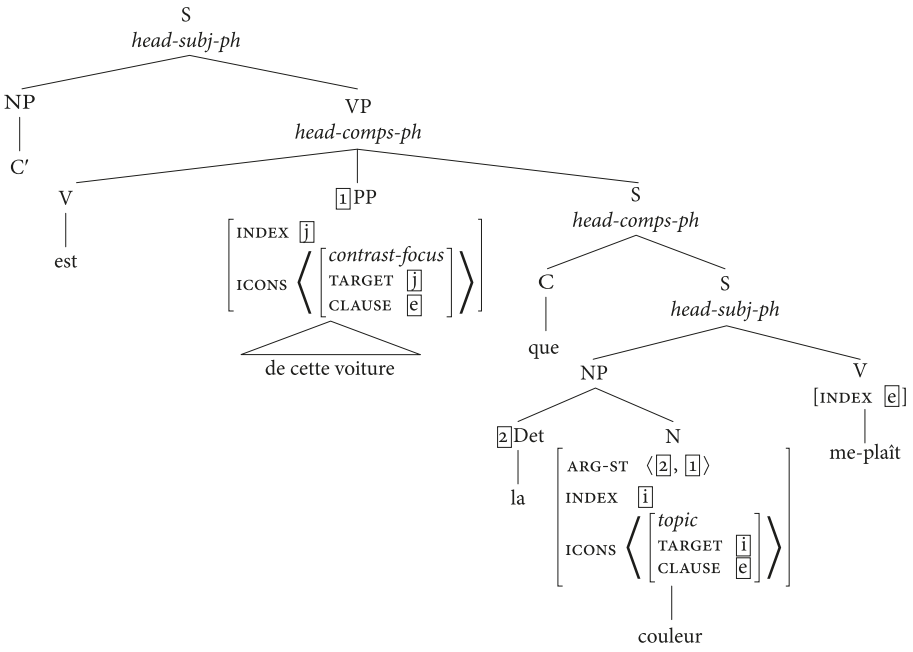

Figure 11. Simplified tree for (51)

For relative clauses, the discourse status of the antecedent with respect to the relative clause is introduced respectively in *head-comp-phrase* when the relative word is a complementizer, and in *head-filler-phrase* when the relative word is pronominal. The relative word does not introduce an *info-str* object in the clause's ICONS list. For subextraction out of an NP, the element in ARG-ST of the head noun is however the relative word – see Figure 12 – such that the constraint (49) cannot be violated. For this reason, the constraint does not apply to relative clauses.

This can explain why relative clauses are considered exceptional by some scholars: the reason is not that they are not "real" extractions constructions (Broekhuis, 2006), but the fact that their semantics and information structure is less constrained than those of interrogative clauses and *it*-clefts.[20]

20. As pointed out by a reviewer, the constraint in (49) does not reflect the gradient acceptability of extractions, and rules out any focalization of the complement of a non-focus subject. In Abeillé *et al.* (2020a), we assume a gradability of information status: an element is more or less focus, or more or less topic, and suggest to reformulate the FBC constraint in (17) as follows: "the more focused an element, the less acceptable it is as part of an unfocused / backgrounded constituent." We leave an HPSG implementation of such gradience for further work. See Chaves and Putnam (2020) for proposals in probabilistic SBCG.

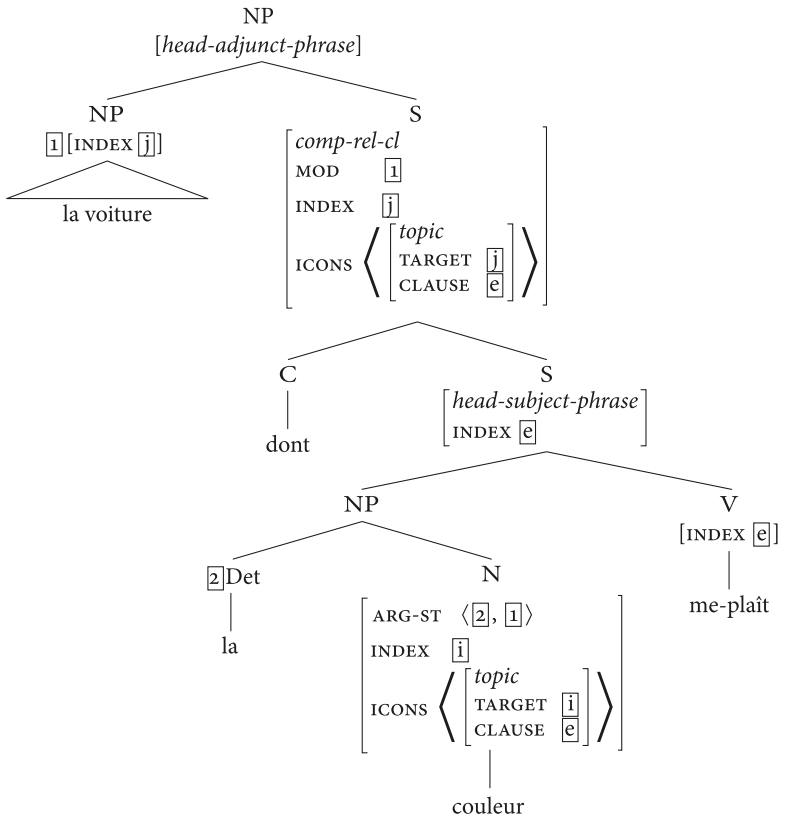

Figure 12. Simplified tree for (39a) (information structure)

6. Conclusion

Following Godard (1988), and contrary to some claims in Transformational Grammar (Tellier, 1991; Sportiche & Bellier, 1989; Heck, 2009), we review some corpora and experimental evidence showing that extraction from subject and objects are both acceptable in French relative clauses. Following Abeillé *et al.* (2020a, 2020b), we show that a difficulty to extract out of nominal subjects, compared with extraction out of nominal objects, may arise in *wh*-questions and *it*-clefts in French. In contemporary Head-Driven Phrase Structure Grammar, constraints on extraction are usually viewed as resulting from processing difficulties (Hofmeister & Sag, 2010; Sag, 2010), pragmatic infelicity or both (Chaves, 2013). Following Godard & Sag (1996) and Abeillé & Godard (2007), we present a syntactic analysis which places no syntactic constraint on extraction out of subjects. We follow Abeillé *et al.* (2020a), who propose a Focus-Background Conflict constraint, which takes into account the information structure of extraction

constructions, and constrains what can be focalized in *wh*-questions or *it*-clefts. We propose a formalization of this constraint in the HPSG framework and show how it accounts for the data in French and can be extended to other Romance languages as well.

Funding

This work is supported by the German Research Foundation (DFG) as part of the research program "Longdistance 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.

Acknowledgements

We thank the anonymous reviewers for their comments and suggestions, as well as Stefan Müller, Berthold Crysmann, Barbara Hemforth, Ted Gibson, Rui Chaves, Olivier Bonami, Gert Webelhuth, Bob Borsley, Jeff Runner, and Danièle Godard. We also thank the audience of the LSA 2019 islands workshop, the 2019 international HPSG conference, AMLAP 2019 Conference, Linguistic Evidence 2020, CUNY 2020.

References

Abeillé, A., Clément, L. & Liégeois, L. 2019. Un corpus annoté pour le français : le French Treebank. *TAL Traitement Automatique des Langues*, *60*(2), 19–43.

Abeillé, A. & Godard, D. 2007. Les relatives sans pronom relatif. In M. Abecassis, Ed., *Le francais parlé, Normes et variations*, p. 37–60. L'Harmattan.

Abeillé, A., Godard, D. & Sabio, F. 2008. Two types of NP preposing in French. In S. Müller, Ed., *The Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar*, p. 306–324, Stanford, CA: CSLI Publications.

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

Abeillé, A. & Godard, D. 1997. The syntax of French negative adverbs. In D. Forget, P. Hirschbuhler & M.-L. Rivero, Eds., *Negation and Polarity*. John Benjamins. https://doi.org/10.1075/cilt.155.02abe

Abeillé, A. & Godard, D. 2002. The syntactic structures of French auxiliaries. *Language*, *72*, 404–452. https://doi.org/10.1353/lan.2002.0145

Abeillé, A., Hemforth, B., Winckel, E. & Gibson, E. 2020a. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition. 204*, Article 104293.

Abeillé, A., Hemforth, B., Winckel, E. & Gibson, E. 2020b. Subject island: PP extraction depends on the construction. Poster at the 33rd Annual CUNY Sentence Processing Conference hosted virtually by the University of Massachusetts, Amherst. URL = https://osf.io/ndtvw/

Abeillé, A. & Winckel, E. 2020. French subject island? Empirical studies of *dont* and *de qui*. *Journal of French Language Studies*. 1–26. https://doi.org/10.1017/S0959269520000137

Aguila-Multner, G. & Crysmann, B. 2020. French clitic climbing as periphrasis. In G. Bîlbîie, Ed., *Interfaces in Romance: a constraint-based approach*. (This volume)

Ambridge, B. & Goldberg, A. 2008. The island status of clausal complements: evidence in favor of an information structure explanation. *Cognitive Linguistics*, *19*(3), 349–381. https://doi.org/10.1515/COGL.2008.014

Bianchi, V. & Chesi, C. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry*, *45*(4), 525–569. https://doi.org/10.1162/LING_a_00166

Bîlbîie, G. 2017. *Grammaire Des Constructions Elliptiques: Une étude comparative des phrases sans verbe en roumain et en français*. Language Science Press.

Bîlbîie, G. & Laurens, F. 2010. Towards a non-elliptical analysis of verbless relative adjuncts. *Bucharest Working Papers in Linguistics*, *1*, 51–67.

Bildhauer, F. 2008. Representing Information Structure in an HPSG Grammar of Spanish. Ph.d. thesis, Universität Bremen.

Bonami, O., Godard, D. & Marandin, J.-M. 1999. Constituency and Word Order in French Subject Inversion. In G. Bouma, E. Hinrichs, G.-J. M. Kruijff & R. Oehrle, Eds., *Constraints and Resources in Natural Language Syntax and Semantics*, p. 21–40. Stanford, CA: CSLI Publications.

Bouma, G., Malouf, R. & Sag, I. A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory*, *19*(1), 1–65. https://doi.org/10.1023/A:1006473306778

Branca-Rosoff, S., Fleury, S., Lefeuvre, F. & Pires, M. 2012. Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000). URL = http://cfpp2000.univ-paris3.fr/

Broekhuis, H. 2006. Extraction from subjects: some remarks on Chomsky's "On phases". In H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster, Eds., *Organizing Grammar*, p. 59–68. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110892994.59

Chaves, R. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics*, *49*(2), 297–344. https://doi.org/10.1017/S0022226712000357

Chaves, R. P. 2012. On the grammar of extraction and coordination. *Natural Language and Linguistic Theory*, *30*(2), 465–512. https://doi.org/10.1007/s11049-011-9164-y

Chaves, R. P. & Putnam, M. T. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford, UK: Oxford University Press.

Chomsky, N. 1973. Conditions on transformations. In S. Anderson & P. Kiparsky, Eds., *A festschrift for Morris Halle*, p. 232–285, New York: Winston.

Chomsky, N. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. 1995. *The Minimalist program*. Cambridge, MA: The MIT Press.

Cinque, G. 1990. *Types of Ā-dependencies*. Cambridge, MA: MIT Press.

Copestake, A. 2001. *Implementing Typed Feature Structure Grammars*. CSLI Lecture Notes. Stanford, CA: CSLI Publications.
Copestake, A., Flickinger, D., Pollard, C. & Sag, I. 2005. Minimal recursion semantics. *Research on Language and Computation*, *3*, 281–332. https://doi.org/10.1007/s11168-006-6327-9
De Kuthy, K. 2002. *Discontinuous NPs in German*. Stanford, CA: CSLI Publications.
De Kuthy, K. 2020. Information structure. In S. Müller, A. Abeillé, R.D. Borsley & J.-P. Koenig, Eds., *Head-Driven Phrase Structure Grammar: The handbook*, Empirically Oriented Theoretical Morphology and Syntax. Language Science Press. To appear.
Destruel, E. 2012. The French *c'est*-cleft: An empirical study on its meaning and use. In C. Piñon, Ed., *Empirical Issues in Syntax and Semantics 9 (Selected papers from CSSP 2011)*, p. 95–112. CSSP.
Diesing, M. 1992. *Indefinites*. Cambridge, MA: MIT Press.
Doetjes, J., Rebuschi, G. & Rialland, A. 2004. Cleft sentences. In F. Corblin & H.D. Swart, Eds., *Handbook of French semantics*, p. 529–552. Stanford, CA: CSLI Publications.
Engdahl, E. & Vallduví, E. 1996. Information packaging in HPSG. In C. Grover & E. Vallduví, Eds., *Studies in HPSG*, p. 1–31, Edinburgh: Centre for Cognitive Science.
Erteschik-Shir, N. 1973. On the nature of island constraints. Ph.D. thesis, MIT.
Gallego, Á. J. & Uriagereka, J. 2007. Sub-extraction from subjects: A phase theory account. In J. Camacho, Ed., *Romance linguistics 2006*, p. 149–162. Amsterdam and Philadelphia: John Benjamins.
Ginzburg, J. & Sag, I.A. 2000. *Interrogative Investigations: the form, meaning, and use of English Interrogatives*. Stanford, CA: CSLI Publications.
Godard, D. 1988. *La syntaxe des relatives en français*. Paris: Ed. du Centre national de la Recherche Scientifique.
Godard, D. 1992. Extraction out of NP in French. *Natural Language and Linguistic Theory*, *10*, 233–277. https://doi.org/10.1007/BF00133813
Godard, D. & Sag, I. 1996. Quels compléments de nom peut-on extraire en français? *Langages*, *122*, 60–79.
Goldberg, A. 2006. *Constructions at Work*. Oxford: Oxford University Press.
Goldberg, A. 2013. Backgrounded constituents cannot be extracted. In J. Sprouse & N. Hornstein, Eds., *Experimental Syntax and Island Effects*, p. 221–238. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139035309.012
Goldsmith, J. 1985. A principled exception to the coordinate structure constraint. In *Chicago Linguistic Society. Papers from the General Session at the Twenty-First Regional Meeting*, volume *1*, p. 133–143.
Haegeman, L., Jiménez-Fernández, A.L. & Radford, A. 2014. Deconstructing the Subject Condition in terms of cumulative constraint violation. *The Linguistic Review*, *31*, 73–150. https://doi.org/10.1515/tlr-2013-0022
Heck, F. 2009. On certain properties of pied-piping. *Linguistic Inquiry*, *40*(1), 75–111. https://doi.org/10.1162/ling.2009.40.1.75
Hofmeister, P. 2011. Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376–405. https://doi.org/10.1080/01690965.2010.492642
Hofmeister, P. & Sag, I. 2010. Cognitive constraints and island effects. *Language*, *86*(2), 366–415. https://doi.org/10.1353/lan.0.0223
Huang, C. 1982. Logical relations in Chinese and the theory of grammar. Ph.d. thesis, MIT.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jiménez-Fernández, A. 2009. On the composite nature of subject islands: A phase-based approach. *SKY Journal of Linguistics*, *22*, 91–138.

Kayne, R.S. & Pollock, J.-Y. 1978. Stylistic inversion, successive cyclicity and nove NP in French. *Linguistic Inquiry*, *9*(4), 595–621.

Kehler, A. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kluender, R. 1991. Cognitive constraint on variables in syntax. Ph.d. thesis, University of California, San Diego.

Kluender, R. 2004. Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher & A. Rodriguez, Eds., *Proceedings of the WCCFL*, volume *23*, p. 101–125, Somerville, MA: Cascadilla Press.

Kluender, R. & Kutas, M. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*, 573–633. https://doi.org/10.1080/01690969308407588

Kolliakou, D. 1999. De-phrase extractability and individual/property denotation. *Natural Language and Linguistic Theory*, *17*, 713–781. https://doi.org/10.1023/A:1006256028441

Krifka, M. 2007. Basic notions of information structure. In G.F. Caroline Féry & M. Krifka, Eds., *Interdisciplinary Studies on Information Structure (ISIS), Working Papers of the SFB 632*, volume *6*, p. 13–56, Potsdam: Universitätsverlag Potsdam.

Kuno, S. 1976. Subject, theme, and the speaker's empathy – a reexamination of relativization phenomena. In C.N. Li, Ed., *Subject and Topic*, p. 417–444, New York, NY: Academic Press.

Kuroda, S.-Y. 1976. Subject. In M. Shibatani, Ed., *Japanese Generative Grammar*, p. 1–16. New York: Academic Press. https://doi.org/10.1163/9789004368835_002

Ladusaw, W.A. 1994. Thetic and categorical, stage and individual, weak and strong. In M. Harvey & L. Santelmann, Eds., *Proceedings of SALT IV*, p. 220–229, Ithaca, NY: Cornell U. DMLL.

Lakoff, G. *et al.* 1986. Frame semantic control of the coordinate structure constraint. In *Chicago Linguistic Society. Papers from the General Session at the Twenty-Second Regional Meeting*, volume *2*, p. 152–167.

Lambrecht, K. 1994. *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Miller, P. & Monachesi, P. 2003. Les pronoms clitiques dans les langues romanes. In D. Godard, Ed., *Les langues romanes, problèmes de la phrase simple*, p. 53–106. Paris: CNRS Éditions.

Miller, P. & Sag, I.A. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, *15*, 573–639. https://doi.org/10.1023/A:1005815413834

Moreau, M.-L. 1971. *L'homme que je crois qui est venu – que*, *qui* : relatifs et conjonction. *Langue française*, *11*, 77–90. https://doi.org/10.3406/lfr.1971.5549

Moreau, M.-L. 1976. C'est : *étude de syntaxe transformationnelle*. Mons: Editions universitaires de Mons.

Müller, S. 2013. The CoreGram project: A brief overview and motivation. In D. Duchier & Y. Parmentier, Eds., *Proceedings of the Workshop on High-level Methodologies for Grammar Engineering (HMGE 2013), Düsseldorf*, p. 93–104.

Pollard, C. J. 1984. Generalized Phrase Structure Grammars, Head Grammars, and natural language. Ph.d. thesis, Stanford University.
Pollard, C. J. & Sag, I. A. 1994. *Head-Driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.
Rizzi, L. 1982. *Issues in Italian syntax*. Dordrecht, Holland and Cinnaminson, N.J., U.S.A.: Foris Publications. https://doi.org/10.1515/9783110883718
Ross, J. R. 1967. Constraints on variables in syntax. Ph.d. thesis, MIT Press, Cambridge, MA.
Sag, I. A. 1997. English Relative Clause Constructions. *Journal of Linguistics*, *33*(2), 431–484. https://doi.org/10.1017/S002222679700652X
Sag, I. A. 2010. English filler-gap constructions. *Language*, *86*(3), 486–545. https://doi.org/10.1353/lan.2010.0002
Sag, I. A. & Godard, D. 1994. Extraction of *de*-Phrases from the French NP. In *Proceedings of the North East Linguistic Society 24*, p. 519–541, Amherst, MA: GLSA.
Song, S. 2017. *Modeling Information Structure in a cross-linguistic perspective*. Berlin: Language Science Press.
Sportiche, D. & Bellier, P. 1989. Le mouvement syntaxique : contraintes et paramètres. *Langages, 95*, p. 35–80.
Sprouse, J., Caponigro, I., Greco, C. & Cecchetto, C. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory*, *34*(1), 307–344. https://doi.org/10.1007/s11049-015-9286-8
Stepanov, A. 2007. The end of CED? Minimalism and extraction domains. *Syntax*, *10*(1), 80–126. https://doi.org/10.1111/j.1467-9612.2007.00094.x
Szabolcsi, A. 2006. Strong versus weak islands. In M. Everaert & H. van Riemsdijk, Eds., *The Blackwell companion to syntax*, p. 480–531. Oxford: Blackwell. https://doi.org/10.1002/9780470996591.ch64
Takami, K. 1992. *Preposition Stranding: From Syntactic to Functional Analyses*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110870398
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*. Dordrecht, The Netherlands and Boston: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-3596-2
Torrego, E. 1985. *On Empty Categories in Nominals*. unpublished ms., University of Massachusetts, Boston.
Uriagereka, J. 1988. On government. Ph.d. thesis, University of Connecticut. URL = https://opencommons.uconn.edu/dissertations/AAI8905987
Uriagereka, Juan. 2012. Spell-Out and the Minimalist Program. Oxford: Oxford Scholarship Online.
van Valin, R. D. J. 1995. Toward a functionalist account of so-called extraction constraints. In B. Devriendt, L. Goossens & J. van der Auwera, Eds., *Complex structures: A functionalist perspective*, p. 26–60. Berlin: Mouton de Gruyter.
Webelhuth, G. 2007. Complex Topic-Comment Structures in HPSG. In S. Müller, Ed., *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, p. 306–322, Stanford, CA: CSLI Publications.
Zaenen, A. 1983. On Syntactic Binding. *Linguistic Inquiry*, *14*, 469–504.

Address for correspondence

Elodie Winckel
Humboldt-Universität zu Berlin
Inst. für deutsche Sprache und Linguistik
Dorotheenstr. 24
10117 Berlin
Germany

Co-author information

Anne Abeillé
LLF, Université de Paris