# Question particles *ça* and *donc* in French: A corpus study*

**Anastasiia Smirnova** ** · **Anne Abeillé** ***
**(LLF, Université de Paris)**

**Smirnova, Anastasiia and Anne Abeillé. 2021. Question particles *ça* and *donc* in French: A corpus study.** *Linguistic Research* 38(2): 239-269. We study two question particles (*donc* and *ça*) in a large French corpus. We show that they are specialized for direct questions, and that wh-*ça* is possible both *in situ* and in fronted position, contrary to previous literature (Cheng and Rooryck 2001). We also show that ellipsis is the most common use of wh-*ça*, with sluicing (Ross 1969) and a new construction that we call *wh-stripping*, while fronting is the most common use of wh-*donc*. In many examples, a non-elliptical question would be impossible or quite different, which seem to cast doubt on a deletion under identity approach of sluices (Ross 1969) and to favor a Direct Interpretation approach (Ginzburg and Sag 2000). Looking at the preceding context, in about half of wh-*ça* and wh-*donc* sluices, the wh-word has a correlate, and in the majority of cases, the correlate is definite. We suggest that the particle *ça* favors elliptical reprise questions, compared to bare wh-words, while the particle *donc* favors biased questions. **(LLF, Université de Paris)**

**Keywords** French, corpus study, wh-question, discourse particle, ellipsis, sluicing

## 1. Introduction

Discourse particles in wh-questions have gained recent attention and their role seems to go beyond emphasis and emotion. They have mostly been studied inside full verbal questions (den Dikken and Giannakidou 2002; Bayer and Obenauer 2011). We focus here on two French particles, *ça* and *donc*, and on their use in elliptical questions (2) (Ross 1969).[1]

---

1 Dutch dialects also have a demonstrative occurring with wh-words in sluices:
   (i) Jef   eid  iemand    gezien,  mo ik  weet  nie   wou da
       Jeff has  someone    seen     but I  know  not   who that
       'Jeff saw someone, but I don't know who.' (Van Craenenbroeck 2010: 13)

Like *the hell* (1) (Rawlins 2009), both discourse particles seem to denote ignorance questions, and are not felicitous with resolved questions, unlike bare wh-words:

(1) a. I don't know who (the hell) would buy that book.
   b. I know who (*the hell) would buy that book. (den Dikken and Giannakidou 2002: 33)

(2) a. Quelqu'un est venu, je me    demande qui   (ça / donc).
   Someone  is   come  I  REFL ask      who (ça / donc)
   'Someone came, I wonder who.'
   b. Quelqu'un est venu et je sais  bien qui  (#ça / #donc).
   Someone  is   come and I   know well  who (#ça / #donc)
   'Someone came and I know who.'

Unlike the connective *donc* ('then') and the pronoun *ça (*'that'*),* the discourse particles *ça* and *donc* are not mobile and are attached to the wh-word, like *the hell* in English.

(3) a. Qui  donc est  venu? ≠ Qui  est donc venu?
   who  donc is   come  ≠ who is donc come?
   'Who actually came?' ≠ 'Who came then?'
   b. Tu  as    vu   qui ça à Paris?
   you have seen who ça  in Paris?
   'Who have you seen in Paris?'
   c. *Tu  as    vu   qui à Paris ça?
    you have seen who in Paris ça?
   'Who have you seen in Paris?'

Unlike the English particle *then* (4a) (Pranav, Hardt, and McCloskey 2020), French *donc* does not presuppose a negative proposition (2a) and (4b).

(4) a. "But that doesn't mean all the time," Balanchine chided her. How many times then? (Santa Cruz database; [152316])
   b. La  porte est  ouverte. Qui  donc est  entré?
   the door is   opened. Who donc is   entered?

'The door is open. Who came?'

From a syntactic point of view, wh-*ça* and wh-*donc* behave differently from the English particle *the hell* (5) (Pesetsky 1987; Ginzburg and Sag 2000; den Dikken and Giannakidou 2002; Huang and Ochi 2004), since the wh-word with both particles can be in situ in French (3b) and (6b).

(5) a. What (the hell) did John buy?
    b. John bought what (*the hell)?

(6) a. Qui   donc as-tu      vu    cet    après-midi?
       who donc have-you seen this   afternoon
       'Who have you met this afternoon?'
    b. Tu    as    vu    qui   donc cet après-midi?
       you   have seen who  donc this afternoon
       'Who have you met this afternoon?'

Obenauer (1976), Cheng and Rooryck (2001) claim that wh-*ça* must be *in situ*, unlike wh-*donc* (6a).

(7) a. Tu as    vu    qui  ça (cet après-midi)?
       you have  seen who ça (this afternoon)?
       'You have met who (this afternoon)?' (Cheng and Rooryck 2001: 16)
    b. *Qui  ça as-tu      vu (cet après-midi)?
        who ça have-you seen (this   afternoon)?
        'Who have you met (this afternoon)?' (Cheng and Rooryck 2001: 16)

If this were true, this would be a counterargument against deletion-based analyses of elliptical questions (sluices), which derive them from fronted wh-questions (Ross 1969; Merchant 2001), since *ça* can occur in sluices (2a). But this has never been tested empirically (Gibson and Fedorenko 2010).

We thus conducted a corpus study on the two particles *ça* and *donc* in French wh-questions, using the contemporary part of the Frantext database (*www.frantext.fr*). This paper can be seen as a contribution to the empirical turn in the study of ellipsis (see Pranav, Hardt and McCloskey (2020) for a corpus study of English sluices based

on the *New York Times*; Kim and Abeillé (2019) for a corpus study of *why*-stripping, and Kim and Kim (2021) for a corpus study of wh-*the hell* questions, based on the COCA, and Miller and Pullum (2014) for a study of exophoric VP ellipsis based on the COCA).

## 2. Previous corpus studies

We briefly review Lefeuvre's (2018) study on French literature after 1700, and Lefeuvre's (2020) study of a smaller corpus of spoken French, before turning to Reinhardt's (2019) corpus of questions in ten French contemporary detective novels. Then we present our own data, using French literature after 1960.

### 2.1 Lefeuvre (2018) and (2020)

Lefeuvre (2018) conducted a corpus study of verbless questions in French literature (Frantext 1700-2010) and found 780 occurrences of wh-*ça?*, with a higher proportion of *comment ça, où ça, qui ça* and *pourquoi ça* (Table 1).[2]

Table 1. Lefeuvre (2018) corpus of verbless wh-*ça?*

| | |
|---|---|
| *Qui ça*? ('who-ça?') | 160 |
| *Quoi ça*? ('what-ça?') | 9 |
| *Quand ça*? ('when-ça?') | 33 |
| *Où ça*? ('where-ça?') | 252 |
| *Pourquoi ça*? ('why-ça?') | 151 |
| *Comment ça*? ('how-ça?') | 175 |
| **Total** | 780 |

She does not compare her data with verbal questions. According to her, the main role of *ça* is to recall an element of the preceding clause and to request information about this element. Thus wh-*ça* often has an antecedent, which can be a personal pronoun (*il* 'he' in (8)) or an NP (*une dame* 'a lady' in (9)).

---

2   Lefeuvre also considered wh-word + *cela*, which we consider a hypercorrected form of *ça*, and which we ignore here.

(8) −Il  était  là,  l'autre?      −Qu... qui  ça? −Le  cuisinier...

    he was  there DEF-other?  Wh  who  ça?    the  cook

    '−Was  the  other  one  there? −Who.. who? −The  cook...' (Gavalda 2004)

(9) −Il y    a  là    une dame qui  demande Monsieur. (...)

    it there has there a    lady  who  asks      Sir

 −Moi!  Qui  ça?

  me    who  ça?

 '−There is a lady there who asks for Sir. (...) −That's me! Who?' (Feydeau 1896)'

Some of these questions with *ça* are reduced to the wh-word (8) and (9), and some have the wh-word with *ça* followed by a previously mentioned segment:

(10)−Ben... et  nous? glapit   Attila.

    ben   and us       screams Attila

 −Qui  ça « nous »?

  who ça «  us   »

 '−Well... and us? screams Attila. −Who «us»? (Buron [de], 1998)

In Lefeuvre (2020), she annotated a smaller spoken corpus ESLO2 (Baude and Duga 2016). She extracted 699 wh-questions from a subcorpus of informal spontaneous speech (meal conversation) and found that 207 (30%) wh-questions were verbless (11a), among which 8% were with *ça* (11b):

(11)a. On devrait plus  avoir à  payer de    taxes alors. −Pourquoi?

    one should more have to pay   INDF taxes then    why

    'One should no longer have to pay taxes then. −Why?' (ESLO2, meal conversation)

  b. Bah je l'   ai   vue.  −Quand ça?

    bah  I  her  have seen    when ça

    'Well I've seen her. −When?' (ESLO2, meal conversation)

She suggests that verbless questions with *ça* are not elliptical but headed by the wh-element with *ça* as a pronominal subject, like *Pourquoi ce bruit? Pourquoi cela?*

('Why this noise? Why this?') (Laurens 2008). The analysis of *ça* as a pronoun is not possible when wh-*ça* is in situ as in (7a). It is also dubious in (9), since a nominal subject is infelicitous (??*Qui une dame?* 'Who a lady?'). Furthermore, if *ça* is the subject of the wh-word, another anaphoric pronoun should not be possible after *ça* as in (10).

We thus consider *ça* as a discourse particle, and (8), (9) and (11b) are thus short questions or sluices.[3] As for (10), we call it *wh-stripping*, since it resembles what Merchant (2011) called *why-stripping*, when the wh-word is followed by a previously mentioned segment: *He left to Rome – Why Rome?*.

Since Lefeuvre's data were not annotated, and may reflect different uses because of different time periods, we turn to a more contemporary corpus.

## 2.2  Reinhardt (2019)

Reinhardt (2019) conducted a corpus study of direct wh-questions in ten XXIst century detective novels from 6 different French authors (Izzo, Lemaitre, Levy, etc), representing non-formal uses. The corpus was coded with several parameters (*est-ce que*, subject inversion, etc). From her data (*https:// hdl.handle.net/11403/interrogatives-in-novels/v1*), we extracted 2282 wh-questions with 7 wh-words of interest (Table 2), disregarding *quel/ lequel/ que* (which cannot occur with discourse particles) and *est-ce que* questions. These data include *qui* ('who') and *quoi* ('what') preceded by a preposition, but not infinitival clauses, and wh-words followed by the discourse particles *ça* and *donc*.

In her data, we observe that fronted (extracted or subject) position represents half of wh-questions (52%), followed by the *in situ* use (26%), except for 3 wh-words: *quoi* for which *in situ* is the most frequent use, followed by ellipsis, *comment* and *pourquoi* for which ellipsis is more frequent than *in situ* use. Overall the ellipsis rate is (17%), which is lower than in Lefeuvre 2020's spoken corpus (section 2.1).

Table 2. Data from Reinhardt's (2019) written corpus

|  | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | *combien* 'how many' | total |
|---|---|---|---|---|---|---|---|---|
| fronted | 84 | 245 | 177 | 50 | 316 | 271 | 63 | 1206 (52%) |
| in situ | 301 | 119 | 76 | 28 | 8 | 46 | 31 | 609 (26%) |
| ellipsis | 132 | 72 | 27 | 14 | 159 | 55 | 8 | 467 (17%) |

---

3   We use 'sluicing' as in Ross (1969): a question reduced to a wh-word or wh-phrase.

| total | 517 | 436 | 280 | 92 | 483 | 372 | 102 | 2282 |
|---|---|---|---|---|---|---|---|---|

We took out the pronominal uses of *ça*, as in *C'est quoi, ça?* ('It is what-ça?'), and found 49 wh-*ça* questions, which all are elliptical:

- Sluicing: 14 *qui ça* (12a), 7 *où ça*, 22 *comment ça* and 3 *pourquoi ça*
- Wh-stripping:1 *où ça*, 1 *comment ça* (12b), 1 *pourquoi ça*

(12) a. Ahoui, qui   ça?

   ah yes who  ça?

   'Ah yes, who that?' (Lévy 2000)

   b. Comment ça,  pas dans une rue?

   how       ça not in   a    street?

   'How come not in the street?' (Bussi 2015)

In her data, we also found 8 wh-*donc*, all in matrix sluices: 3 *qui donc*, 2 *quoi donc*, 3 *pourquoi donc* (13).

(13) Et    pourquoi donc?

   and  why        donc

   'And  why  then?' (Lévy 2000)

It seems that both discourse particles favor ellipsis, and sluicing in particular (93% of wh-*ça* and 100 % of wh-*donc* are sluices). This could be explained by the fact that such particles play a role in discourse cohesion, and that ellipsis relies on discourse context more heavily than non-elliptical utterances (Ginzburg and Sag 2000; Kehler 2002). In order to test this hypothesis, we now turn to a larger corpus, and will explore further differences between wh-*ça* and wh-*donc* and compare them with bare wh-words.

## 3.  Our corpus of contemporary French

We focus on contemporary (written) French, using the Frantext database (*www.frantext.fr*) after 1960 (1001 texts, 61.2 M words), searching for a wh-word + *ça*, wh-word + *donc*, and the same wh-words followed by?. *Ça* and *donc* cannot occur with

a weak pronoun (*Que ça? 'What-ça?', *Que donc? 'What-donc?') nor with a determiner (*Quel ça homme? 'Which-ça man?', *Quel donc homme? 'Which-donc man?').

Like *the hell* (Pesetsky 1987), *ça* and *donc* cannot modify an anaphoric pronoun (*lequel* 'which'):

(14) a. Mary kissed one of the boys tonight, but I have no idea which (*the hell).

　　b. Marie a 　embrassé un des 　　garçons. *Lequel ça?

　　　Marie has kissed 　one of-DEF 　boys. 　　which ça?

　　　'Mary kissed one of the boys. Which one?'

We thus searched for 7 wh-words: *combien* ('how many'), *comment* ('how'), *pourquoi* ('why'), *quand* ('when'), *qui* ('who'), *quoi* ('what') and *où* ('where'), followed by *ça, donc* or? (Table 3).

We included infinitival clauses (*Pourquoi donc continuer à vivre?* 'Why-donc go on living?'), but excluded cases where the wh-word is preceded by a preposition (*d'où* 'from where', *jusqu'à quand* 'until when', *pour qui* 'for whom', etc) because we did not find any with *ça*. Moreover, we leave aside exclamative uses (*Comment ça !* 'How-ça!', *Quoi donc !* 'What-ça!').[4]

Table 3. Wh+*ça/ donc/ ?* in Frantext (1960-2010)

| | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | *combien* 'how many' | total |
|---|---|---|---|---|---|---|---|---|
| + *ça* | 3 | 91 | 116 | 13 | 23 | 235 | 0 | 481 |
| + *donc* | 106 | 237 | 152 | 88 | 198 | 35 | 3 | 819 |
| +*?* | 4566 | 1379 | 939 | 570 | 2474 | 850 | 254 | 11032 |
| total | 4675 | 1707 | 1207 | 671 | 2695 | 1120 | 257 | 12332 |

We observe that the most common bare wh-words (in final position) are *quoi, pourquoi* and *qui*. They were also the most common wh-words (all uses) in Reinhardt's corpus (Table 2).

We also observe that the two discourse particles are not used with the same wh-words: while *ça* is mostly used with *comment* ('how'), *où* ('where') and *qui* ('who') (as in Reinhardt's corpus), *donc* is mostly used with *qui* ('who'), *pourquoi* ('why'), *où* ('where') and *quoi* ('what'). They are never (or almost never) used with *combien* ('how

---

4　See Ginzburg (2019) for a corpus study of English exclamative sluices.

many'), as in Lefeuvre's and Reinhardt's corpora.

## 4.  A new corpus study of wh-*ça*

In our corpus, we found 481 occurrences of wh-*ça* (Table 3), with a higher proportion of *comment ça* ('how-ça') and *où ça* ('where-ça'), and a lower proportion of *quoi ça* ('what-ça') and *pourquoi ça* ('why-ça'), and no *combien ça* ('how many-ça').

Then, we annotate a large sample of wh-*ça* for subject/extraction/in situ/ellipsis, taking all occurrences but only a random sample of 104 for *comment ça* ('how ça') and of 100 for *où ça* ('where ça'). Overall, we annotated 334 occurrences of wh-*ça*, among which 329 are elliptical (Table 4). We observe that ellipsis (sluicing and wh-stripping) is the main use of wh-ça questions (98.5%).

Table 4. Annotated uses of wh-*ça* in Frantext (1960-)[5]

| + *ça* | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| subject | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| extracted | 0 | 0 | 1 | 0 | 1 | 1 | 3 |
| in situ | 0 | 0 | 1 | 1 | 0 | 0 | 2 |
| sluicing | 3 | 65 | 86 | 11 | 21 | 38 | 224 (67%) |
| stripping | 0 | 26 | 12 | 1 | 1 | 65 | 105 (31%) |
| total | 3 | 91 | 100 | 13 | 23 | 104 | 334 (100%) |

### 4.1 Non elliptical uses of wh-ça

Contrary to previous literature (Obenauer 1976; Cheng and Rooryck 2001), we found 3 examples of wh-*ça* extracted (15), both without subject inversion, among which two have a nonstandard *que* ('that') insertion (15ac).[6]

(15) a. Où    ça qu' ils  vont ces  deux copains-là?
         where ça that they go   these two  pals     there?
         'Where are going those two pals?' (Sabatier 1985)

---

5   Our data are available on the website *http://www.llf.cnrs.fr/ressources.*

6   (15c) is a case of reason-*comment* which expresses surprise for an unexpected event (Desmets and Gautier 2009; Brunetti, Tovena, and Yoo 2020).

    b. Et  pourquoi ça vous feriez      pendre mon  papa, monsieur le   duc?
       and why          ça you  do-COND hang  my    dad    sir        the duke?
       'And why  would  you  hang  my  dad, Mr Duke?' (Queneau 1965)
    c. Comment ça, s'il te plaît, que  je ne     pourrais   pas le savoir?
       how         ça please       that I  NEG  can-COND  NEG it  know?
       'Excuse me, how  I  could  not  know it?' (Koltès 1985)

We  suggest  that  the  unacceptability  of  (7b)  may  be  due  to  clitic  subject  suffixation
(Huot  1986;  Miller  1991),  and  that  nominal  subject  inversion  may  also  block  wh-*ça*
fronting  (16).

   (16) a. Où      ça ils    vont ces     deux copains-là?
       where ça they  go    these two    pals     there?
       'Where  are  going  those  two  pals?'
    b. #Où     ça vont  ces    deux copains-là?
       where ça go     these two pals     there?
       'Where  are  those  two  pals  going?'

We  suggest  that  this  may  be  due  to  a  register  clash,  since  clitic  suffixation  or  subject
inversion  is  a  mark  of  formal  interrogative  clauses  in  French  (Thiberge  2020),  whereas
wh-*ça*  belongs  to  an  informal  register.
    We  also  found  two  *in situ*  examples  (17):

   (17) a. −Mon  papa il    est mort. −Ton  papa il est  mort, quandça?
        my    dad he  is dead    your dad  he is   dead  when ça?
       '−My  dad died. −When  did  your  dad die?' (Mauriac 1961)
    b. T'     avais appris  à tirer  où     ça?
       You had    learned  to shoot  where ça?
       'Where  had  you  learned  to shoot?' (Boudard 1982)

We  conclude  that  wh-*ça*  cannot  be  ruled  out  with  extraction,  contrary  to  previous
literature,  and  that  it  is  not  specialized  for  *in situ*  uses.

## 4.2 Elliptical uses of wh-ça

As for ellipsis, we found matrix sluices, especially in dialogs, but also in a few monologs (18b), as well as 1 embedded sluice, which is a case of indirect speech (18a).

(18) a. Elle me dit,   vous  savez, je vous  connaissais avant   de vous rencontrer.
        She me says  you   know I you   knew       before  of  you  meet
        Étonné,   je demande  comment   ça?
        surprised I ask       how         ça
        'She says, you know, I knew you before I met you. Surprised, I ask how come?' (Doubrovsky 2011)

  b. Ma femme?  Quelle femme?  Où     ça, ma femme?
        my wife?       which wife?     where ça  my wife
        'My wife? Which wife? Where my wife?' (Pennac 1995)

Root sluicing is the predominant use of wh-*ça* (67% in Table 4); it is almost 100% with *pourquoi ça*, *quoi ça* and *quand ça*; 86% with *où ça* ('where ça') and 71% with *qui ça (*'who ça').

Contrary to (Merchant 2001: 110), who claims that preposition omission is not possible in French sluices, since French is not a preposition stranding language, we found 2 examples of preposition omission (*de* 'of'):

(19) a. −L'   adresse  de  Rosine Portinari, tu  l'  as      pas? −Qui ça?
          the   address of  Rosine Portinari you it have  NEG   who ça
        'You don't have the address of Rosine Portinari? −Who?' (Thérame 1985)

  b. « La ville de Jaufré Rudel ! » « Qui   ça? »
          the city  of Jaufré Rudel        who ça
        '« The city of Jaufré Rudel! » « Who? »' (Garat 1984)

The interpretation of the sluice requires contextual information. We thus annotate whether the wh-word has an explicit correlate (*quelqu'un* 'someone') (20), an implicit correlate (21a) or no correlate (21b) (*sprouting*) (Ross 1969). The results are illustrated in Table 5.

(20) Quelqu'un me l' a    dit. − Qui  ça?
　　　someone  me it has said  who  ça?
　　　'Someone told me that. − Who?' (Vargas 2008)

(21) a. Bien sûr,   je copie.  − Qui ça?  − Damascius.
　　　　of   course I  copy     who ça?    Damascius
　　　　'Of course, I copy. − Who? − Damascius.' (Roubaud 2002)

　　　b. Mlle Dreyfus ne  travaille plus    ici.  Elle  nous a   quittés.
　　　　Ms  Dreyfus NEG work      anymore here she   us    has left
　　　　− Quand  ça?
　　　　　when   ça?
　　　　'Ms. Dreyfus does not work here anymore. She has left us. − When?'
　　　　(Gary 1974)

Table 5. Correlates of wh-sluices + *ça*

| + *ça* | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| definite correlate | 3 | 52 | 21 | 2 | 0 | 2 | 80 (35%) |
| indefinite correlate | 0 | 10 | 4 | 2 | 0 | 1 | 17 (8%) |
| implicit correlate | 0 | 2 | 0 | 0 | 0 | 0 | 2 (1%) |
| no correlate | 0 | 1 | 61 | 7 | 21 | 35 | 125 (56%) |
| total sluicing | 3 | 65 | 86 | 11 | 21 | 38 | 224 (100%) |

Almost half (44%) of wh-*ça* sluices have an antecedent with a correlate for the wh-word. 100% occurrences of *quoi ça* have a correlate, while there is none for *pourquoi ça* ('why-ça') and only 8% for *comment ça* ('how-ça'), which is mostly used in *sprouting* (Ross 1969). Following Ginzburg and Sag (2000), we distinguish between direct-sluices (22a), which are information questions ('Who came?'), and reprise sluices (22b), which are clarification questions ('Who do you mean by Jo?').

(22) a. − Someone came.    − Who?
　　　b. − Jo        phoned − WHO? (Ginzburg and Sag 2000: 334)

In direct sluices, the wh-word has an indefinite or implicit correlate: *on* ('one'), *quelqu'un* ('someone'), *ailleurs* ('somewhere else'), *quelque part* ('somewhere'), *une fois* ('once')...) (23), while in reprise sluices it usually has a definite correlate (a personal

pronoun, a proper name, or a definite NP): *Jean-Paul Sartre, à Paris* ('in Paris') (24). Most wh-*ça* sluices with a correlate have a definite correlate, and are thus reprise sluices (81%): 100% for *quoi ça*, 83% for *qui ça*.

(23) a. J'ai     dit   que Thérèse était ailleurs. − Où     ça? a     demandé Le Petit.
         I have said that Thérèse was elsewhere  where ça? has asked     Le Petit
         'I said that Thérèse was somewhere else. − Where? asked Le Petit.'
         (Pennac 1999)

     b. Vous avez habité ailleurs? − Oui, une fois... − Quand ça?
         you   have lived  elsewhere yes   once          when ça?
         'Have you lived elsewhere? − Yes, once... − When?' (Le Clézio 1963)

(24) a. Quand il   était au     Havre, Jean-Paul Sartre s'   y      exerçait    bien,
         when   he was in-DEF Havre Jean-Paul Sartre REFL there practiced  well
         à   la   boxe. [...] − Qui ça?
         in the boxing        who ça?
         'When he was in Le Havre, Jean-Paul Sartre was practicing boxing. [...]
         − Who?' (Guyard 2011)

     b. Et vous ne   voyez pas   quelque chose  comme un dissident ou deux,
         and you NEG see    NEG something   like     a  dissident or  two
         à  Paris ? − Où    ça?
         in Paris     where ça?
         'And don't you see something like a dissident or two in Paris? − Where?'
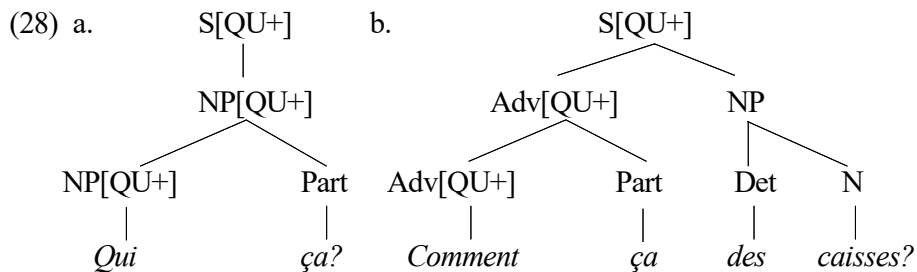         (Kristeva 1990)

Sometimes, a verbal question would be possible with wh-*ça*: for example in (21b), *Elle vous a quittés quand ça*? ('When did she leave you?') or in (23a) *Thérèse était où ça*? ('Where was Thérèse?'), but not always: *\*Qui ça vous l'a dit*? ('Who told you that?') in (20). Most of the time, its form would be quite different from that of the preceding utterance: for example in (24b), the dialog pronoun *vous* ('you') should be switched to *je* ('I'), and the negation should be dropped: *Et je verrais un dissident ou deux où ça*? (And I would see a dissident or two where ça?). In (18b) there is no corresponding preceding verbal sentence (the speaker does not have a wife). More generally, the unavailability of an identical verbal question casts doubt on the derivation of wh-*ça* sluices from non-elliptical sources.

We also found cases of what we call *wh-stripping* (section 2.1). Contrary to Yoshida et al. (2015), who claim that stripping (with a non-interrogative remnant) is limited to *why* across languages, we found that wh-stripping is possible with 5 of our wh-words (Table 4), and very common with *comment ça* ('how-ça'). Like reprise sluices, wh-stripping questions the form which has just been used. The reprise fragment can be of any category: an adjective (25a), a noun, a preposition (25b), even a weak pronoun (26a), an imperative (26b) or a whole sentence:

(25) a. C'est curieux, [...].  Pourquoi ça, curieux?
        it is curious          why      ça curious?
        'It's curious, [...]. Why curious?' (Del Castillo 1981)
    b. Avant, au moins, il y    avait les mariages[...]. −Avant? Quand ça, avant?
        before at least  it there had the marriages     before? when ça before?
        'Before, at least, there were marriages [...]. −Before? When, before?')
        (Sollers 1987)
(26) a. Ils   ont  arrêté   Paul. [...] −Qui ça, ils?
        they have arrested Paul         who ça they?
        'They have arrested Paul. [...] −Who they?' (Clavel 1968)
    b. Laisse, dis- je.  −Comment ça, laisse?
        leave   said I     how       ça leave?
        'Let it go, I said. −How come, let it go?' (Pennac 1995)

Dialog pronouns can be switched (*vous – je* 'you – I') (27a) or not (27b). Notice also the change of interrogative form in (27a) (a *si* ('whether')-question to reprise a questioning declarative) and the indefinite change in (27c).[7]

(27) a. Vous croyez? dit  le  type,  −Comment ça si je crois?
        you  believe say the guy     how      ça if I  believe?
        'You think so? says the guy, −What do you mean by I think so?'
        (Benoziglio 1974)
    b. B: Qui  est là?   D: C'est moi, c' est moi.  B: Qui ça, moi?
          who is  there    it is me    it is me        who ça me?

---

7   Most cases of wh-stripping with *comment ça* ('how ça') (25b, 26) have a reason interpretation ('how come')
    (Brunetti, Tovena, and Yoo 2020).

'B: Who's there? D: It's me, it's me. B: Who me?' (Ionesco 1963)

c. Elles ont   emporté un certain nombre de   caisses. – Quoi!   Comment
   they have taken    a   certain number of   boxes       what   how
   ça, des   caisses?
   ça  INDF boxes?
   'They took  a   certain number of  boxes. – What!  How come the boxes?'
   (Garat 2010)

Therefore, we conclude that matrix sluice is the main use of wh-*ça* (67%), followed by wh-stripping (31%).

In most cases of wh-stripping, a non-elliptical counterpart would not be felicitous. When a verbal question is possible, its form be quite different from that of the preceding utterance: for example in (26b), a modal should be added: *Comment ça vous voudriez que je laisse?* ('How would you want me to let it go?'), while in (26a), the weak pronoun should be turned into a strong pronoun with a copula and a relative maker added: *Qui ça sont ceux qui ont arrêté Paul?* ('Who are those who have arrested Paul?').

We thus suggest an analysis with a Direct interpretation approach (Ginzburg and Sag 2000; Culicover and Jackendoff 2005): the content of the sluice is inferred from that of the antecedent clause, and there is no need to posit a syntactic structure for the unpronounced material (i.e., at the ellipsis site) and a wh-phrase in sluicing can thus be treated as a categorial phrase projection of the wh-expression itself. In this approach, any category can be projected into an elliptical clause when it functions as a focus establishing constituent (FEC), and the resulting clause inherits the rest of its interpretation from the question under discussion (QUD). In HPSG terms, we use a Head-fragment-phrase (Ginzburg and Sag 2000), which can be unary (sluicing) or binary (stripping) (Kim 2017; Kim and Abeillé 2019; Kim and Nykiel 2020). The particle is adjoined to the wh-word (Kim and Kim 2021).

(28) a.            S[QU+]        b.                S[QU+]

             NP[QU+]                    Adv[QU+]          NP

      NP[QU+]        Part       Adv[QU+]      Part     Det     N

        *Qui*          *ça?*       *Comment*      *ça*    *des*   *caisses?*

## 5.  A corpus study of wh-*donc*

In our corpus, we found 819 occurrences of wh-*donc* (Table 3), with a higher proportion of *qui donc* ('how-donc'), *pourquoi donc* ('why-donc') and *où donc* ('where-donc'), and a lower proportion of *comment donc* ('how-donc'), and *combien donc* ('how many-donc').

We use the same type of annotation for the wh-*donc* subcorpus as for the wh-*ça* one (subject, extraction, in situ, ellipsis), taking a representative sample, with all *quand donc* ('when-donc') and *comment donc* ('how-donc'), but reducing the number of other wh-*words* to a random 100. Overall, we annotated 522 occurrences of wh-*donc* (Table 6).

Table 6. Annotated uses of wh-*donc* in Frantext (1960-2010)

| + *donc* | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| subject | 4 | 81 | 0 | 0 | 0 | 0 | 85 (16%) |
| extracted | 5 | 9 | 79 | 83 | 55 | 27 | 258 (49%) |
| in situ | 3 | 0 | 0 | 1 | 0 | 0 | 4 |
| sluicing | 88 | 10 | 20 | 4 | 45 | 7 | 174 (33%) |
| stripping | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| total | 100 | 100 | 100 | 88 | 100 | 34 | 522 (100%) |

Let us now look at their distribution. We can see that, unlike wh-*ça*, ellipsis is not the most common use with wh-*donc,* but nevertheless represents one third of all the annotated wh-*donc* occurrences (33%).

## 5.1 Non elliptical uses of wh-*donc*

Most cases of wh-*donc* are fronted (subject or extracted) (65%): the extracted cases represent 94% for *quand donc (*'when-donc'), 79% for *où donc (*'where-donc'), 79% for *comment donc (*'how-donc*')*, 55% for *pourquoi donc* ('why-donc'). However, for *quoi donc (*'what-donc'), the most frequent use is elliptical (88%).

The vast majority of *qui donc* ('who-donc') are in subject position (81%), which under certain analyses can be considered as a case of extraction (29). The example in (29b) illustrates a nonstandard *que* ('that') insertion, as it was with wh-*ça* (15a,c).

(29) a. Qui donc  avait  parlé?

   whodonc  had talked

   'Who  had  talked?'  (Kessel  1967)

   b. Qui donc  qu'  est  malade?  C'est  vous, madame Joffrette?

   whodonc  that is  ill      it is  you  missis  Joffrette

   'Who  is  ill?  Is  it  you,  Mrs  Joffrette?'  (Fallet  1973)

We observe that *quoi donc* ('what donc') can also be a subject (30a), while bare *quoi* ('what') cannot (30b) (Abeillé and Godard 2008).

(30) a. Quoi donc m' occuperait    mieux dans le   désert  de  ma  vie?

   what donc me occupy-COND better in    the desert  of  my  life

   'What  would  occupy  me  better  in  the  desert  of  my  life?'  (Duras 1977)

   b. *Quoi  m' occuperait    mieux dans le   désert  de ma vie?

    what  me occupy-COND better in    the  desert  of  my  life

   'What  would  occupy  me  better  in  the  desert  of  my  life?'

We also found many wh-*donc* examples with an infinitival verb (31), while there is no example like this with wh-*ça*. They often are self-addressed or deliberation questions, and (31a) includes the answer (after *sinon* 'unless).

(31) a. Où     donc la situer, d'ailleurs, sinon  dans la   raison elle-même [...]?

   where donc it place  anyway    if-not  in    the  reason itself

   'Where  to  place  it,  anyway,  if  not  in  reason  itself  [...]?'  (Foucault 1961)

   b. [...] comment doncse    remettre à  vivre comme si de rien    n'  était?

        how       doncREFL restart  to  live  as     if of nothing  NEG  was

    '[...]  how  to  get  back  to  living  as  if  nothing  had  happened?'  (Romilly 1993)

   c. [...] Et    pourquoi donc s'      indigner?

        and  why       donc  REFL  be-angry?

    '[...]  And  why  be  angry?'  (Moret 1973)

We found one example with no verb at all (32), which can be analyzed as a non elliptical verbless clause, with *quand donc* ('when-donc') as the head and the following

NP as its subject (Laurens 2008).

(32) Quand  donc un  bon  lit, une  nuit  sans      trouble [...]?
  when   donc a    good bed a     night without trouble?
  'When a good bed, a night without trouble [...]?') (Mauriac 1996)

We only found two occurrences of wh-*donc* in embedded clauses, and they are both indirect speech:

(33) a. je me    demandais: où    donc se    cache  l' essence qui  anime
     I  REFL asked       where donc REFL hides  the essence that animates
     ce    corps?
     this  body?
     'I was wondering: where is the essence that animates this body?'
     (Margerie 1985)
   b. [...] demandé comment donc  grand-maman avait pu   apprendre à
       asked    how     donc grandma      had   can teach     to
       écrire à   des  jeunes  enfants.
       write  to  INDF young   children
       '[...] wondered how grandma could have taught young children to write.'
       (Roubaud 1993)

Most of the times, the questions with fronted wh-*donc* are not genuine information questions. They often contain the answer, which may be introduced by *sinon* ('unless') (31a), or are just followed by it (29b). They can also be rhetorical or self-addressed questions, as in (34a).

(34) a. Quand donc serais -je   grand? Je  le voulais  tellement, du
     when  donc be-COND-I big    I   it wanted   so-much  from-DEF
     fond     de ma rage de  vivre.
     bottom of  my rage to  live
     'When will I grow up? I wanted it so much, in my rage to live' (Jardin
     1986)

b. Et  vous allez les    récupérer quand  donc, ces    deux  jours?
and you   go    them collect   when   donc these two    days
'And  when  are  you  going  to  get  them  back,  these  two  days?' (Clavel 1962)

c. [...] alors elle entrevoit quoi donc? une absence?  une présence?
      so    she glimpses what donc? an   absence? a    presence
'[...] so,  what  does  it  glimpse?  an  absence?  a  presence?' (Sartre 1961)

According to our corpus, the only wh-words with *donc* that are found *in situ* are *quand* ('when') (34b) and *quoi* ('what') (34c). These in situ questions occur as a discourse continuation (with *et* 'and', *alors* 'so'), and seem to be biased: (34b) suggest it would be difficult to get those days back (maybe never), and in (34c) the speaker goes on suggesting possible answers.

## 5.2 Elliptical uses of wh-*donc*

Even if it is lower than with wh-*ça*, a high proportion (33%) of wh-*donc* are elliptical. Almost all elliptical wh-*donc* uses are sluices (98%), unlike wh-*ça* with which one third were wh-stripping. *Quoi donc* and *pourquoi donc* are most common among sluices. All wh-*donc* sluices are matrix sluices.

(35) a. C'est  de Colombie.   −Quoi donc?    −Le  café.
       it is    of Colombia     what donc?      the coffee
'It's  from  Colombia. − What?         − The  coffee.' (Sabatier 1966)

b. Mrs. Killarney, vous devriez      cacher ça. − Pourquoi  donc?
mrs  Killarney  you  must-COND hide   it     why       donc?
'Mrs.  Killarney,  you  should  hide  this. − Why?' (Queneau 1962) ;

c. Ce soir    -là, [...] je me   suis saoulé. − Ah bon? Et  où     donc,
this evening there    I REFL am  drunk    ah so    and where donc
monsieur Rima?
mister    Rima?
'That  evening, [...] I got drunk. − Really? And where, Mr Rima?' (Vergne 1985)

As with wh-*ça*, sluices with wh-*donc* may have an overt correlate (35a), an implicit correlate or no correlate (*sprouting*) (35b,c).

Among the 174 sluices, 88 (51%) have a correlate for the wh-*donc* word, mostly with *quoi donc* ('what donc') (78%), 0% with *comment donc* ('how donc') and *pourquoi donc* ('why donc') (sprouting), which is a bit higher than with wh-*ça* (44%).

Looking at wh-*donc* sluices with a correlate, we annotated whether it is definite, indefinite (*quelque chose* 'something') (36a) or implicit (36b) (Table 7).

(36) a. Pas tout, Malaussène, j'ai     quelque chose à ajouter. −Quoi donc?

       not all    Malaussène I have   something      to add       what donc?

       'Not all, Malaussène, I have something to add. −What?' (Pennac 1990)

   b. Alors? J'attends. −Quoi donc?

       so     I wait      what donc

       'So? I am waiting. −For what?' (Le Clézio 1963)

Table 7. Correlates of wh-sluices + *donc*

| + *donc* | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| definite | 35 | 7 | 6 | 1 | 0 | 0 | 49 (28%) |
| indefinite correlate | 21 | 0 | 2 | 2 | 0 | 0 | 25 (14%) |
| implicit correlate | 13 | 0 | 1 | 0 | 0 | 0 | 14 (8%) |
| no correlate | 19 | 3 | 11 | 1 | 45 | 7 | 85 (49%) |
| total sluicing | 88 | 10 | 20 | 4 | 45 | 7 | 174 (100%) |

As with wh-*ça*, we found some cases of direct sluices with an indefinite correlate, which can be explicit: *quelque chose* ('something'), *autre chose* ('something else'), *ailleurs* ('elsewhere'), *on* ('one') (36a) or implicit (36b). But, as with wh-*ça*, we found a majority of reprise sluices (with a definite correlate) (56% sluices with a correlate), even if it is lower than with wh-*ça* (81%). Often the definite correlate is a pronoun: *celui-là* ('that one') (37a), *elle* ('she') (37b), *ils* ('they') (37c):

(37) a. Qui c'est celui-là?       −Qui donc? −Celui    qui   se     débattait

       who it is that-one-there    who    donc? that-one who REFL struggled

       tout à l'heure...?

       earlier

'Who's that one? – Who? – The one who was struggling earlier...?'
(Mauriac 1978)

b. Comment elle s'    appelle? – Qui  donc?  – La   vache.

   how       she REFL call          who donc?    the  cow

   'How is she called? – Who? – The cow.' (Sabatier 1974)

c. [...] c'est là    qu' ils   crèchent? – Qui donc? – Les académiciens, là-bas.

      it is there that they live         who donc?  the academicians  there

   '[...] is it there that they live? – Who? – The academicians, there.'
   (Mauriac 1984)

When the source clause is a wh-question, a full verbal question would be difficult:
*Qui c'est qui donc?* ('Who it's who donc?') (37a), *Comment qui donc s'appelle?*
('How who donc is called?') (37b). Also, when the correlate (*ils* 'they') is inside an
it-cleft, as in (37c), a full verbal question would be difficult:*? C'est là que qui donc
crèche?* ('It is there that who donc lives?').

Finally, we found one case with *alors* ('then') and a polarity mismatch between the
source clause and the sluice. The sluice in (38) means *Qui donc alors en avait la
surveillance?* ('Who then had her under supervision?') and not *Qui donc alors n'en avait
pas la surveillance?* ('Who then didn't have her under supervision?'):
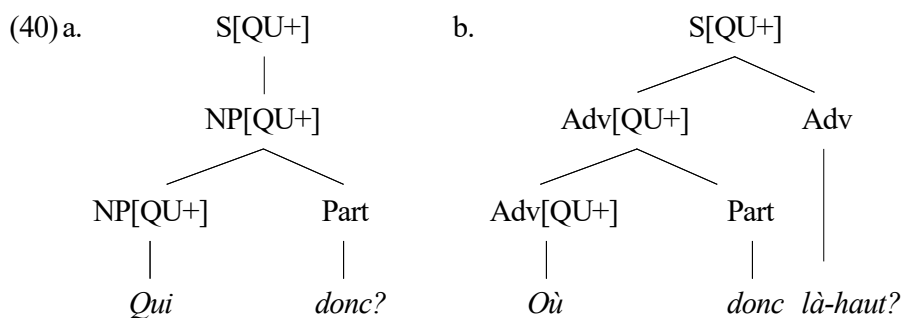
(38) – Il   n'    en      avait pas  la  surveillance, dit  Ouroz.

      he NEG of-her had  NEG the supervision  said Ouroz

     – Qui donc alors?   – La   servante.

      who donc then?       the  servant

     'He didn't have her under supervision, said Ouroz. – Who did then? – The
     servant.' (Kessel 1967)

Regarding wh-stripping with *donc*, we only found one case with *où* ('where') (39).

(39) Tu   as   mis des  gens    là-haut? – Où    donc là-haut?

    you  have put INDF people  up there     where donc up there

    'Did you put people up there? – Where up there?' (Giono 1982)

As with wh-*ça* (section 4.2), most of the times, a non-elliptical counterpart is not

available, or would have a very different form from that of the preceding utterance. We thus suggest an analysis with a Direct interpretation approach (Ginzburg and Sag 2000; Culicover and Jackendoff 2005), as for wh-*ça* (section 4.2); the particle is adjoined to the wh-word and the elliptical clause is a fragment, with its content inferred from that of the antecedent clause:

(40) a.

```
            S[QU+]                        b.            S[QU+]
              |                                         /    \
          NP[QU+]                              Adv[QU+]       Adv
           /    \                               /    \         |
      NP[QU+]    Part                      Adv[QU+]    Part     |
         |        |                          |          |      |
        Qui     donc?                        Où        donc  là-haut?
```

## 6.   General discussion

We first compare the use of the two question particles *ça* and *donc*, and then compare them with bare wh-words.

### 6.1 Comparison between wh-*ça* and wh-*donc*

Contrary to Lefeuvre (2018), we found that ellipsis is not the only use of wh-*ça*, and both *in situ* and extracted uses, even if they are rare, cast doubt on her subject analysis of *ça*. Contrary to Reinhardt's corpus (section 2.1), we found that sluicing is not the only use of wh-*donc*.

We also found that both particles only occur in matrix questions (or after a speech verb), which may be due to their emotive potential.

They do not occur with the same wh-words. *Pourquoi donc* and *qui donc* are the most frequent wh-words with *donc*, while *comment ça* and *où ça* are the more frequent words with *ça* (Table 3). *Comment* is the most common word with *ça*, and has a special 'how come' interpretation, while *comment donc* is rare.

Looking at the use of the wh-word with a particle, we also found that both wh-*ça* and wh-*donc* are very rarely used *in situ*. The most common use of wh-*donc* is fronting

(extracted or subject) (65% in our sample), followed by ellipsis (33%) (table 6), while the most common use of wh-*ça* is ellipsis (sluicing or wh-stripping) (98.5% in our sample) (table 4). As for sluicing per se, it represents 33% of wh-*donc* and 67% of wh-*ça*, and they are (almost) all matrix sluices. Looking at sluices in our corpus, *quoi donc* is the most common wh-word in sluices (30a), while *quoi ça* is almost non existent. On the other hand, *qui ça* is common in sluices while *qui donc* is rare. In sluices, *qui ça* mostly has a definite correlate (Table 5), while it is not necessarily the case for *quoi donc* (Table 7). We also found a new kind of ellipsis that we called *wh-stripping*, which is quite common with wh-*ça* (31% in our sample) while it is almost non existent with wh-*donc*.

A possible explanation of the differences between wh-*ça* and wh-*donc* is that *ça* tends to point more to the preceding context (favoring sluices with correlates and reprise questions), while *donc* is more apt to point to the discourse continuation. Indeed, many questions with *donc* are not ignorance questions: some are rhetorical questions (implying a negative answer, such as *no one, nowhere*), as in (34b), while 11 of them contain the answer, with a *sinon* ('unless') continuation (31a) and (41a) or a comparative (41b):[8]

(41) a. Qui   donc sinon   la   France?
        who  donc unless  the  France
        'Who  else unless France?' (de Gaulle 1970)

    b. [...] où      donc la   France pourrait  -elle les     affirmer mieux qu'
          where donc the France can-COND  she them affirm    better than
       à   Strasbourg?
       in  Strasbourg
       '[...] where else could France affirm them better than in Strasbourg?' (de Gaulle  1970)

Looking at sluices with a correlate, they are mostly reprise sluices, with a definite correlate, with both particles, especially with wh-*ça*: 81% of wh-*ça* sluices (with a correlate) and 56% of wh-*donc* sluices (with a correlate) have a definite correlate for the wh-word. Since wh-stripping is used with a reprise fragment, this suggests that in their elliptical uses (sluicing or wh-stripping), both particles favor reprise questions (55% of

---

8   This is possible with *ça* but much rarer: *Qui ça, on, sinon les chers officiers aux mains pures*? ('Who ça, one, unless our dear officers with pure hands?') (Beauvoir 1963).

wh-*ça* and 29% of wh-*donc*, including wh-stripping and sluicing with a definite correlate). We now look at bare-wh words in order to test this hypothesis.

## 6.2 Comparison with bare wh-words + *?*

In order to compare wh-*ça* and wh-*donc* with bare wh-words, we searched for bare wh-words in final position, followed by a question mark in Frantext after 1960 (12332 occurrences, cf. Table 3), since both particles occur with direct questions. We then annotated 532 bare wh-questions, taking 5% out of all bare wh-word occurrences. As illustrated in Table 8 below, their most common use is sluicing (67%), followed by the *in situ* use. We did not find bare wh-words in fronted (extracted/subject) position nor in wh-stripping, since we only searched for wh-words followed directly by the question mark. Yet, we can observe that the *in situ* use is much more frequent with bare wh-words compared to wh-*ça* or wh-*donc*.

Table 8. Annotated uses of wh-? in Frantext (1960-)

| +? | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| in situ | 99 | 16 | 30 | 11 | 1 | 17 | 174 |
| sluicing | 129 | 53 | 17 | 12 | 122 | 25 | 358 |
| total | 228 | 69 | 47 | 23 | 123 | 42 | 532 |

As for sluices, we only found matrix sluices given our *?* query. But the distribution of bare wh-words in sluices is different: *quoi?* is the most frequent bare sluice, as it is with *quoi donc* (Table 6), while it is very rare with *quoi ça* (Table 4). The interpretation of *quoi* with and without particle may be somewhat different. The main use of bare *quoi* ('what*)* sluice is to question an inanimate entity (42), with a few reprise uses (with *ce* ('this') as the correlate) (43a), and one speech act use (*pardon?*) (43b):

(42) Attends, y    a    un truc qui    va    pas du tout, là ... − Quoi?
     wait    there have a    stuff which goes NEG at all    there    what
     'Wait, there's something going really wrong out there... − What?' (Gavalda 2004)

(43) a. On arrête  maintenant,  d'accord, avant  que ce soit       définitif. −
　　　we stop     now             ok         before that it be-SUBJ  final
　　　Quoi? Quoi définitif?
　　　what   what final
　　　'We're stopping now, ok, before it's final. −What? What final?' (Angot
　　　2006)

　　 b. Je croyais  que  t'   étais parti... − Demain. − Quoi? − Je pars
　　　I  thought that  you was  left        tomorrow what    I   leave
　　　demain.
　　　tomorrow
　　　'I thought you had left... − Tomorrow. − What? − I'm leaving tomorrow.'
　　　(Gavalda 2004)

On the other hand, *quoi donc* does not have this speech act use (35a).

　　 Bare *pourquoi?* ('why') is also frequent as a sluice (Table 8), like *pourquoi donc*,
while *pourquoi ça* is more rare. Bare *pourquoi* ('why') sluice has two main uses:
question about reason/purpose (44) or about the speech act ('why are you telling/asking
me that?') (45), like *pourquoi donc* (35b)*,* while *pourquoi ça* mainly asks about
reason/purpose (only 1 example of reprise use in (25a):

(44) Pourquoi  tu    te      lèves?    − Oh... Je vais  lui   donner
　　 why          you  REFL stand-up   oh    I  go    him give
　　 son biberon ... − Mais pourquoi?
　　 his bottle          but   why
　　 'Why are you getting up? − Oh ... I'm going to give him his bottle... − But
　　 why?' (Gavalda 2000)

(45) T'   es  croyant? − Pourquoi? − Pasqu' en  prime  t'    auras
　　 you are believer    why          cause  on  top     you have-FUT
　　 la  bénédiction d'  Allah !
　　 the blessing      of  Allah
　　 'Are you a believer? − Why? − Cause as a bonus you'll have the blessing of
　　 Allah!' (Lasaygues 1985)

Bare *comment?* ('how') sluice has three main uses: question about manner / identity

(46a), most of the time, speech act (*pardon?*) (47), or reason ('how come') (48). The two last uses would be possible with *comment ça*, but not the first one (46b). On the other hand, the reprise use of *comment ça* (26b) (27ac) would be difficult with bare *comment*.

(46) a. Lorsque mes livres parurent, en octobre, je fus   saisi  d' une panique
         when    my  books appeared in October I   was seized of a    panic
         telle que  je ne    parlais que  de les    détruire.
         such that I  NEG   talked  that of  them destroy
         Mais comment?
         but   how
         'When my books got published, in October, I was seized with such panic that I was only talking about destroying them. But how?' (Althusser 1985)
      b. Lorsque mes livres parurent, en octobre, je fus saisi d'une panique telle que je ne parlais que de les détruire. Mais comment donc? #Mais comment ça?

(47) On cherche l'  emplacement du     lac de La Dame, du     Lac de Chandler.
     we search    the location          of-DEF lac de La Dame  of-DEF Lac de Chandler
     On pensait y    aller demain... −Hein? Quoi? Comment? J'en   étais sûr !
     we thought there go   tomorrow   uh    what   how       I of-it was sure
     'We are looking for the location of Lac de La Dame, of Lac de Chandler. We were thinking about going there tomorrow... −Huh? What? How? I was sure of it!' (Manœuvre 1985)

(48) Ce mec  avait une sacrée avance. Mais comment, merde, comment?
     this guy had  a    cursed lead    but  how       shit   how
     'This guy had a hell of a head start. But how come, shit, how come?' (Page 1982)

We also annotated the correlate of the bare wh-words in all sluices (Table 9).

Table 9. Correlates of bare wh-sluices in Frantext (1960-)

| +? | *quoi* 'what' | *qui* 'who' | *où* 'where' | *quand* 'when' | *pourquoi* 'why' | *comment* 'how' | total |
|---|---|---|---|---|---|---|---|
| definite | 11 | 7 | 0 | 1 | 2 | 0 | 21 (6%) |
| indefinite correlate | 17 | 18 | 7 | 1 | 0 | 0 | 43 (12%) |
| implicit correlate | 0 | 0 | 0 | 0 | 0 | 0 | 0 (0%) |
| no correlate | 101 | 28 | 10 | 10 | 120 | 25 | 294 (82%) |
| total sluicing | 129 | 53 | 17 | 12 | 122 | 25 | 358 (100%) |

In sluices, most bare wh-words (82%) don't have a correlate. When they have one, it is usually an indefinite correlate (12%). The reprise uses of bare wh-sluices (with a definite correlate) are very rare (6%). By comparison, 51% of occurrences of wh-*donc* sluices have a correlate and 44% of wh-*ça* sluices, and most of the time it is a definite correlate (tables 5 and 7). Furthermore, 100% of *quoi ça* and 98% of *qui ça* sluices have a correlate against only 22% for bare *quoi* and 47% for bare *qui*.

In conclusion, our data suggest that the discourse particle helps the anaphoric relation with a correlate, and with *ça* it is most of the time a definite correlate, hence a use of as a reprise question more than as a standard information question.

Compared with English, among 4700 sluices in the Santa Cruz database (data extracted from the New York Times newspaper) (Pranav, Hardt, and McCloskey 2020), 27% are root sluices and 35% have an overt correlate.[9] If we now turn to the corpus study of wh-*the hell* by Kim and Kim (2021): they found over 10 000 occurrences of wh-*the hell/ the heck/ the fuck/the devil* in the COCA, mostly in fiction texts. 85% of them are matrix questions, and there is an overrepresentation of *what the hell* (49a) (62%), compared to other wh-words. Contrary to previous literature (Ginzburg and Sag 2000; Merchant 2001), they also found examples of elliptical uses (49b) and (49c).

(49) a. What the hell're those for? (COCA 2009 FIC)

b. Suddenly a human hand punches through the wall. CHRIS: What the hell? (COCA 2002 FIC)

c. "My God," he said. "They survived. How the hell?" (COCA 2000 FIC)

They annotated 2112 examples of sluicing and sprouting (2000 with *what the hell*)

---

9   Probably due to the formal register of their corpus (extracts from the *New York Times*), they only found one example with *the hell*. The only example in the Santa Cruz sluicing database is a matrix sluice (with swiping): *I said "Leave the car there? What the hell for? Why don't I just keep it?".*

and found that the majority of them had an exophoric antecedent (49b). *Why*-stripping cases are rarer but attested (50a), mostly with *not* (50b).

(50) a. " ... A hundred ten years ago, the first colonists brought along frozen bird embryos-mostly chicken embryos, I understand. " "Why the hell chickens?" "For eggs... and meat..." (COCA 1991 SPOK)

b. How about Detroit? −Detroit? Why the hell not Detroit? (COCA 2015 FIC)

Contrary to our French particles, the English particles (*the hell/ the heck/ the fuck/the devil*) denote total ignorance of the speaker. In French, we found that the major use of wh-*ça* is for reprise questions, and the major use of *wh-donc* for biased questions.

## 7.   Conclusion

Contrary to previous theoretical claims (Obenauer 1976; Cheng and Rooryck 2001), we found that both wh-*ça* and wh-*donc* are compatible with extracted and *in situ* positions in French. We also show that they are specialized for matrix questions, like German particles (Czypionka et al. 2021), which may be explained by their emotive force.

We show that their distribution is quite different. While *ça* mostly combines with adjuncts (*où* 'where' and *comment* 'how'), *donc* often combines with arguments (*qui* 'who' and *quoi* 'what'). While *ça* is mostly used in elliptical questions, *donc* is mostly used in fronting verbal questions.

We also found out that both particles favor elliptical questions, which may be related to the role they play in discourse cohesion. Ellipsis, as a source-antecedent relation, indeed plays a major role in discourse cohesion (Halliday and Hasan 1976; Kehler 2002). In particular, both French particles favor sluices with a correlate (44% of wh-*ça* and 51% of wh-*donc*), while only 18% of sluices with a bare wh-word have a correlate (in our sample). These sluices with *ça* or *donc* also often have a definite correlate, and are thus used as reprise questions, which is not the most common use of bare wh-words.

We also discovered a new elliptical construction with *ça* which we call *wh-stripping* (with a reprise fragment), which is very rare with *donc*. This casts doubt on the claim

that wh-stripping is restricted to *why* cross-linguistically (Yoshida et al. 2015). In this construction, the wh-question is a reprise question.

We conclude that the main use of wh-*ça* is reprise questions. On the other hand, the main use of wh-*donc* seems to signal a biased question (rhetorical, including the answer or self-addressed), but further work on the answers is necessary to confirm this hypothesis.

# References

Abeillé, Anne and Danièle Godard. 2008. *Tout, ça, quoi*: Remarques sur les mots légers en français. In Dan van Raemdonck (ed.), *Modèles Syntaxiques: La syntaxe à l'aube du XXIe siècle*, 85–111. Bruxelles: Peter Lang.

Baude, Olivier and Céline Dugua. 2016. Les *ESLO*, du portrait sonore au paysage digital, *Corpus*, 15. http://journals.openedition.org/corpus/2924.

Bayer, Josef and Hans-Georg Obenauer. 2011. Discourse particles, clause structure, and question types. *The Linguistic Review* 28(4): 449–491.

Brunetti, Lisa, Lucia M. Tovena, and Hiyon Yoo. 2020. French questions alternating between a reason and a manner interpretation. *Linguistic Vanguard.* In press.

Cheng, Lisa L.-S. and Johan Rooryck. 2001. Licensing wh-in-situ. *Syntax* 3(1): 1–19.

Czypionka, Anna, Maribel Romero, and Joseph Bayer. 2021. Question-sensitive discourse particles at the interfaces of syntax, semantics and pragmatics – An experimental approach. *Glossa: A Journal of General Linguistics* 6(1): 24.

den Dikken, Marcel and Anastasia Giannakidou. 2002. From *hell* to polarity. *Linguistic Inquiry* 33: 31–61.

Desmets, Marianne and Antoine Gautier. 2009. 'Comment n'y ai-je pas songé plus tôt?', questions rhétoriques en *comment. Travaux de Linguistique* 58(1): 107–125.

Gibson, Edward and Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Science* 14: 233–234.

Ginzburg, Jonathan. 2019. Exclamative sluices: Evidence for semantically-based ellipsis resolution. *Sinn und Bedeutung* 2019, Osnabrück.

Ginzburg, Jonathan and Ivan A. Sag. 2000. *Interrogative investigations: The form, meaning and use of English interrogatives.* Stanford: CSLI Publications.

Halliday, Michael A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Huot, Hélène. 1986. Le subjonctif dans les complétives: Subjectivité et modélisation. In Mitsou Ronat and Daniel Couquaux (eds.), *La Grammaire Modulaire*, 81–112. Paris: Minuit.

Culicover, Peter and Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Huang, C.-T. James and Ochi Masao. 2004. Syntax of the hell: Two types of dependencies. In Keir Moulton and Matthew Wolf (eds.), *Proceedings of the 34th Conference of the North Eastern Linguistic Society (NESL)*, 279–293. Amherst: University of Massachusetts GLSA.

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

Kim, Jong-Bok. 2017. Why-stripping in English and Korean: A direct interpretation approach. *Studies in generative grammar* 27(4): 733–751.

Kim, Jong-Bok and Anne Abeillé. 2019. Why-stripping in English: A corpus-based perspective. *Linguistic Research* 36(3): 365–387.

Kim, Jong-Bok and Jungsoo Kim. 2021. Agressively non-D-linked wh-phrases and ellipsis: A direct interpretation approach. *Journal of Linguistics*. Under review.

Kim, Jong-Bok and Joanna Nykiel. 2020. The syntax and semantics of elliptical constructions: A direct interpretation perspective. *Linguistic Research* 37(2): 327–358.

Laurens, Frédéric. 2008. French predicative verbless utterances. In Stefan Müller (ed.), *Proceedings of the 15th International Conference on head-driven phrase structure grammar*, 152–172. Stanford: CSLI Publications.

Lefeuvre, Florence. 2018. Les interrogatives partielles averbales: Le cas de « proforme interrogative + ça / cela. In Marie-José Béguelin, Aidan Coney, and Alexander Guryev (eds.), *L'Interrogative en français*, 183–208. Bern: Peter Lang.

Lefeuvre, Florence. 2020. Les interrogatives partielles dans un corpus de théâtre contemporain. *Langages* 217(1): 23–38.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. New York: Oxford University Press.

Merchant, Jason. 2011. Ellipsis. In Artemis Alexiadou, Tibor Kiss, Miriam Butt (eds.), *Handbook of contemporary syntax*, 2–36. Berlin: Walter de Gruyter.

Miller, Philip. 1991. *Clitics and constituents in phrase structure grammar*. Outstanding dissertations in linguistics. New York: Garland Press.

Miller, Philip and Geoffrey Pullum. 2014. Exophoric verb phrase ellipsis. In Philip Hofmeister and Elisabeth Norcliffe (eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 5–32. Stanford: CSLI Publications.

Obenauer, Hans-Georg. 1976. *Études de syntaxe interrogative du français: Quoi, combien et le complémenteur*. Tübingen: Niemeyer.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In Eric Reuland and Alice ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, Mass.: MIT Press.

Pranav, Anand, Dan Hardt, and James McCloskey. 2021. The Santa Cruz sluicing dataset. *Language* 97(1): 68–88.

Rawlins, Kyle. 2009. Extreme ignorance questions. *LSA Annual Meeting* 2009, San Francisco.

Reinhardt, Janina. 2019. *Regularity and variation in French direct interrogatives*. PhD Dissertation. University of Konstanz. Corpus: https://hdl.handle.net/11403/interrogatives-in-novels/v1.

Ross, John R. 1969. Guess who? In Robert I. Binnick, Alice Davison, Georgia M. Green, Jerry L. Morgan et al. (eds.), *Proceeding of the Fifth regional meeting of the Chicago Linguistic Society*, 252–286. Chicago: Chicago Linguistic Society.

Thiberge, Gabriel. 2020. *Acquisition et maîtrise des interrogatives partielles en français: La variation sociolinguistique comme outil interactionnel*. PhD Dissertation. Université de Paris.

Yoshida, Masaya, Chizuru Nakao, and Ivan Ortega-Santos. 2015. The syntax of *why*-stripping. *Natural language and linguistic theory* 33(1): 323–370.

**Anastasiia Smirnova**
Master student
LLF
Université de Paris
8 rue Albert Einstein, Paris, 75013, France
E-mail: anastasiia.smirnova@etu.u-paris.fr

**Anne Abeillé**
Professor
LLF, CNRS
Université de Paris
8 rue Albert Einstein, Paris, 75013, France
E-mail: abeille@linguist.univ-paris-diderot.fr