

FRENCH WH-WORDS + ÇA A CORPUS STUDY

ANNE ABEILLÉ (U PARIS)

ANASTASIIA SMIRNOVA
(U PARIS)

JONG-BOK KIM (KYUNG
HEE U)

OUTLINE

Previous corpus
studies

Previous studies on
wh+ça

Our corpus study,
examples and results

Statistical analysis,
conclusion

INTRODUCTION

Discourse particles affect the distribution and interpretation of wh-questions.

French *-ça* ('that'):

Obenauer (1974), Cheng and Rooryck (2001) claim that wh-words+*ça* only occur in situ :

- (1) Tu as vu qui *ça* (cet après-midi) ? 'You have met who-*ça* (this afternoon)?'
- (2) *Qui *ça* as-tu vu (cet après-midi) ? 'Who-*ça* have you met this afternoon?'

But they can occur in sluices:

- (3) — J'ai vu quelqu'un. 'I saw someone.'
- Qui *ça* ? 'Who-*ça*?'

⇒ A challenge for deletion-based analyses of sluices (Ross 1969, Merchant 2001)

INTRODUCTION

Two goals:

1. Distribution of *wh+ça* in a large corpus.
2. Comparing sluices and full questions
 - sluices are shorter: easier to produce;
 - sluices are elliptical: more difficult to understand.

Hypothesis 1: Sluicing is easier with *wh*-words with a propositional content (*pourquoi, comment*).

Hypothesis 2: Sluicing is easier with *ça* (*ça* comes from a demonstrative pronoun hence more anaphoric potential).

PREVIOUS CORPUS STUDIES

Santa Cruz Ellipsis Project ([English](#))

- New York Times sub-corpus of the English Gigaword (Second Edition) corpus
- 4700 annotated examples of sluicing
- Quantitative results:
 - 35% sluices with overt correlate
 - 27% root sluices

Decreasing frequency:

reason (why) > degree (how much) > entity (what/who) > manner (how) > temporal (when) > locative (where)

Root sluices: degree > entity > manner > reason > locative > temporal

PREVIOUS CORPUS STUDIES ON WH+ÇA

Lefevre F. (2018) (French)

- Frantext 1700-2010
- 780 occurrences of sluicing with *ça* (final position); no *combien ça* (how-much-ça)

decreasing frequency :

où *ça* ('where') > comment *ça* ('how') > qui *ça* ('who') > pourquoi *ça* ('why') > quand *ça* ('when') > quoi *ça* ('what')

Lefevre 2020: Smaller spoken corpus ESLO2 (meal conversation):

30% of wh-questions are sluices, among which 8% with *ça*

(4) - bah je l'ai vue 'I saw her'

- quand *ça* ? 'when-ça?' (ESLO2-REPAS)

OUR CORPUS STUDY

Frantext (1960-2010)

- Literary corpus of contemporary French
- 1001 texts, 61,2 M words

Ça cannot occur with a weak (**Que ça?* ‘what *ça*’) or anaphoric pronoun (**lequel ça* ‘which-*ça*’) nor with a determiner (**Quel ça homme?* ‘Which-*ça* man?’)

Two sub corpora :

1. We searched for *qu’il/ qu’où/ comment/ où/ combien/ pourquoi/ quand* + *ça* => 481 occurrences
2. We filtered out the use of *ça* as a subject or dislocated pronoun (*Où ça fait mal ?* ‘where does it hurt?’ *C’est quoi, ça ?* ‘what is this ?’). => 335 occurrences annotated (no *combien ça*)
3. We searched for the same wh-words in final position (except for *combien*) => 10778 wh-+? (in situ or sluice)
4. We extracted 5% of each bare wh-word total => 532 occurrences annotated

CORPUS SEARCH RESULTS

Wh-word	Quoi 'what'	Qui 'who'	Où 'where'	Quand 'when'	Pourquoi 'why'	Comment 'how'	Total
wh- + <i>ça</i>	3	91	116	13	23	235	481
wh- + ?	4566	1379	939	570	2474	850	10778
Total	4569	1470	1055	583	2497	1085	11259

Table 1: Wh-word + *ça* and wh- + ? in Frantext 1960-2010

ANNOTATED CORPUS DATA

Wh-word	+ça				Total	bare		Total	Grand Total
	extract	insitu	sluicing	stripping		insitu	sluicing		
comment ('how')	1		38	65	104	16	26	42	146
où ('where')	1	2	85	12	100	30	17	47	147
pourquoi ('why')	1		21	1	23	1	122	123	146
quand ('when')		1	11	1	13	11	12	23	36
qui ('who')			65	26	91	16	53	69	160
quoi ('what')			4		4	99	129	228	232
Total	3	3	224	105	335	173	359	532	867

Table 2: Annotated occurrences of wh-word + *ça* and wh- ?

MAIN FINDINGS ON WH+ÇA

Extraction of *wh+ça* is rare (3 occ) but possible (without subject inversion) contra Obenauer (1974), Cheng & Rooryck (2001), Sales 2009

(5) « Et pourquoi ça vous feriez pendre mon papa, monsieur le duc ? »

‘And why-ça would you hang my dad, Mr. Duke?’ (Queneau 1965)

(6) « Où ça qu'ils vont ces deux copains-là ? »

‘Where-ça do they go these two friends ?’ (Sabatier 1985)

(7) « Comment ça, s'il te plaît, que je ne pourrais pas le savoir ? »

‘How-ça, please, could I not know that?’ (Koltès 1985)

Very rare uses of *wh+ça* in situ (3 occurrences):

(8) « — Mon papa il est mort. — Ton papa il est mort, quand ça ? »

‘My dad has died. — Your dad has died, when-ça ?’ (Mauriac 1961)

(9) « T'avais appris à tirer où ça ? »

‘You’d learnt to shoot where-ça?’ (Boudard 1982)

MAIN FINDINGS ON WH+ÇA

Main uses of *wh+ça* are sluicing and stripping (98% of *wh+ça*)

(10) — *Ça te dirait de partir ? — Où ça ? — Ailleurs.*

‘Would you like to leave? — Where-ça? — Elsewhere.’ (Page 1982)

Very rare embedded sluices (4 occurrences) (reported speech)

(11) *Elle me dit, vous savez, je vous connaissais avant de vous rencontrer. Étonné, je demande comment ça ?*

‘She tells me, you know, I knew you before I met you. Surprised, I ask how-ça?’ (Doubrovsky 2011)

Comment ça: more frequent with stripping (64% of *comment ça*)

(12) « — *Je ne me suis pas posé la question. Il fallait que je l'écrive. — Comment ça, il fallait ?* »

‘I did not ask myself. I had to write it. — How-ça, you had to?’ (Groult 2008)

MAIN FINDINGS ON WH+ÇA SLUICES

- Ginzburg and Sag (2000)

Direct sluices

(14) « Quelqu'un me l'a dit. — Qui ça ? »
'Someone told me that. — Who-ça?' (Vargas 2008)
(indefinite correlate)

(15) « — Bien sûr, je copie. — Qui ça ? — Damascius. »
'Of course, I copy. — Who-ça? — Damascius.' (Roubaud 2002)
(implicit correlate)

(16) — Mlle Dreyfus ne travaille plus ici. Elle nous a quittés.
— Quand ça ?
'Ms Dreyfus doesn't work here anymore. She has left us.
— When-ça?' (Gary 1974) *(no correlate : sprouting)*

Reprise sluices

(17) « — Attention, elle est bouddhiste ! vous avertit Melchior. — Qui ça ? — Claire. »
'Be careful, she is a Buddhist! Melchior warns you.
— Who-ça? — Claire.' (Buron 1998)

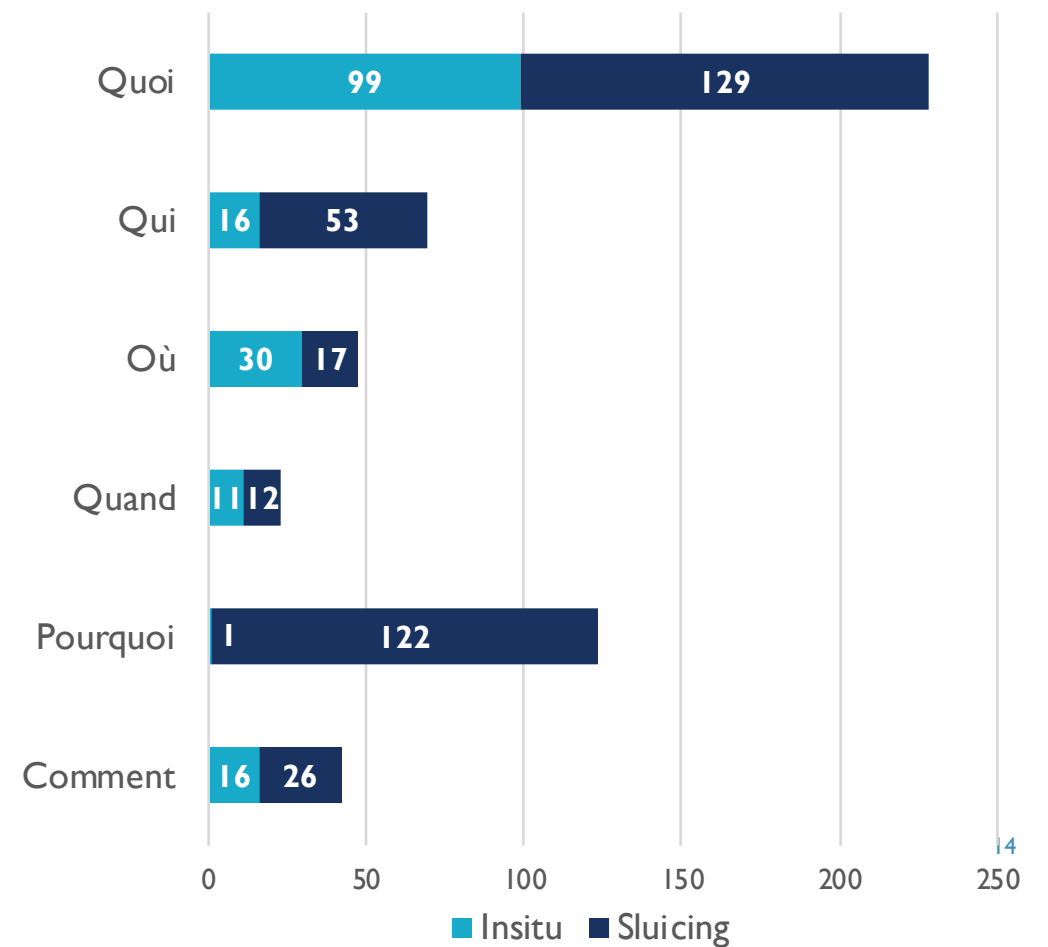
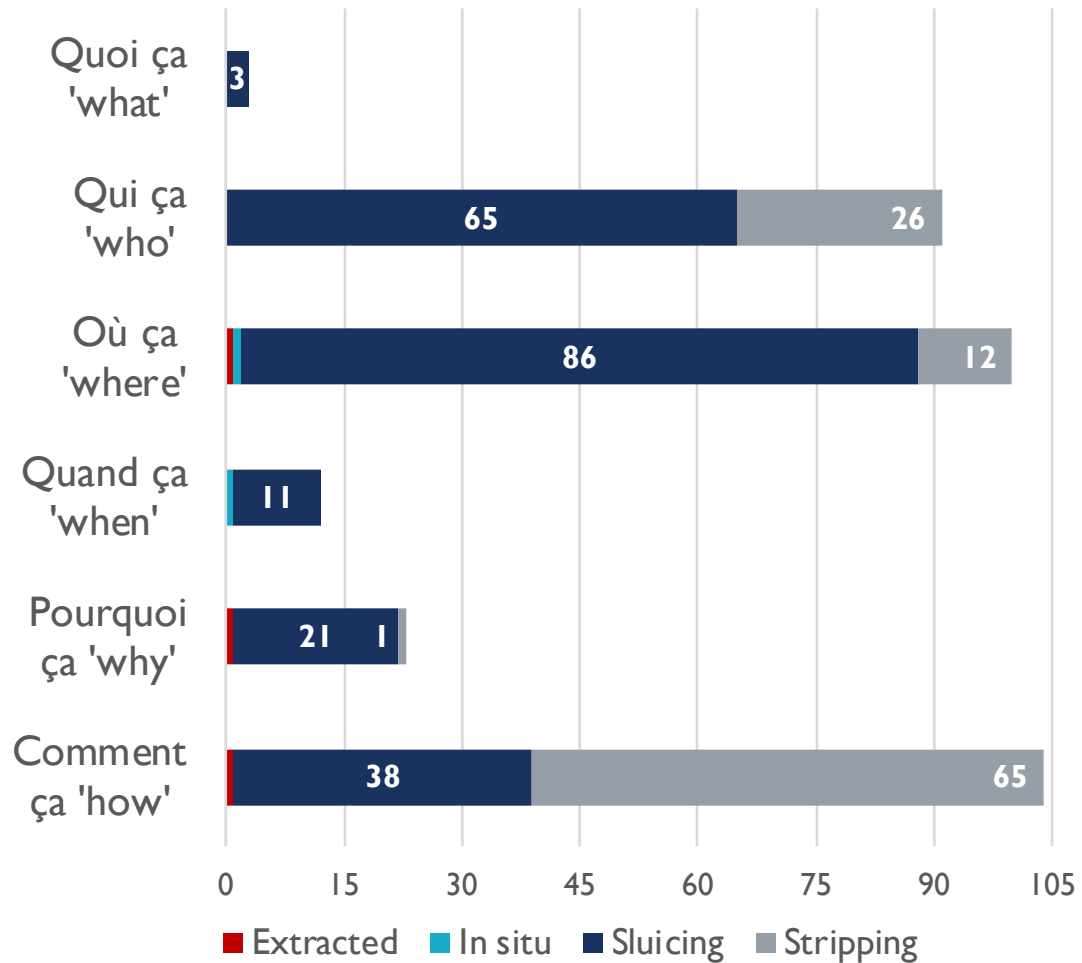
ANNOTATION SCHEME

Each example was annotated with following tags (variables):

- wh-word
- presence of *ça*
- type of use (sluicing, in situ) + extracted, stripping (for wh+*ça*)
- fragment (only for stripping)
- correlate (only for sluicing)
- embedded

Example	Author	Wh-word	Wh-word_bin	Correlate	Correlate_bin	Ça_bin	Use	Fragment	Embedded
— On dirait qu'il date d'hier, n'est-ce pas ? demande Walter. — Quoi ça ? — Le monument... Il est encore tout frais.	Mertens	quoi	yes	il	yes	yes	sluicing	no	no

RESULTS



RESULTS ON (ROOT) SLICES (BARE WH- WORDS)

Quoi > pourquoi > qui > comment > où > quand

Quoi ‘what’. Two uses: question about an inanimate entity (18) or speech act use (‘pardon?’)(19)

(18) « Attends, y a un truc qui va pas du tout , là ... — Quoi ? »

‘Wait, there’s something going wrong out there... — What?’ (Gavalda 2004)

(19) « Qu' est-ce qu' y a ? — Je croyais que t' étais parti ... — Demain . — Quoi ? — Je pars demain . »

‘What’s wrong? — I thought you had left... — Tomorrow. — What? — I’m leaving tomorrow.’ (Gavalda 2004)

Pourquoi ‘why’. Two uses: question about reason/purpose (20) or about the speech act (‘why are you telling me that?’) (21)

(20) « Nina? Pourquoi tu te lèves ? — Oh ... Je vais lui donner son biberon ... — Mais pourquoi ? »

‘Nina? Why are you getting up? — Oh ... I'm going to give him his bottle ... — But why?’ (Gavalda 2000)

(21) « Si t' y tiens vraiment on ira chez Mimouni . T'es croyant ? » « Pourquoi ? » « Pasqu'en prime t' auras la bénédiction d' Allah ! »

‘If you really want to, we'll go to Mimouni. Are you a believer? — Why ? — Cause as a bonus you will have the blessing of Allah!’ (Lasaygues 1985)

RESULTS ON (ROOT) SLICES (BARE WH- WORDS)

Quoi > pourquoi > qui > comment > où > quand

Comment ‘how’. Two uses: question about manner (22) or speech act (pardon?) (23)

(22) Lorsque mes livres parurent , en octobre, je fus saisi d'une panique telle que je ne parlais que de les détruire. Mais comment ?

‘When my books got published, in October, I was seized with such panic that I was talking only about destroying them. But how ?’ (Althusser 1985)

(23) « On cherche l'emplacement du lac de La Dame, du Lac de Chandler . On pensait y aller demain... — Hein ? Quoi ? Comment ? J'en étais sûr ! »

‘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)

RESULTS ON WH+ÇA

Comment ça > où ça > qui ça > pourquoi ça > quand ça > quoi ça

Où ça ‘when’, **qui ça** ‘who’, **quand ça** ‘when’. The anaphoric particle *ça* helps sluicing.

Pourquoi ça ‘why’. Already associated with a propositional content without *ça*.

Comment ça ‘how’. More clarification questions ‘how come’ (24) than manner questions.

(24) «— Eduardo, son père ? ! Oh ! s'esclaffa-t-il avec une joie mauvaise. Mais ce n'est pas son père ! Oh, non !
— Comment ça ?»

‘Eduardo, his father? Oh! he laughed with evil joy. But it is not his father! Oh, no! — How-ça?’ (Férey 2012)

Very rare uses of **quoi ça** (redundant if reprise *quoi*) (25)

(25) « C'est quoi ça ? hurla le chef, et l'autre Quoi ça ? »

‘What is that ? yelled the chief, and the other What-ça?’ (Benoziglio 1974)

RESULTS ON SLICES WITH WH+ÇA

Root sluicing is the predominant use of wh+ça: 67%

- almost 100% with *pourquoi ça*, *quoi ça* and *quand ça*
- 85% with *où ça*
- 60% with *qui ça*

Sluice (decreasing frequency): **où ça > qui ça > comment ça > pourquoi ça > quand ça > quoi ça**

Comment ça : stripping more frequent than sluicing (clarification question on any preceding segment)

(26) « — Job a filmé l'existence entière de Matthias, c'est tout. — Comment ça, l'existence entière ? »
'Job filmed the whole existence de Matthias, that's all. — How-ça, the whole existence?' (Pennac 1992)

(27) « C'est pas un visage, c'est un chiffon. Ça nous est interdit, les visages. Comment ça interdit ? »
'It is not a face, it is a cloth. It is forbidden for us, the faces. How-ça forbidden?' (Ajar 1975)

RESULTS ON SLICES WITH AND WITHOUT +ÇA

Overt correlate with *wh+ça* slices: 45%

- almost 100% with *qui ça* and *quoi ça*

Sluice with overt correlate:

quoi ça > qui ça > quand ça > où ça > comment ça > pourquoi ça

Hypothesis: +*ça* helps anaphoric binding with an explicit correlate

Overt correlate with bare *wh*-slices: 18%

- 22% with *quoi*
- 47% with *qui*

WH-SLICES+ÇA AND BARE WH-SLICES WITH CORRELATES

wh-word	+ça			bare		
	total sluices	w/ correlate	%	total sluices	w/ correlate	%
comment ('how')	38	4	10,5%	26	0	0,0%
où ('where')	85	25	29,4%	17	7	41,2%
pourquoi ('why')	21	0	0,0%	122	2	1,6%
quand ('when')	11	4	36,4%	12	2	16,7%
qui ('who')	65	64	98,5%	53	25	47,2%
quoi ('what')	4	4	100,0%	129	28	21,7%
Total	224	101	45,1%	359	64	17,8%

Table 3: Proportions of wh-slucices+ça and wh-?-slucices with correlates

STATISTICAL MODELING

Logistic (binomial) model

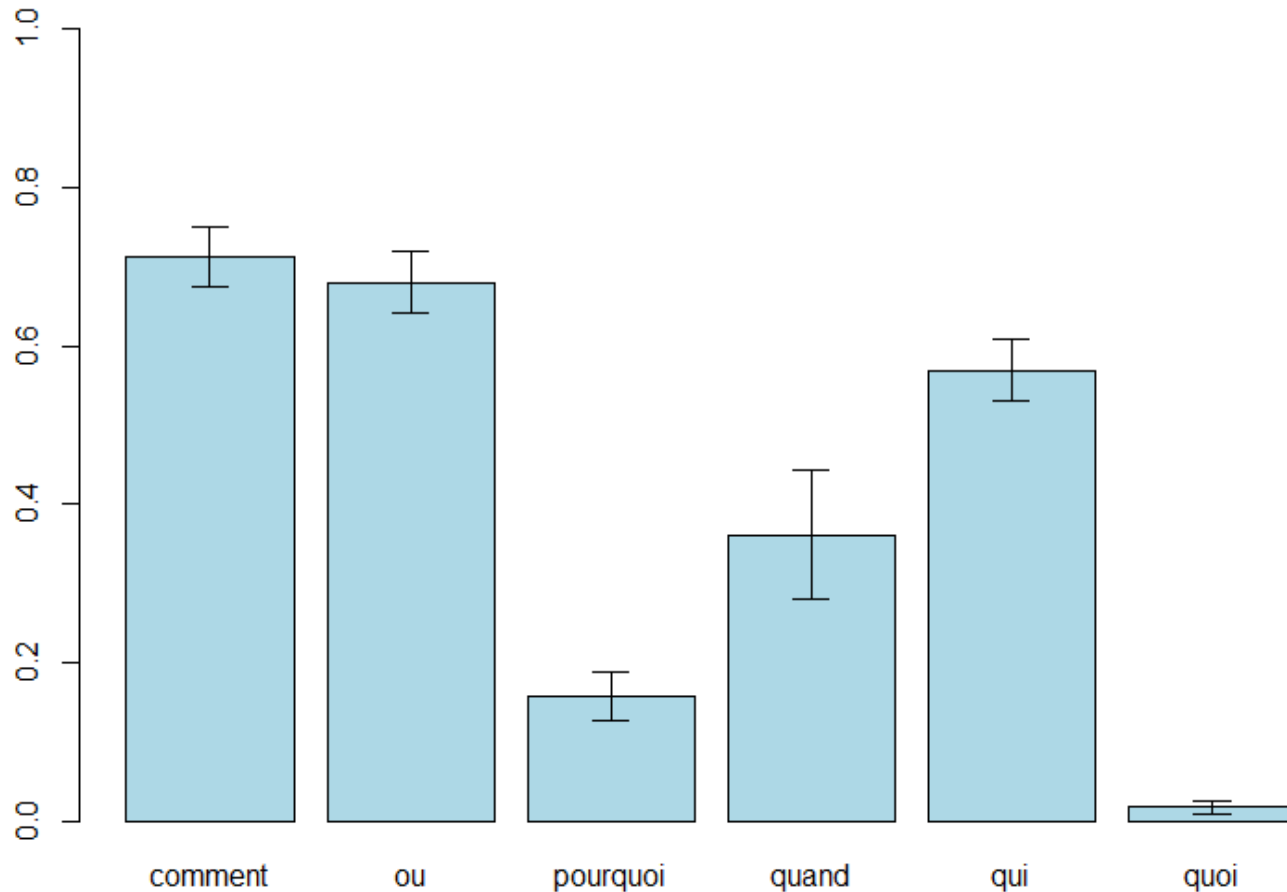
- allows to model dependent variable that gives outputs between 0 and 1 for all values of X (unlike linear regression model which can result in predicted values that are below 0 or above 1);
- we used lme4, sciplot, stats packages for R.

Logistic function :

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

where β_i is the weight corresponding to X_i estimated by using the maximum likelihood method.

STATISTICAL MODELLING



➤ *Ça* is most frequent with *comment* ('how'), *qui* ('who') and *où* ('where'), as opposed to *quoi* ('what').

- 71% comment *ça*
- 68% où *ça*
- 56% qui *ça*
- 36% quand *ça*
- 16% pourquoi *ça*
- 1% quoi *ça*

Figure 1. The proportion of *ça* depending on the wh-word.

STATISTICAL MODELLING

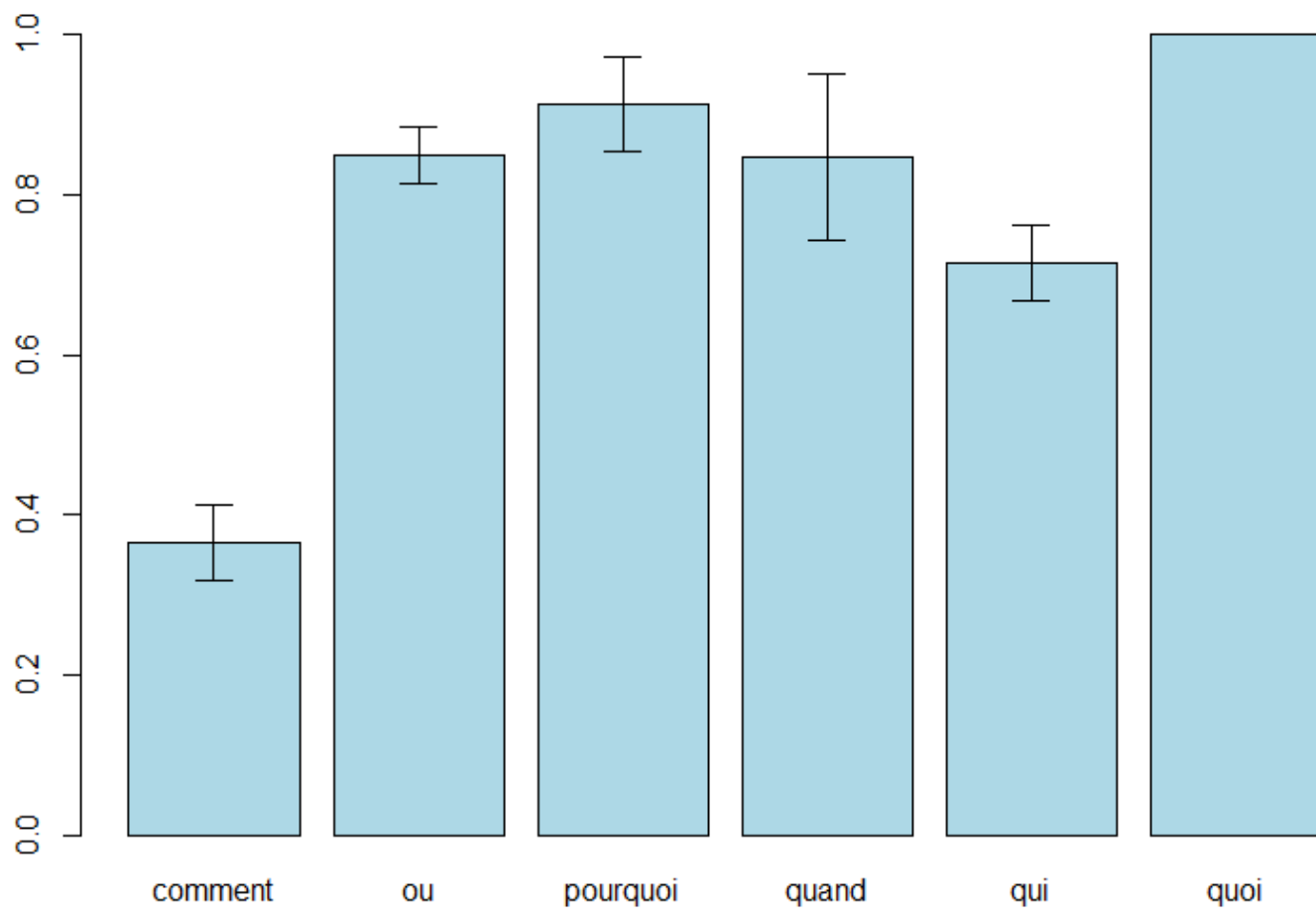
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.9067	0.1828	4.960	7.07e-07	***
où	-0.1517	0.2544	-0.596	0.550920	
pourquoi	-2.5834	0.2916	-8.859	< 2e-16	***
quand	-1.4773	0.3922	-3.767	0.000166	***
qui	-0.6300	0.2427	-2.596	0.009443	**
quoi	-4.9498	0.5364	-9.228	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Intercept = *comment* : the most frequent with ζa
- *Où* ζa : almost the same frequency as *comment* ζa
- All the other wh-words are less frequent with ζa compared to *comment*

SLUICING VS OTHER USES WITH WH+ÇA



Sluicing :

- 100% quoi ça
- 91% pourquoi ça
- 85 % où ça
- 85% quand ça
- 71 % qui ça
- 36% comment ça

Figure 2. Sluicing vs other uses depending on the wh-word + ça

SLUICING VS OTHER USES WITH WH+ÇA

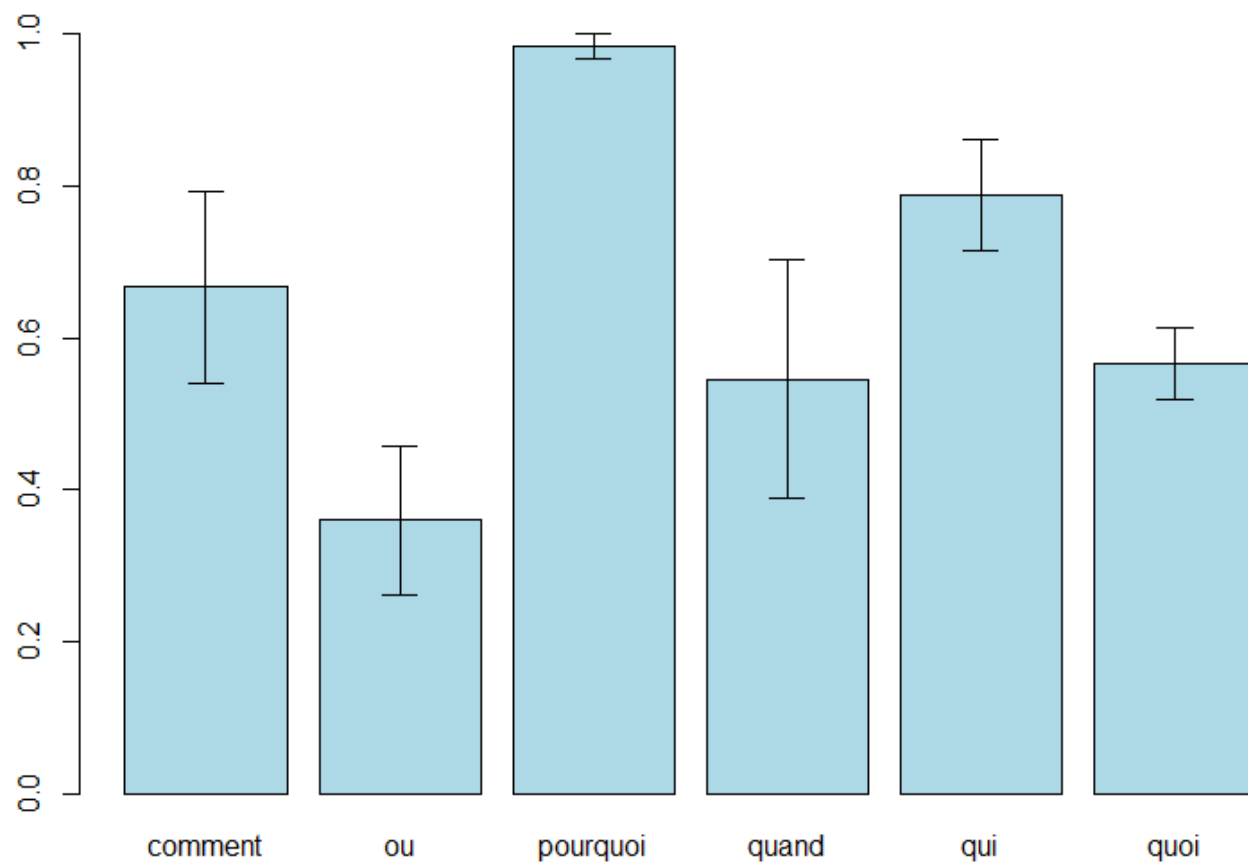
Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.5521	0.2036	-2.711	0.006707	**
MOT_QUou	2.2867	0.3463	6.604	4.01e-11	***
MOT_QUpourquoi	2.9034	0.7673	3.784	0.000154	***
MOT_QUquand	2.2568	0.7952	2.838	0.004540	**
MOT_QUqui	1.4684	0.3087	4.756	1.97e-06	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- *Quoi* is eliminated since 100% sluicing
- Intercept = *comment*
- *Où, pourquoi, quand, qui* + *ça* : higher frequency of stripping compared to *comment ça*

SLUICING VS IN SITU WITH BARE WH-WORD



In situ :

- 64% où ?
- 46% quand ?
- 44 % quoi ?
- 34% comment ?
- 22% qui ?
- 2% pourquoi?

Figure 3. Sluicing vs in situ depending on the wh-word

SLUICING VS IN SITU WITH BARE WH-WORD

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.6931	0.5477	1.266	0.20569	
où	-1.2685	0.6882	-1.843	0.06529	.
pourquoi	3.4177	1.1471	2.979	0.00289	**
quand	-0.5108	0.8165	-0.626	0.53156	
qui	0.6190	0.6938	0.892	0.37224	
quoi	-0.4261	0.5797	-0.735	0.46232	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Intercept = *comment*
- *Pourquoi ? Qui ?* : higher frequency compared to *comment* ?
- *Où ? Quand ? Quoi ?* : lower frequency compared to *comment* ?

CONCLUSIONS

- Wh+ça is possible in extraction;
- Main use of wh+ça is root sluicing (and stripping);
- Distribution of wh-words in sluicing is different with and without ça;
- More wh+ça-slucices with an overt correlate;
- To be explained by the semantics and pragmatics of the wh-word;
- To be explained by the anaphoric status of ça.

Future work:

- A closer look at sluices' source/antecedent
- More comparison with full questions
- Comparison with other discourse particles (*donc* 'then')
- Comparison with spoken corpora

WH-STRIPPING WITH WH+ÇA

The reprise fragment can be of any category:

an adjective(10a), a noun, a preposition(10b), even a weak pronoun(11a), an imperative (11b) or a whole sentence:

(10)a. —C'est curieux, fis-je en tournant mon regard vers la ville. —Pourquoi ça, curieux? ('Whyça curious?')(Del Castillo 1981)

b.—Avant, au moins, il y avait les mariages, les naissances, les vraies crises, les guerres, les révolutions, [...].

—Avant ? Quand ça, avant ? ('Before? When-ça, before?') (Sollers 1987)

(11) a.—Ils ont arrêté Paul. [...] —Qui ça, ils?('Who-ça,they?')(Clavel 1968)

b.—Laisse, dis-je. —Comment ça, laisse ? ('Leave (it), I said -How-ça, leave ?') (Pennac 1995)

There can be vehicle change (12a) or not (12b). Change to an embedded clause in (12a) and indefinite change in (12c).

(12) a.—Vous croyez ? dit le type, —Comment ça si je crois ?

('You believe (so)? the guy said -How-ça if I believe?') (Benoziglio 1974)

b .—:Qui est là? — C'est moi, c'est moi. — Qui ça,moi? ('Who-ça,me?') (Ionesco 1963)

c.—Elles ont emporté un certain nombre de caisses.—Quoi ! Comment ça, des caisses ?

('They took a certain number of boxes -What! How-ça, some boxes?') (Garat 2010)

MORE ON SLUICING WITH WH+ÇA

Most of the time, it is difficult to reconstruct a full verbal question

Some preposition omissions (French is NOT a preposition stranding language)

— *L'adresse de Rosine Portinari, tu l'as pas?* — **Qui ça?**

‘the address of Rosine Portinari, you dont have it ? — Who ça ?’ (Thérôme 1985)

⇒ *L'adresse de qui ça ?* ‘the address of whom ?’

Some nominal antecedents

A ses yeux, désormais, il n'y avait plus qu'une seule issue. Le départ des Juifs. Où ça? Ailleurs.

‘To his eyes there was only one issue. The Jews’ departure. Where-ça ? Elsewhere (Jonquet 1998)

⇒ ‘where should the Jews go ?’

Some ‘island’ contexts

Alors, qu'est-ce qu'ils font ?- Qui ça ? ‘so what do they do ? Who-ça?’

⇒ **Qu'est-ce que qui fait?* ‘*What who do ?’