The French Wh Interrogative System: 
Est-ce que, Clefting?

by

Sandrine Tailleur

A thesis submitted in conformity with the requirements 
for the degree of Doctorate in Philosophy

Department of Linguistics 
School of Graduate Studies
University of Toronto

© Copyright by Sandrine Tailleur 2013
Abstract

This thesis revisits the variation inherent to the French wh interrogative system. In Laurentien French, there are many ways to ask wh questions, all of which are said to have the same general interpretation. By looking at different types of data, historical as well as contemporary, this thesis puts forward three main findings/proposals: i. the high degree of variation is due to the use of wh est-ce que, which appeared in Old French as an interrogative cleft (Rouquier 2003); ii. between Old and Modern French, wh est-ce que has gone through a typical cycle of grammaticalisation (as described by Roberts & Roussou 2003, van Gelderen 2008a, b), while the free relative of the Old French cleft remained unchanged; iii. today’s LaF wh system is dominated by the wh est-ce que and variants (over 98 percent of use – Elsig 2009), which lack traditional wh movement. It is proposed that wh est-ce que and variants appear in a construction containing an atrophied clefted CP adjoined to a main clause containing a wh operator, whereas the variant traditionally called wh in situ is generated in a structure in which the wh element is interpreted and spelled-out in the position of Merge.

In addition to its theoretical contribution, this thesis helps to bring together theoretical and applied linguistics, since it makes use of different types of data, both historical and synchronic (oral and written corpora, experimental studies and grammaticality judgements). Moreover, the
conclusions raise important questions about the realities of diglossia in the French diaspora: wh interrogative variants are divided according to fundamental structural differences; some have wh movement (high, formal register) and others do not (vernacular and neutral register). Finally, this thesis also contributes to the theories of oralité (Gadet 1992), since it sheds light on a complex system of variants found exclusively in vernacular speech.
Acknowledgments

First and foremost, I would like to thank Yves Roberge for his good words, his help, his patience, which were constant for the three years during which he was my supervisor. Diane Massam and Ileana Paul were also always very encouraging and supportive; all three members of my committee were extremely helpful, and I am convinced that I could not have finished without them. I am also very grateful for the Linguistics Department at the University of Toronto. The atmosphere, created by both faculty and students, contributed to a working environment that is both stimulating and most motivating. A special thanks to Elizabeth Cowper, Jack Chambers, Keren Rice, Alana Johns and Elaine Gold for their special contribution to my well-being as a grad student. Anne-Marie Brousseau, Susana Bejar and, most of all, Arsalan Kahnemuyipour, have also showed me anything but encouragements and kindness.

My student colleagues; their constant presence in our lounge has been the excuse for many hours of serious and challenging, but also light and entertaining, conversations. A special thanks to Julia Sue, Maria Kyriakaki, Derek Denis, Chris Spahr, Alex Motut, James Byrnes, Eugenia Sue, Sarah Clarke, Liisa Duncan, Michelle St-Amour, Bridget Jankowski, Beth Macleod, Richard Compton, Kenji Oda, Will Oxford, Safieh Moghaddam, LeAnn Brown and Annick Morin, who each helped in their own ways, and were all equally important to me.

Merci à ma famille : mom, dad, Karl, Francine, Guillaume, et surtout Annie-Jo, qui ont toujours été là pour moi, même loin, même sans toujours comprendre ce que je faisais qui me rendait si nerveuse… Le sentiment de les avoir derrière moi, me soutenant peu importe mes décisions, est irremplaçable.

And finally, the special, special thank yous. I have already mentioned Yves Roberge, but he also deserves a special, special thank you. France Martineau, my mentor, my supervisor, my colleague and my friend; I owe her much of who I am professionally, and personally. Anne Mackay, who was always there for me, and who now understands me more than ever; Aude Patry et Alex Genois, for being such great friends (and great hosts!); Justin and Renee, for just being wonderful friends and neighbours; and, last but not least, Ailis Cournane, my friend, roommate, colleague, co-author, without whom none of this Toronto adventure would have been possible.
And Matt: thanks for being there, for understanding that the dirtier the house was, the more productive I was being... This thesis is part yours and you have been an important piece of it, from the very beginning.
# Table of Contents

Acknowledgments ........................................................................................................ iv

Table of Contents ........................................................................................................ vi

List of Tables ................................................................................................................ ix

List of Figures ............................................................................................................... xi

List of Appendices ....................................................................................................... xii

Chapter 1 ..................................................................................................................... 1

1 Introduction .............................................................................................................. 1

Chapter 2 ..................................................................................................................... 6

2 Description of the French Interrogative System ....................................................... 6

2.1 Types of French described .................................................................................. 6

2.1.1 Dialectal varieties ....................................................................................... 6

2.1.2 Types of data and registers ......................................................................... 7

2.1.3 Variable context .......................................................................................... 8

2.2 Laurentian French variants: their evolution ....................................................... 9

2.2.1 Wh fronting with inversion variants ........................................................... 11

2.2.2 Wh fronting without inversion variants ..................................................... 12

2.2.3 Wh fronting with reinforcer variants ......................................................... 14

2.2.3.1 Est-ce que variants: ALF data ............................................................... 24

2.2.3.2 Corpus study: wh+est-ce que vs. wh+complementizer ......................... 32

2.2.4 Wh in situ variants ....................................................................................... 41

2.2.4.1 Intervention effects and in situ .............................................................. 43

2.2.5 Embedded contexts ..................................................................................... 47

2.2.6 Summary: what does this all mean? ............................................................ 52

Chapter 3 ..................................................................................................................... 54
3 What is ‘est-ce que’? ........................................................................................................... 54
  3.1 The status of ‘est-ce/c’est’ .......................................................................................... 54
    3.1.1 Grammaticalisation of est-ce/c’est: from XP to X^0 ........................................... 57
    3.1.2 Contemporary French: actual status of est-ce/c’est ............................................ 63
  3.2 French declarative clefting ......................................................................................... 68
    3.2.1 Definition and evolution of clefting ................................................................. 69
    3.2.2 Interpretation of clefts and relatives ............................................................ 72
      3.2.2.1 Prepositions and interpretation ............................................................... 72
    3.2.3 Syntactic approaches to French clefts and relative clauses ......................... 76
  3.3 Summary ..................................................................................................................... 80
Chapter 4 ......................................................................................................................... 81
  4 Proposal – Accounting for Variation ........................................................................ 81
    4.1 Previous accounts of optional wh movement ................................................... 82
    4.2 Proposal ................................................................................................................ 86
      4.2.1 The structure of wh est-ce que interrogatives ............................................. 86
      4.2.2 The structure of wh-est-ce que interrogatives: diachronic evolution .......... 92
      4.2.3 Accounting for in situ .................................................................................... 108
        4.2.3.1 The in situ variant and negation ............................................................. 110
        4.2.3.2 Covert movement? ................................................................................. 111
          4.2.3.2.1 Island restrictions and adjuncts ......................................................... 111
          4.2.3.2.2 Simpson (2000): long distance licensing of wh in situ ................. 116
      4.2.3.3 Licensing wh in situ – interfacing syntax and prosody? ......................... 119
      4.2.3.4 Summary .................................................................................................. 122
    4.2.4 Distributional facts about est-ce que questions ............................................. 122
    4.2.5 Que or quoi? ..................................................................................................... 127
4.3 Typological considerations: French as an in situ language ........................................ 131
4.4 Interrogation in LaF: unified system? ......................................................................... 134
4.5 Sociolinguistic data and linguistic theory .................................................................. 136
  4.5.1 The diglossic approach ......................................................................................... 137
4.6 The syntax of speech: l’oralité .................................................................................. 142

Chapter 5 ....................................................................................................................... 145
5 Conclusion ..................................................................................................................... 145
References ....................................................................................................................... 148
Appendix ......................................................................................................................... 163
List of Tables

Table 1. Context of usage for each wh- word, with comp only and with est-ce que.................... 37

Table 2. Raw numbers for each wh- word with comp only and with est-ce que....................... 40

Table 3. Grammaticalisation process of ce in est-ce/c’est.................................................... 58

Table 4. Grammaticalisation of est and ce in terms of Feature Economy................................ 62

Table 5. Lexical element insertions between the wh word and the complementizer (by period)......... 66

Table 6. Vocabulary items available for insertion in two different stages of LaF grammar..... 127

Table 7. Variants of yes/no (total) interrogatives (based on Elsig 2009: 104, OH corpus (20th century) only).......................................................................................................................... 134

Table 8. Distinctions between French High and Low varieties (H vs. L).......................... 139

Table 9. Distribution of interrogative variants within the LaF diglossic situation............... 141
List of Figures

**Figure 1.** Axes of study .............................................................................................................................. 3

**Figure 2.** Map of inversion occurrences (including *est-ce que*) vs. pro-drop and subject-verb variant (Tuaillon 1975) ........................................................................................................................................ 26

**Figure 3.** Map of ‘verb as a second functional element’ variants (Tuaillon 1975) ................................. 28

**Figure 4.** Map of subject-verb variants (including wh+comp) (Tuaillon 1975) ...................................... 29

**Figure 5.** Evolution of the French interrogative system with *où* ‘where’ ............................................ 32

**Figure 6.** Order of attestation by wh- word .................................................................................................. 35

**Figure 7.** Order of attestation of wh *est-ce que* in Old and Middle French ........................................ 35

**Figure 8.** Graph of the distribution of wh *est-ce que* according to context and wh word ............ 39

**Figure 9.** Graph of the distribution of wh+comp according to context and wh word ...................... 39

**Figure 10.** Possible elements appearing in fronted wh- interrogatives in Laurentian French... 126

**Figure 11.** Representing a diglossic situation ............................................................................................ 140
List of Appendices

Appendix I. Complete list of contemporary LaF subject and object interrogative variants ..... 163
Chapter 1

1 Introduction

The French wh interrogative system exhibits a number of variants. A subset of these are exemplified in (1) to (6).

1. Où vas-tu?
   where go-you

2. Où tu vas?
   where you go

3. Où est-ce que tu vas?
   where is-it that you go

4. Où c’est que tu vas?
   where it-is that you go

5. Où que tu vas?
   where that you go

6. Tu vas où?
   you go where

‘Where are you going?’\(^1\)

French allows its wh to be fronted (1-5), to stay in situ (6), to be ‘reinforced’ with various elements (3-5), to appear bare (1, 2 and 6), and to trigger inversion of the subject and the verb (1), or not (2). There exist a number of publications concerning the French interrogative system, in particular wh interrogation (Elsig 2009, Lefebvre 1982, Mathieu 2004, Munaro, Polletto & Pollock 2001, among others). These studies range in focus from sociolinguistic evidence to theoretical syntax, but, crucially, only Elsig (2009) and Lefebvre (1982) have addressed the

---

\(^1\) The context of use and details of each of these variants will be discussed in Chapter 2.
system as a whole (many studies on certain particularities of the wh system in French were published by Quebec linguists during the 1980s; apart from Lefebvre (1982), we can mention Roberge (1983), (1986), and Barbarie (1982), among others). Ever since Cheng’s (1991) typological contribution, interrogative systems have generally been viewed crosslinguistically as binary; languages either allow wh movement, or they leave all wh words in situ. French, as seen above, allows both, which seems to contradict not only Cheng’s typology, but also the general Minimalist principle that states that optionality is not allowed (in this case, optionality of movement) (Chomsky 1995). Many researchers have focussed their work on the wh in situ variant, which is very unexpected in a language traditionally treated as having wh movement (Cheng & Rooryck 2000, Shlonsky 2009, Mathieu 2004, Boskovic 2000, to name a few).

The present study also aims to explain the unexpected behaviour of the French wh interrogative system, but in a different way. It has been shown that the most common of all wh constructions in Laurentien French (henceforth LaF) is est-ce que and its variants (examples 3-5 above); they represent over 80 percent of all usage of wh interrogatives (Elsig 2009, based on two oral corpora, more details in Chapter 2). In comparison, the in situ variant represents less than 10 percent of usage (Elsig 2009). Can this great variation in usage be attributed to the specialisation of each variant to very specific registers, or does it reflect a change in progress? I believe that by answering this question, I will be able to provide a more enlightened analysis of the system. Moreover, by considering the wh est-ce que the primary variant of the system, which is in line with the observed LaF usage, we may arrive at a different explanation from what has been proposed previously.

To answer our questions, we address the problem historically: where (and when) do all these variants originate? As illustrated in Figure 1, I adopt an approach characterised by two main axes of study. Synchronically, the whole system will be compared to crosslinguistic systems (the horizontal axis) to see how the wh system, as it is used in French today, compares with other possible systems of wh. Diachronically and language internally, the variants will be studied to verify how the system came to be what it is today.
Because the only data available to us from past languages are written texts, examples of questions in such texts – poetry, prose, administrative documents, religious documents, etc. – can be found only in a very limited number. The lack of extensive data from prior to the 20th century does not allow us to explain modern data in a satisfactory way. This fact is one of the main reasons why we have difficulty finding a unified analysis of the French wh system: the historical information is crucial to understand each variant. This work will look at each variant, both individually and with respect to the other possible variants, by assembling old and new information on their evolution and usage, and thus will provide an overview of the wh interrogative system, explaining its complexity.

This thesis contains three main parts; one exploring the evolution of all variants, answering the question: why so many?; one describing and explaining the behaviour of the est-ce que in wh interrogatives, as well as its declarative equivalent c’est; and finally, one proposing an analysis for the wh est-ce que construction, based on both diachronic facts and synchronic properties of the construction. This work as a whole provides a complete overview of the system, placing today’s most common variant at the center, to explain and formalise the variation.

Chapter 2 begins by establishing the geographic and linguistic limits of the variation described. We are mainly concerned with LaF, although all states of language from before the 20th century are from France, due solely to data availability. The evolution, usage, and main proposals concerned with each variant is then described. A big part of this chapter is also dedicated to a more in depth study of the wh complementizer variant (seen in 5), since it may be the case that this particular variant has a slightly different behaviour than the other wh est-ce que variants. We will see that wh est-ce que variants have existed since almost the beginning (i.e., Old French),
and they are now the majority variant used in LaF, while the fronted wh with inversion, the standard variant used in written registers, has virtually disappeared from the vernacular wh interrogative system.

To understand the realities of the majority variant, Chapter 3 details the properties of the *est-ce que* element. Although its grammaticalisation since Old French has been well attested (Rouquier 2002, 2003, Marchello-Nizia 1995, Elsig 2009, and others), very few theoretical details on its exact grammaticalisation trajectory have been provided (Druetta 2002, 2003). Using the theories of grammaticalisation cycles put forward by Roberts & Roussou (2003), who use Minimalist theories to explain grammaticalisation, we show that the wh *est-ce que* construction exhibits all the properties of a completed grammaticalisation cycle (section 3.1). I also address the issues of clefting and relative clauses in declarative contexts (section 3.2), since I am showing that the second part of a wh *est-ce que* question, i.e., the headless relative according to the proposals I follow (Clech-Darbon et al. 1999), has changed little if at all in many centuries of evolution.

Finally, Chapter 4 contains the proposal put forward following the facts unravelled in the two previous chapters. I begin by reviewing the previous analyses of the apparent optionality that is inherent to the LaF wh interrogative system, and then propose that the wh *est-ce que* variants in today’s LaF do not show wh movement. In place of movement, they have a complex adjoined element that is in fact a remnant of the Old French full cleft. I argue for a lack of covert movement (section 4.2.3), and show that the in situ variant is only a small reanalysis away from a construction such as wh *est-ce que*. Such a proposal makes the full LaF wh interrogative system very coherent and cohesive, which was not always the case in other contributions that treated the in situ variant as odd and unexpected. While proposing a non wh-movement system for French, this thesis provides the necessary evidence for such an innovative analysis to be accepted. Although the fact that French possesses a variant of clefted wh interrogatives is not new (Shlonsky 2009), the present work represents the first attempt to explain the variation inherent to the French wh system through a clefting analysis lacking traditional wh movement altogether.

This thesis adopts a methodology aimed at reconciling corpus-based linguistics with generative theory, which is an approach that has rarely been explored. Although most of the data, except for what is presented in section 2.2.3, comes from other sources, I have purposefully tried to base the
proposal on data taken from corpora and experiments (Elsig 2009, Adli 2006). This results in giving a picture of the system that is much more in line with the actual speakers’ usage. Moreover, this thesis contributes to the theory of oralité first put forward by Gadet (2007), since the variants described and explained are all taken from a vernacular register and are very specific to a certain state of language. We will see that we are putting forward a non-wh movement analysis only for the wh est-ce que construction and its variants; however, studies of usage have shown that those variants are used by LaF speakers over 98 percent of the time in conversational situations. This leads to the hypothesis that LaF probably only has the non-wh movement construction in its system, and that the system that still has wh movement and inversion is part of a standard, formal, ‘international’ system, separated from that of LaF (section 4.5).

This thesis contributes not only to the syntactic knowledge of the left periphery over time, but also to theories of diglossia and the study of the French language in general. Through the use of many different types of data, I hope to have been able to portray a faithful picture of the wh interrogative system in LaF and, just as importantly, to have shown that considering diachronic and synchronic facts is sometimes necessary to get the theoretical answers we seek.
Chapter 2

2 Description of the French Interrogative System

This section offers an in-depth look at the wh-interrogative system as it behaves today, based on new and established data. There is also a summary of analyses that have been proposed to account for the data structurally. The emphasis will be on the many different forms that the variant wh *est-ce que* can have; I will show that it is not all variants that are just ‘phonetic reductions’ of *est-ce que*. Indeed, section 2.2.3 shows evidence that some variants developed a function of their own, within the interrogative system but with influence from the embedded context and the clefting constructions. I will start by giving a complete description of all variants that are available to French speakers. For each variant, I give a description of its historical evolution and its usage, as well as an overview of how the variant has been analysed theoretically. To begin this section, I include a note on the type of French described throughout this thesis, since information on dialectal and regional variation is crucial to get a thorough understanding of the phenomena under study. There is also a small section describing the scope considered for this work: the constructions that are part of this study, as well as the constructions that have been left out for various reasons.

2.1 Types of French described

2.1.1 Dialectal varieties

This work will focus on variants of the wh system that are available to Laurentien French speakers. Laurentien French (henceforth LaF) has been described as the variety of Canadian French spoken in Quebec, mainly, and in the provinces that received their French population through an immigration wave that came from Quebec at the beginning of the 20th century: Ontario, Manitoba, Saskatchewan, Alberta, and, to a lesser extent, British Columbia (Mougeon & Beniak 1989, Baronian & Martineau 2009). The main variety described is the Quebec French variety, however it is assumed that the other Laurentian varieties do not differ in a dramatic way from Quebec French, and that the main differences found are due to the degree of contact with English and to language usage restrictions (see Mougeon & Beniak 1991 for the scale of language usage restriction). We do not have much information on the wh interrogative system used in Laurentian varieties west of the province of Quebec, but when the differences are
notable, they will be included in the descriptions of the variants. Likewise, when I have access to the information, I will include a few notes on the wh interrogative systems of other North American French varieties: Acadian varieties (French spoken in the Maritime provinces of Canada, New-Brunswick, Nova-Scotia, Prince-Edward Island, Newfoundland as well as Iles-de-la-Madeleine (QC)) and the Cajun varieties (French spoken in the Southern United States, mainly in Louisiana).

When describing the historical state of the language, however, the only variety of French that is available to us is the one from the European continent, used by scribes of different regions. We will include a few notes on regional characteristics whenever possible, but in general, when describing French before the 20th century, published texts from France are our references.

2.1.2 Types of data and registers

I will be working with different types of data, ranging from metalinguistic documents such as historical grammars, to modern spoken sociolinguistic interviews. I will rely as little as possible on judgements from native speakers, although judgements will inevitably be included, especially in the analysis section. I will instead try to focus on first hand data (corpora of written and spoken speech) whenever possible. The description, which is almost the entirety of Chapter 2, is concerned with the system as a whole, all registers included. However, special attention is given to spoken varieties, since oral data exhibit more variation than written data, at least for the variable under study. Given the limitations of historical data, comparisons will be made between historical written usage and modern spoken usage; whenever possible, a bridge will be established between these two types of data by looking at written genres known to represent, more or less faithfully, spoken language for the 18th and 19th century (just before oral data started being available), such as theater. Details of the types of data discussed will be given in each section.

It is important to note that describing any state of the French language is particularly challenging. French linguists often argue that French speakers have access to more than one ‘French’ grammars (Massot 2010, Zribi-Hertz 2011, Rowlett 2007). The gap between what is considered spoken or colloquial French and the written or standard variety, is especially important when it comes to the French language (compared to the situation of English, for instance). Rowlett (2007), for example, makes the distinction in his grammar between Modern French, which would
be the ‘artificial’ variety that was inherited from the Classical period, practically unchanged, and Contemporary French, which would be the more natural variety that has evolved throughout the twentieth century. Regardless of the name given to the two distinct French grammars, we have to keep this fact in mind when describing the French wh system in general, and usage in particular. For example, it is a well accepted fact that French speakers do not use the preverbal “ne” in negation anymore, nor do they use inversion when asking a wh question (in Contemporary French, to use Rowlett’s (2007) terminology). However, all French speakers learn in school that they have to use both of these elements in writing, and they are still in contact with registers of written or spoken language that would make use of the ‘ne’ or of the inversion variant (Rowlett’s (2007) Modern French). Of course this discordance between two systems is not unique to French, but it is definitely present to a greater extent than in other languages, hence the two separate complete grammars. I will discuss the consequences of such an assumption in Chapter 4 (section 4.5), where I will describe more extensively the diglossic approach put forward by Zribi-Hertz (2011).

2.1.3 Variable context

I will consider for this analysis all possible variants of wh interrogatives, uttered in matrix context (direct questions - 7). I will also consider those same – or slightly different – variants, uttered in embedded contexts (indirect questions - 8).

7. Quand est-ce que tu viens?
   when is-it that you(sg.) come
   ‘When are you coming?’

8. Tu m’as dit que tu venais quand?
   you(sg.) told me that you(sg.) came when
   ‘When did you tell me that you were coming?’

2 The exact status of embedded in situ is presented in section 2.2.5, since its interpretation is controversial.
Multiple wh questions, when many arguments are questioned within the same utterance (9), are left out of this analysis.

9. Quand est-ce que tu m’as dit que tu m’as donné quoi?
   when is-it that you me has said that you me-have given what
   ‘When did you tell me that you gave me what?’

It has been shown that such constructions are not acquired in the same way by children (Strik 2007), given their higher relative complexity. Therefore, I cannot assume that they have been through the same historical evolution as the other contexts (historical information on multiple wh is quasi-inexistant, since it is very rarely used). Moreover, the behaviour and restrictions on such questions are very different from the simple matrix wh questions; for instance, in situ is the norm for the second wh in such questions, as exemplified in (9) with quoi ‘what’. I refer the reader to Bošković (2002) and Jakubowicz & Strik (2008), among others, and I leave this issue for further work.

2.2 Laurentian French variants: their evolution

This section will go over each wh interrogative variant available to a LaF speaker. Each variant will be described thoroughly, with details on its evolution, its usage in matrix and embedded contexts, and its structure. I will conclude this section with a summary of the state of the system today, with an explanation, based on the facts described for each variant, as to why the system is now so complex. This section answers the obvious question: why so many variants?, and it provides a complete description of the evolution and usage of the different variants involved in the French wh interrogative system. We will see that the multitude of variants can be attributed to the parallel evolution of the wh est-ce que and the fronted wh with inversion, which have been ‘competing’ for a place in the interrogative system since Old French.

In order to get the necessary historical information, three main types of documents were used:

i. metalinguistic documents: historical grammars and dictionaries;

ii. linguistic atlas: “Atlas Linguistique de la France” (ALF) (Gillieron & Edmond 1902-1910);
iii. literary texts corpus: Frantext (ARTFL, University of Chicago).

The first type of document was consulted in order to determine which variants were available to speakers for each period of the history of French (Old French, Middle French, (pre-)Classical French and Modern French), and what contemporary grammarians were saying about the usage of each of these variants. The linguistic atlas search was conducted mainly to determine the state of the system at the turn of the twentieth century, which is important to distinguish the variants that are LaF innovations as opposed to the variants that came to Canada from France with the settlers during the New France period. The hypothesis is that if a variant is widespread in the ALF, it probably means that it entered the system before the settlers came to Canada, meaning that it is not a LaF innovation. It is important to note however that linguistic atlases are synchronic recordings of regional variation, and although they are a very important source to explain, or rather reconstruct, historical processes, data coming from this type of source need to be carefully contextualised. Linguistic atlases are more traditionally used to explain word formation and etymology, since most atlases are almost exclusively lexical (see Lauwers 1998), but studies such as Tuaillon (1975), which described the evolution of the interrogative variants with où, prove that it is possible to use atlases for morphosyntactic variables as well. Of course, any diachronic interpretations that are posited based on linguistic atlases are, although often quite accurate, only hypotheses. It is necessary to bring facts from other types of data to corroborate anything found in a linguistic atlas, which is why this work also includes a corpus study, a study of metalinguistic works, and a more typological investigation. Of course, when dealing with historical data, there is no medium that represents a perfect window to an ancient state of the language; it is only by combining different types of data that we can arrive at more conclusive statements.

This leads us to the last type of data consulted, the ‘real’ data taken from written corpora. Of course, since we are dealing with historical data, we have to work with a written, often quite formal, state of language. The corpus that was consulted for this study is the Frantext database, which is part of the "Project for American and French Research on the Treasury of the French Language (ARTFL)". It contains more than 2600 French texts of different genres, covering a

3 Tuaillon’s study is discussed in detail in section 2.2.3.1.
period from Medieval to Contemporary French. Given the findings from the two other types of sources, the search within the corpus was limited to the wh *est-ce que* and wh+complementizer variants only. This will become clearer following the presentation of results in the next section.

This section is divided in the following way: first, each variant – fronting with inversion, fronting without inversion, fronting with reinforcer and in situ – is presented with a summary of findings from the metalinguistic documents, from Old French to Modern French. Findings from (socio)historical linguistic research are also included in this part. Section 2.2.3.1 presents the ALF data, based on Tuaillon’s (1975) study of map 25, as well as original studies of all maps containing wh words. The following section (2.2.3.2) presents the results of the Frantext database search.

### 2.2.1 Wh fronting with inversion variants

The oldest and most standard of the possibilities for asking a wh question in French is the fronting of the wh, which triggers the inversion of the subject and the verb (complex inversion if the subject is a nominal DP, such as in 11).

10. Où va-t-il?
   where goes-t-he
   ‘Where is he going?’

11. Où Jean va-t-il?
   where John goes-t-he
   ‘Where is John going?’

4 Note that wh interrogatives that question the subject, such as *Qui est venu?* ‘Who came?’, do not behave in the same way as others. *Qui* ‘who’ has a special behaviour in French: it is incompatible with inversion, and it is not clear if it can stay in situ or not, since the subject is always pre-verbal. This type of questions will be ignored in the present chapter, but I will come back to this issue in Chapter 4, section 4.2.5.

5 The [t] is merely orthographic and phonetic. It appears with inversion, but only between two vowels to avoid hiatus.
12. Qui veux-tu voir?
   who want-you see
   ‘Who do you want to see?’

13. Quand irons-nous au théâtre?
   when go-we at theatre
   ‘When are we going to the theatre?’

In Old French, this way to ask a question was the only one available (est-ce que was also present, but it was only used in clefted questions, i.e. with a different interpretation, not in ‘regular’ ones (Rouquier 2002)).

This variant is nowadays very rare in oral LaF. Elsig (2009) reports that fronted wh with inversion corresponded to 2.5 percent of the total number of occurrences (24/959) in the corpus he studied (p. 147), which is composed of oral interviews with Ottawa-Hull French speakers (a LaF variety).6 We will return to the theoretical relevance of movement that triggers inversion in the analysis section. For now, we will focus on the variants that lack inversion, which are the ones that LaF speakers choose in their daily conversations. It is usually understood that the evolution of the French interrogative system is characterised by a regularisation to SVO word order (Rouquier 2002, Elsig 2009, Druetta 2003), which is why the inversion variants have now almost disappeared from the spoken language. All other variants – fronting, fronting with reinforcer and in situ – preserve the regular SVO word order.

2.2.2 Wh fronting without inversion variants

The second variant is the fronting of the wh word without inversion, and without any use of a ‘reinforcer’ such as est-ce que. On the surface, it looks like the wh word fronts somewhere within CP, checking the wh feature, leaving the declarative subject-verb word order untouched.

14. Où tu vas?
   where you go

6 Elsig uses two corpora hosted at the University of Ottawa: the Ottawa-Hull French corpus (Poplack 1989) and the corpus Récits du français québécois d’autrefois (Poplack & St-Amand 2007).
‘Where are you going?’

15. Qui t’as vu?
    who you-have seen
    ‘Who have you seen?’

16. Comment tu vas?
    how you go
    ‘How are you?’

17. Quand mes parents sont arrivés?
    when my parents are arrived
    ‘When did my parents arrive?’

This variant seems to be quite marked throughout the evolution of the French wh system, at least until the Modern French period. Ayres-Bennett (2004: 51) notes: “non-inversion [in partial interrogatives] therefore plays a marginal role”. She also mentions the possibility that this variant might be marked semantically, and have a function of confirmation and surprise or indignation (Prüßmann-Zemper 1986: 118-119, cited in Ayres-Bennett 2004: 51-52). However, Ayres-Bennett herself gives counterexamples where the non-inversion questions seem to be “straight questions seeking information”, from the speech of young Louis XIII (1605-1610) in Journal d’Héroard (Ernst 1985):

18. pouquoy **papa fai** cela
    why **daddy does** that
    ‘Why is Daddy doing that?’
    (14.8.1605, Ayres-Bennett 2004: 51)

Moreover, the corpus search that she performed confirmed that wh fronting without inversion was, in fact, marginal during this period; it represents 0.4 percent of all partial interrogatives

7 Note however that this example may not be ideal: Zuckerman & Hulk (2001) have shown that in today’s Quebec French, fronting without inversion is significantly more used by children acquiring language than by adults. It may therefore be the case that this example, which is the speech of a 6 to 10 year old child in the 17th century, is not representative of the adult speech for the same period.
occurrences in Héroard (2 out of 493), and 0.7 percent (2 out of 271) in her corpus of model dialogues (Ayres-Bennett 2004: 53) (fronting with reinforcers (est-ce que) represents between 6 and 7 percent of all occurrences).

As for how it is used today, it seems that although it cannot be qualified as ‘marginal’, its usage is far from representing the majority of occurrences. Elsig (2009: 147)) found that it represented 17 percent of all partial interrogatives in Ottawa-Hull French, and Coveney (2002: 189) found that it represented 26 percent of all occurrences in France French. The relative formality of this variant is also hard to determine in today’s LaF. Elsig considers that this variant is equivalent to the one with reinforcers, and that it only lacks the est-ce que. 8

2.2.3 Wh fronting with reinforcer variants

The third variant is the most common one in today’s spoken LaF. The wh word is fronted, but it does not trigger inversion, instead appearing alongside an “interrogative reinforcer”, traditionally est-ce que (lit. ‘is-it that’).

19. Où est-ce que tu vas?
   where is-it that you go
   ‘Where are you going?’

20. Qui est-ce qui vient demain?
   who is-it that comes tomorrow
   ‘Who is coming tomorrow?’

21. Quand est-ce que le travail sera fini?
   when is-it that the work will be done
   ‘When will the work be done?’

8 Note that Elsig (2009: 194) analyses all occurrences of wh+reinforcer as lexical variant of the wh. For him, comment is just a variant of comment-est-ce, for example. I will address his proposal more extensively in Chapter 3.
This variant is the most common only if we take into account that it includes all reduced and modified variants of the est-ce que, such as the complementizer only and the inverted c’est que (which are actually more common than the ‘traditional’ est-ce que – Elsig 2009: 152).

22. Où que tu vas?
   where that you go
   ‘Where are you going?’

23. Comment c’est que tu dis ça?
   how it-is that you say that
   ‘How do you say that?’

24. Qui (ce) qui vient demain?
   who (it) that come tomorrow
   ‘Who is coming tomorrow?’

25. Qu’est-ce tu veux?
   what-is-it you want
   ‘What do you want?’

Modern linguists seem to agree that in Old French, the meaning of the periphrastic est-ce que was one of stress, or emphasis (Foulet 1919, 1921; Moignet 1973, Jensen 1990, all cited in Rouquier 2002: 102). It was similar to a cleft construction, therefore not exactly in competition with the other interrogative variant of the time, wh+inversion. However, the special emphasis meaning seems to have disappeared by the 17th century; Fournier (1998) mentions a (rare) example of this meaning during the classical French period: “Indeed, for Thésée, qu’est-ce que j’entends? ‘what is it that I hear?’ is not a simple variant of ‘what do I hear?’, but should rather be interpreted as ‘what is the nature of what I am hearing? How can I classify what I am hearing?’” [my translation]9 (pp. 127-128).10 Importantly, she also adds: “This original value had

9 “Ainsi dans la bouche de Thésée, qu’est-ce que j’entends? n’est pas une simple variante de qu’entends-je?, mais doit s’interpréter comme : « quelle est la nature (impie, sacrilège) de ce que j’entends? quel nom donner à ce que j’entends? ».”
by then mostly disappeared, as the reinforced form was being used in a higher frequency (with any interrogative word)” [my translation]¹¹ (p. 128). I will not argue against this hypothesis, since the existence of this emphatic meaning has never been challenged by linguists; I will rather try to understand what happened after this period, once the construction had lost its marked meaning, and became a variant of regular wh- interrogation.

On the possible causes of the spread of this variant, Rouquier (2002) (and the authors she cites) argues that the development of est-ce que was linked to word order; it was a way to preserve the traditional SVO order, since it was no longer the predicate that bore the inversion, but the est-ce que construction itself.

“Il permet [le marqueur est-ce que] […], d’étendre davantage encore aux structures interrogatives l’ordre désormais dominant SVO : en effet, c’est est ce qui porte l’inversion et non plus le prédicat lui-même.”

“It [est-ce que] allows […], to spread the by then dominant SVO word order to interrogative structures: indeed, it is est-ce that bears the inversion, no longer the predicate itself.” [my translation]


“le développement de est-ce que comme marque interrogative intégrant un ce cataphorique atone, et qui se généralisera au XVIème siècle, permettra de maintenir l’ordre SN-SV en déclarant la modalité de la proposition à l’initiale de celle-ci, selon une tendance amenant à extraposer les morphèmes signifiant la modalité et à les dégager des relations d’incidence interne à la proposition.”

¹⁰ Fournier’s example could also be argued against, when taken out of context: does it really mean ‘what is the nature of what I am hearing?’, rather than more simply just ‘what am I hearing?’. If Fournier is right, this sentence would only push the disappearance of the emphatic meaning a century later than what is usually proposed. We leave this judgment to the reader’s discretion.

¹¹ “Cette valeur originelle est cependant largement estompée, à proportion de la fréquence d’emploi de la forme renforcée (avec n’importe quel terme interrogatif).”
“the development of *est-ce que* as an interrogative marker, which involves a cataphoric unstressed *ce* ‘it’, and which became generalised during the 16th century, makes it possible to maintain the NP-VP order by announcing the modality of the sentence at the beginning, following a tendency to strand morphemes expressing modality and to remove them from the internal proposition.” [my translation]

(Buridant 2000 : § 588)

*Est-ce que* in wh- questions appeared during the 12th century with *que* (*qu’*) ‘what’ and *qui* ‘who’, and the exact date of other attestations vary according to the wh- word (Rouquier 2002: 99). According to Rouquier, *quel* ‘which’ and *comment* ‘how’ are attested with *est-ce que* during the 12th century,12 *lequel* ‘which one’ during the 13th, *pourquoi* ‘why’ mid-14th, *où* ‘where’ at the end of the 15th, and finally *quand* ‘when’ appeared with *est-ce que* only at the end of the 16th century.13 Wh *est-ce que* questions never represented more than 9 percent of the wh-interrogatives before the 18th century (which is similar to what Kaiser (1980) found: 6.9 percent in the second half of the 15th century). This seems to contradict Brunot, who stated that periphrastic constructions became more and more used (*‘de plus en plus usuels’,* t. II: 319) as early as the 16th century. None of the studies mentioned included indirect interrogatives, which might partially explain the low numbers. Brunot seems to consider all instances of wh *est-ce que*, but unfortunately there are no numbers for indirect questions in historical French, although we know that they were possible, since we find examples (very few) in grammars. A more exhaustive quantitative study would be necessary to understand fully the use of the periphrastic variant with wh- constructions.

The appearance of *est-ce que* in yes/no questions is much more recent, since Rouquier (2002) found occurrences of it only during the 16th century (p. 98).

Middle French is the main period during which the periphrastic form extended its usage to a broader range of contexts. Brunot (1905-1938, vol. I) states that it is obvious in texts that the

12 During this period, the *ce* was not obligatory, which was linked to the possibility of pro-drop in Old French (Rouquier 2002: 100; Brunot t. I: 259).

13 Details of these findings by Rouquier are presented in section 2.2.3.2, when compared to the first attestation of the wh + complementizer variant in the 18th and 19th centuries.
periphrastic form will take over the simple (wh-) form (p. 449). He says that there are many occurrences at the beginning of the 14\textsuperscript{th} century, and by the 15\textsuperscript{th} they are everywhere:

"Au XV\textsuperscript{e}, on rencontre à foison les nouvelles formes, particulièrement dans le Myst. Du V. Test.: quesse que jo? (4771); qui esse qui m’a frappé (4763); qu’essem que vous avez? (4577); [...] ou esse que nous logerons? (Farce des gens nouv., Pic., Sot., I, 131); cf. Farce du Munyer (Fourn., Th., 167)."

“In the 15th century, we see the new forms everywhere, especially in the Myst. Du V. Test. [religious play mainly based on the Old Testament]: quesse que jo? (4771); qui esse qui m’a frappé (4763); qu’essem que vous avez? (4577); [...] ou esse que nous logerons? (Farce des gens nouv., Pic., Sot., I, 131); cf. Farce du Munyer (Fourn., Th., 167)."\textsuperscript{14} [my translation]

(Brunot 1905-1938, vol. I: 449)

This quote highlights the fact that the est-ce que variants are mainly associated with a spoken register, since they were found mainly in various types of plays, in which characters are speaking to each other, rather than in other types of written documents. Interestingly, he also mentions examples in non-interrogative contexts, and the examples that he gives are both in free relatives and in indirect questions. Rouquier’s arguments go in the same direction; est-ce que is by then used with more wh- words. However, she cites the results of the Kaiser corpus (Kaiser 1980, cited in Rouquier 2002: 101), which is a corpus containing farces from the second half of the 15\textsuperscript{th} century, 1450-1500. Kaiser found that the periphrastic construction accounts for only 6.9 percent of the total number of interrogatives, the majority of which were with the wh- words qui and que.

Clearly there are some discrepancies between the actual statistics that are available from sociolinguistic works and comments from historical grammarians. As seen above, it seems that Brunot is exaggerating the relative weight of this construction for this period, or at least that we should not interpret his words as meaning that this form is the majority one in Middle French; Kaiser’s numbers clearly show that although the form is used, it does not seem to be generalized yet;\textsuperscript{15} although no longer grammatically restricted, it seems that it still only appears in certain

\textsuperscript{14} Note that a farce is a type of play.

\textsuperscript{15} Unfortunately, Rouquier does not give any raw numbers for Kaiser’s corpus, so we do not know how many occurrences in total this 6.9 percent represents.
literary genres. Brunot is however not the only to mention the spread of *est-ce que* in Middle French. Marchello-Nizia (1995) states: “When it comes to the way of marking interrogation, two characteristics of Middle French are to be remembered: the wide usage of *qui* ‘who’, [...] but mainly the rapid development, at least in certain texts (mostly theater, but not exclusively), of the construction *...est-ce que*.”\(^{16}\) It therefore seems to be the case that although the construction does not represent the majority of the interrogatives in Middle French, the number of contexts in which it can appear still increases considerably (all wh-words, both direct and indirect interrogatives, etc.).

Fronted wh- with subject-verb inversion was still the dominant variant used in texts during the (pre-)Classical French period (16\(^{th}\)-17\(^{th}\) centuries). Most authors talk of the periphrastic form in terms suggesting that it was then a standard variant, now part of the interrogative system. We can find affirmations such as the following, both for the 17\(^{th}\) century: “They [the wh *est-ce que* variants] are extremely frequent, with all interrogative words and in all [literary] genres”\(^{17}\) [my translation] (Fournier 1998: 127); “This turn of phrase is still very much used in popular [vernacular] language” [my translation]\(^{18}\) (Haase 1935: 92). There are some coeval sources from this period, especially for the 17\(^{th}\) century, and they do not seem to give this form much thought, or if they do they certainly do not condemn its usage. According to Brunot, grammarians notice it in the 16\(^{th}\) century, without judging it badly (Maupas, Brunot t. II: 319), and at the end of the 17\(^{th}\) century, only the “grammairiens de second ordre” (t. IV: 868) ‘second-rate grammarians’ advise not to use it.

Ayres-Bennet (2004) offers a corpus study of interrogatives for the 17\(^{th}\) century. Her conclusions go in the same direction as what was mentioned above: the *est-ce que* form is quite widespread, but it is still far from being the majority variant for wh- interrogatives. She mentions (p. 55) that out of the 1220 examples of wh *est-ce que* questions that she found for the century as a whole,

\(^{16}\) "En ce qui concerne la façon de marquer l’interrogation, deux traits sont à retenir qui caractérisent le moyen français : l’emploi élargi de *qui*, [...] et surtout le développement rapide, au moins dans certains textes (œuvres théatrales essentiellement, mais pas uniquement), de la tournure *...est-ce que...?*.”

\(^{17}\) "Elles sont extrêmement fréquentes, avec tous les mots interrogatifs et dans tous les genres”.

\(^{18}\) “Cette tournure est encore très usitée dans le langage populaire.”
more than half (64 percent) are occurrences of *qu’est-ce que* ‘what-is-it-that’. The other wh-words are not nearly as common with *est-ce que*, but they still appear: *comment* ‘how’ (4.3 percent), *pourquoi* ‘why’ (4 percent), *où* ‘where’ (3.4 percent) and *combien* ‘how many/much’ (2 percent). The wh-questions represent approximately 80 percent of all questions with *est-ce que*. As for the comparison between the different variants that were available for wh-questions during the 17th century, it is still clearly the inversion that dominates, representing 91.5 percent of all occurrences. The periphrastic form accounts for 7.6 percent of her data, with the remaining percentage being the intonation variant (p. 58).

The same results are observable in Elsig’s study (2008) of wh-interrogatives from the 15th to the 17th centuries.19 Once again, inversion is the dominant variant, and wh *est-ce que* represents less than 9 percent, depending on the text (p. 144). What is interesting is that the two texts showing the highest percentage of use of the periphrastic construction are a farce from the 15th century and a parody from the 17th century;20 6.3 percent and 8.9 percent of wh *est-ce que* respectively (p. 144). This provides more evidence towards the hypothesis that this construction was associated with vernacular speech, which is why we do not see much of it in texts from this period. It seems to be the case that this variant originated from below, entering vernacular speech before being used in literary contexts. More evidence for this hypothesis will follow, when the results of the linguistic atlas and corpus studies are presented (sections 2.2.3.1 and 2.2.3.2).

In none of the studies consulted is there a mention of a possible reduction of the *est-ce que* form, except for Old French when the *ce* could be moved or deleted. It is possible that there were a few examples of them; however, even if there are any, they are probably not very numerous, because they did not elicit any comment from the researchers.

During the 18th century, some examples of orthography and usage could make us believe that a reanalysis might be on its way, and maybe even at least partially completed in the speech of the lower class:

19 Note that none of the texts that Elsig included in his corpus are used in Ayres-Bennet’s study; the two researchers have completely different corpora.

Ousque. – Louvet s’amuse à placer cette synérèse de où est-ce que dans la bouche d’une femme du peuple : « Ousque vous disiez, Monsieur? » (Faublas, IV, 150).

Ousque. ‘where-it-that’ – Louvet puts this form, an example of syneresis, in the mouth of a lower class woman: « Ousque ‘where-it-that’ you said, Sir? » (Faublas, IV, 150). [my translation] (from Brunot t. VI: 1437; example also cited in Séguin 1972: 79)

Brunot also gives a special description of the use of the periphrastic construction during the 18th century, and even if the description of the exact usage is quite vague, it still gives us a good idea of the spread of its use:

En dépit de toutes les prescriptions des puristes, ces formes gagnaient du terrain dans la langue écrite, à plus forte raison étaient-elles communes, on pourrait dire normales, dans la langue parlée. […] Plus populaires encore sont les tours conservés de l’ancien usage où l’ordre des mots reste normal, type : « A quoi que ça vous sert »? (Père Duch. Royal., Avis du P. Duch., p. 3.); […] « Je demande … si c’est qu’on veut faire du Louvre une ménagerie » (Jean-Bart, IX, p. 5). Rien ne montre mieux comment les règles qui enchaînaient la langue écrite restaient alors étrangères à l’usage courant. Il y avait des siècles que ces expressions s’étaient formées. On les trouve au XVIe, dans la haute poésie elle-même : « où c’est, helas! Où c’est que je voy nos tyrans » (Garnier, Porcie, act. II. La Nourrice); […] Condamnées, ces périphrases vécurent « en marge »; le poissard me manqua pas de s’en servir : « quoiqu’ tout ça veut dire »? « quoi qu’i me d’mande » (Vadé, La Grenouill., t. III, p. 291; 4e Bouq. poiss., Ib., p. 265.)?

“Despite all purists’ prescriptions, these forms were spreading in the written language, one could even say that they were common, almost normal, in the spoken language. […] Even more popular were the constructions from the old usage in which word order stays normal, like : « A quoi que ça vous sert »? (Père Duch. Royal., Avis du P. Duch., p. 3.); […] « Je demande … si c’est qu’on veut faire du Louvre une ménagerie » (Jean-Bart, IX, p. 5). No other construction can show better how rules that were governing written language could stay ignored by the common usage. It had been centuries since these constructions were formed. One could find them in the 16th century, in poetry : « où c’est, helas! Où c’est que je voy nos tyrans » (Garnier, Porcie, act. II. La Nourrice); […] Condemned, these ‘periphrases’ were marginal; of

---

21 It could also be argued as a case of syncope (segment dropped) instead of syneresis (segment fusion).
course, the poissard genre made good usage of them: « quoiqu’ tout ça veut dire »? « quoi qu’i me d’mande » (Vadé, La Grenouill., t. III, p. 291; 4° Bouq. poiss., Ib., p. 265.)?" [my translation].

(Brunot 1905-1938 vol.10: 361, emphasis my own)

We do not exactly know what he means when he mentions the “tours conservés de l’ancien usage”, because never before had anyone mentioned the wh- word appearing with only the complementizer, without the est-ce or c’est (bolded passage at the beginning of the quote). Brunot’s way of introducing the example makes it look like a very common one, although there are not much information available to corroborate his views (see below). In fact, we have seen that there were examples of fused orthography for this form as early as the 15th century, so by the 18th century, it is very likely that the grammaticalisation of the form was almost completed (we will explore the grammaticalisation of the form in detail in Chapter 3).

Concerning the wh+comp[lementizer] variant as exemplified in the quote above, there are very little evidence available to determine its provenance. What we do know however is that there was in Old French a higher number of conjunctions constructed with wh- words: Old French speakers could construct conjunctions such as quant que (or quancque, quanque) (Kuntsmann 1990: 14). Since we do not possess any document that officially mentions the possibility of the wh+comp construction before Modern French (18th c.), we are confronted with two opposing hypotheses: the modern wh+comp can be derived from the interrogative est-ce que (by phonetic reduction), or it could be derived from Old French conjunctions, introducers of embedded clauses. In other words, wh+comp, which can be today used both in embedded and in matrix contexts, could have originated in either of these two contexts. We do not have any studies which focus exclusively on these complex relative pronouns, although we know that the process wh word + que was quite productive in Old French. Kuntsmann (1990) gives us very useful information about their usage.

First of all, it seems that the complex relatives were not used as interrogative pronouns. The interrogative pronouns were similar to today’s inventory, although there were a few more, remnants of Latin’s inventory – qui, cui (meaning lequel ‘which’), que, quoi, dont (meaning où ‘where’), ou (also meaning où ‘where’), quel, lequel, quanz (meaning ‘how much’), combien (meaning ‘how much’ or ‘how big/small’), and a few rare/dialectal forms, quantel (same as combien), queien (meaning ‘which’), queinement (meaning comment ‘how’), quellement
(meaning ‘in which manner’). The difference between pronouns and adverbs is not clear at this stage, but for our purposes, it is not crucial to make the exact distinction, since there is only one position for wh words in today’s LaF. So it seems that the wh+comp construction could not be used as a direct interrogative pronoun/adverb, only as a conjunction.

Another difficulty that presents itself when looking at wh+comp in written forms of older versions of French (and one that is probably behind the fact that grammarians like Brunot are not clear when it comes to the wh+comp variant), is that there was the possibility, still present today in fixed expressions and written registers, to have the complementizer *que*, often following immediately a wh word, introducing a clause in the subjunctive mode.

26. *Je ne sais quoi que ce soit.*

I neg know what-that-it-would be

‘I don’t know anything.’

This version of wh+comp is probably derived from Old French wh+*que* conjunctions, but it does not correspond to the interrogative variant that we are studying here. Today, examples like (26) are only used in indefinite contexts, whereas the interrogative version, direct or indirect, of wh+comp is most of the time very specific, and does not require the subjunctive mode, as in (27).

27. *Dis-moi un peu comment que tu t'y prendrais!*

tell-me a little how that you yourself-cl. take

‘Tell me how you would address this!’

(Huysmans, J.K., 1879, *Les sœurs Vatard*. – Frantext)

In the example above, the complementizer is clearly redundant, fullfilling no grammatical function. In (26), the *que* introduces an indefinite clause, which is prescriptively conjugated in the subjunctive mode. Examples such as (26) are very common in Old French, therefore a search in Old French document would have to take into account the fact that occurrences like (26) would have to be discarded, since they are not instances of the construction that interests us. Section 2.2.3.2 aims at comparing the use of the wh+comp variant with the more standard wh *est-ce que* in interrogative contexts; I hope in this way to provide more evidence towards one of the two hypotheses outlined above as to the origins of the wh+comp variant.
The next two sections present the findings taken from a linguistic atlas and from a corpus search, but the period examined is only (early)Modern French; in Old French, the wh est-ce que was not in competition with the wh+comp, it was not even in competition with a regular wh interrogative, since, as mentioned earlier in this chapter, it fulfilled a very specific function of emphasis that the fronted wh with inversion could not carry out. More details are given in the following sections.

2.2.3.1 Est-ce que variants: ALF data

Since most of the historical information that we have on the state of the language before the 20th century is from written, mostly published, texts, it is crucial to look at other types of data to complete the picture of the phenomenon under study. Moreover, we have seen in the previous section that there is evidence to suggest that the wh est-ce que construction was associated very early with vernacular speech. As a consequence, we looked at data from the Atlas Linguistique de la France (Gilliéron & Emond 1902-1910) (ALF), and we are fortunate enough to have had access to a detailed study that was published on one of the maps that interest us: Map 25, Où vas-tu?. Tuaillon’s (1975) analysis of this particular linguistic atlas map provides a number of implications relevant to the development of the interrogative system. This section gives a summary of Tuaillon’s proposals, along with our own findings using other types of data.

One cannot straightforwardly determine which variant of a pool of variants is the most used by speakers of a certain dialect using only data from a linguistic atlas. As Tuaillon (1975) puts it, it all depends on the distribution of the data. If regular patterns are observable, it is very likely that the form given by the speaker at a certain point on the map really is the more widely used variant (p. 83). The following description of Tuaillon’s observations will provide much evidence in favour of the idea that dialectal geography data can give invaluable diachronic linguistic information.

Map 25 from the ALF represents the answer to the question ‘how do you say where are you going in your region?’. Unsurprisingly, the informants’ answers vary a lot, and by examining the word order of the various answers, Tuaillon established an order of evolution for each determined variant. He divided all the variants into three main groups: the Occitan variants (without any subject pronouns), the ‘verb-second’ variants, and the subject-verb variants. For the purpose of this work, I will leave aside his description of the Occitan variants, for which I give
an example in (28), because such variants are not used today in the parts of the French Diaspora that concern us here (see Elsig 2009).

28. Où vas? (Occitan, Nièvre, Wallonie)
   where going(2\textsuperscript{nd} sing)
   ‘Where are you going?’ (Tuaillon 1975: 81)

The ones that interest us are the variants from the other two groups, which vary according to the presence or absence of subject-verb inversion. The second group of variants includes all the occurrences in which the verb occupies the ‘second functional position’ (“deuxième unité fonctionnelle”, p. 90), which means that the verb immediately follows the wh- pronoun. Note that he considers the element est in est-ce que a verb, so wh est-ce que occurrences are also included in this group. The two structural variants in this group are subject-verb inversion (exemplified in (29)) and wh est-ce que forms (exemplified in (30)).

29. Où vas-tu?
   where go-you
   ‘Where are you going?’

30. Où est-ce que…? (Tuaillon 1975: 91)
   where is-it that…
   ‘Where (are you going)?’

Finally, the rest of the variants fit into the category ‘subject-verb’, and they are illustrated in examples (31) to (34).

31. Où que tu vas?
   where that you go

32. Où que c’est que tu vas?
   where that it-is that you go

33. Où ce que tu vas?
   where (it) that you go
34. Où tu vas?
   where you go
   ‘Where are you going?’ (Tuallion 1975: 95)

In total, excluding the Occitan subjectless occurrences, he analysed six different variants. It is interesting to note that those variants are not different from what we find today in various dialects of French (see Lefebvre 1982, Elsig 2009). He tried to determine the evolution of the variable, which one(s) is/are conservative and which one(s) is/are innovative. Figure 2 below shows the map presenting the repartition of the inversion occurrences in regard to the other variants.

**Figure 2. Map of inversion occurrences (including est-ce que) vs. pro-drop and subject-verb variants**

Tuallion 1975: 80 (Carte I)
The diagonally striped areas represent regions where we find the inversion variant *ou vas-tu* and *ou est-ce que tu vas*. The vertically lined area represents pro-drop regions, and the white area represents regions where we find different variants of subject-verb order.

As we can see, the division is quite clear: the South is faithful to the Occitan subjectless characteristics, the North has a general subject-verb order, and the inversion variants occur in the central lateral peripheries. Such a clear division between the forms suggests that the distribution is not accidental. If all variants were used in an equivalent way, we could not observe such a clearly divided map. Since the subject-verb sequence occupies the whole Oïl domain (including Paris), it is probably safe to say that it is in fact the dominating and spreading variant. The conservative variant, subject-verb inversion, is found only in peripheries (“elle forme des aires marginales cohérentes” (p. 82)). As Tuaillon puts it: “une parfaite égalité d’occurrence n’aurait pas pu donner un aspect géographique aussi ordonné et aussi intelligible” (p. 83) (‘an equivalent usage [between different variants] could not have given such an ordered and coherent geographical space’ [my translation]).

The next map shows the details of the occurrences which Tuaillon calls “verb as a second functional element”. It examines in detail the distribution of all the variants included in this category, including the wh *est-ce que* variants.

---

22 Note that this map does not show all the variants described above. The inversion occurrences are only of the ‘*ou vas-tu*’ type, and the subject-verb (called “séquence progressive” by Tuaillon (p. 80)) occurrences include the variants exemplified in (31), (32), (33) and (34).

23 “It [the subject-verb inversion variant] constitutes marginal areas located in an orderly way” [my translation].
Figure 3. Map of ‘verb as a second functional element’ variants

Once again, the diagonally striped parts represent subject-verb inversion regions (as exemplified in (29) above), but in this case, the wh est-ce que variant is in the horizontally striped areas. The vertical lines are the pro-drop regions, and the circled communities are the ones that gave Où ce que tu vas? as an answer. This map shows very interesting facts, especially when it comes to the variants with est-ce que. We can see that where we find où est-ce que are places that are close to a region that uses inversion. For example, we find in Lorraine (north-west) a great number of est-ce que forms, and this region is specifically in between two areas where inversion is still dominant (Tuaillon 1975: 94). For Tuaillon, this means that the est-ce que construction ‘is the link’ between the verb-subject and subject-verb variants: it is the intermediate construction. Therefore, this map suggests that on the French interrogation system timeline, the form with est-ce que comes right after the form with subject-verb inversion, which is quite compatible with what we have described in the previous section.

Finally, the last of Tuaillon’s maps reproduced here gives us the details of the dominant variants.
Figure 4. Map of subject-verb variants (including wh+comp)

In this map, the area covered with vertical lines represents the regions where the wh+comp (où que tu vas?) is used. The circled municipalities are the ones using où que c’est que tu vas?; the checkered regions are the ones using où ce que tu vas?; the white regions are the inversion ones (either où vas-tu? or où est-ce que tu vas?); and the few regions covered with diagonal lines are the fronted wh- without inversion, or intonation variant, où tu vas?.

This map is by far the most revealing for our purposes. It shows that the variant où que tu vas, i.e. the wh+comp form, is the most common one in France at the end of the 19th, beginning of the 20th century.\(^{24}\) Let us now examine each of the variants of the map in detail.

---

\(^{24}\) If we think about the origins of the variant in Canada, this is a major fact; if this form was the most spread one during this period, especially in the central and northern parts of the country, it might be
First, the circled variant, the reinforced *ou que c’est que tu vas?*, is located exclusively in the central northern part of the country. Tuaillon does not comment on the origin of this variant, but it is a variant that is also quite common in Canada today (Lefebvre 1982). Since a good number of migrants from this area came to colonize New France, this variant might have been transferred to Canada during this period. Unfortunately, the map gives no information that would allow us to confirm this hypothesis, so we will leave the question open for the moment.

The checkered variant, *ou ce que tu vas?*, is also only found in isolated pockets, but contrary to the preceding variant, we can find it all over the country, except in the Occitan region. Tuaillon (p. 92) brings up the question of whether this variant comes from the majority variant *ou que tu vas?*, with only the addition of some sort of antecedent to the *que*, or if it is a reduction of the *ou est-ce que tu vas?* form. His answer is that, at least in the Eastern part of the region, the *ou ce que* variant is found near *ou est-ce que* regions, and that it might mean that the two are related, but since there are pockets in the westernmost part of the country without any *est-ce que* form in the neighbouring regions, there is no way to confirm this hypothesis. Indeed, it seems to me that the absence of the longer form in the westernmost regions calls into question the assumed reduction relationship between the two constructions. The evidence is at any rate not sufficient to support such a claim.

The more simple subject-verb variant, *ou tu vas?*, is found all over, but in isolated pockets. It is quite unpredictable, and even appears in the middle of inversion regions. Tuaillon (p. 95) states that this variant, subject-verb without any *mot-outil* ‘tool-word’, is a more recent one, along with the in-situ form which, although we hear it quite often today, especially in France, does not appear on the linguistic map from the beginning of the century. We have seen however in the previous section that this non-inversion variant was in fact found in the seventeenth century, albeit marginally. The emergence of this particular variant does in fact date back from a couple of centuries earlier than the in situ variant.

---

evidence for the hypothesis that French people immigrating to Canada during the New France period brought this variant with them. Therefore, it would not constitute a Canadian innovation.
Finally, there is the majority variant, où que tu vas?, with que, which Tuaillon calls the mot-outil, because it does not really have any meaning or grammatical function, as mentioned earlier. It is hard to deduce where this particular form comes from, because since it is found in a very large region of the country, it is in close proximity with all the other variants. Moreover, it is a variant that has never really spread into more formal speech, so it is difficult to find in written language. It is somewhat unusual that a variant can be so spread geographically, even if it has never entered a more formal state of language. Its geographic spread at the beginning of the century suggests that it is probably quite old, but given its familiar status, it is very hard to track its evolution; we just cannot find occurrences of it in older historical texts, as we saw in the previous section. This fact is the main reason that a corpus search was conducted; since wh+complementizer is so widespread in the ALF, where does it come from, and when exactly did it start to be used? If the non-inversion variant was there marginally during the 17th century but is still not very common in the ALF over two centuries later, what can we say about wh+comp? The next section answers these questions, where the results of the Frantext database search are presented.

Although Tuaillon does not mention it explicitly, his article lets us assume that all forms are derived from each other, or that they all come from the same source form. However, it is difficult to confirm using only the information contained in the map. All we can deduce is the order of appearance of each of these forms, and the question of whether they are all the results of the evolution of a unique variant will have to be answered using other kinds of data. Right now, we can say that based on our findings on the inversion and the est-ce que variants, the latter definitely does not come from the former. They appeared contemporaneously during the Old French period, fulfilling different functions. As for the non-inversion variant, it is hard to know if it really is, in fact, a non-inversion variant, or if it is instead a non-est-ce que variant. As mentioned earlier, Elsig (2009:194) proposes the latter option, but I will reserve judgement until I present my proposal in Chapter 4.

Figure 5 shows the timeline of the order of appearance of each of the forms that we find on the map. The one in parentheses is the one for which we have some evidence, but we would need to verify its status using different kind of data.
Figure 5. Evolution of the French interrogative system with où ‘where’

subject-verb inversion > wh- est-ce que > (wh- ce que) > wh- que > wh- subject-verb > subject-verb wh-

As for the other wh words, their attestation in the ALF for each of these variants varies greatly according to the nature of the wh word. There is no certainty that all wh words would follow the same path, although the main variants, that is, inversion, est-ce que and in situ, are assumed to have appeared in the same order for all wh words, although not necessarily during exactly the same period.

The next section provides more information on the status of the wh est-ce que and wh comp variants at the time where they started to be in official competition in texts.

2.2.3.2 Corpus study: wh+est-ce que vs. wh+complementizer

Following the findings from the ALF and from metalinguistic documents, it is appropriate to include data from a written corpus, in order to measure the usage based on actual data. A search was conducted with two main goals in mind:

i. to find the first attestation of the wh+comp construction in written texts for each wh word;
ii. to compare its usage with the usage of wh est-ce que.

The extraction was conducted in the following way: each wh word was searched for accompanied by each version of the complementizer: que, qu’ and qui (when it applied). The total number of occurrences found in the search was recorded, and each occurrence was sorted. The occurrences preceding a verb in the subjunctive mood were rejected, as well as the ones in which the wh word and the complementizer were separated by punctuation. Of course searching in this way will not allow us to determine if there can be any elements between the wh word and the complementizer, but given the limited number of occurrences retrieved, it is fair to believe that if those two elements can in fact be separated by something, it would be quite rare to find in texts, and would not significantly affect the overall results. For comparison reasons, the same

---

25 In this figure, ‘x > y’ means ‘x comes before y’ or ‘x appears earlier than y’.
method was used to extract all occurrences of wh est-ce que; each wh word was searched for accompanied by est-ce que, est-ce qu’ and est-ce qui (where it applied).\textsuperscript{26}

The context of variation was carefully delimited. The two interrogative contexts – direct and indirect – are fairly straightforward. An example of each is provided in (35) and (36) respectively.

35. Où que tu vas?
   where that you go
   ‘Where are you going?’

36. Je me demande où qu’il va.\textsuperscript{27}
   I myself wonder where that-he goes
   ‘I am wondering where he is going.’

Those contexts always allow all the different variants to appear: the est-ce que variant, the ce que variant, the bare variant (only the wh-pronoun) as well as the complementizer only variant. For the moment, I will consider all wh-pronouns as being equivalent,\textsuperscript{28} and I will later explore the possibility of some of them favouring the appearance of the complementizer (or other variants) over others. The two contexts represented in (35) and (36) form together what I will call the interrogative context, since in both cases the wh element bears a wh feature that needs to be checked. Also note that direct questions as in (35) represent the only matrix context considered; the two other contexts presented are in embedded positions ((36) and (37)).

The other context considered here is the one of free relatives, as in (37).

\textsuperscript{26} Only the occurences in which the reinforcer appears in its full form, est-ce que, were retrieved. We do not have any results on variants of est-ce que such as c’est que, ce que, etc.

\textsuperscript{27} Note that the appearance of the complementizer in such contexts is associated with a more informal register, which means that the subject clitic pronoun il would probably be pronounced as [i], with the final consonant not pronounced.

\textsuperscript{28} I exclude que, since it never appears by itself as a wh-element in French, neither in matrix nor in embedded contexts.
37. Tu vas où que tu veux.
   you go where that you want
   You go wherever you want.

A similar construction, or one that might appear similar, is the full (or bound) relative, as in (38).

38. L’homme que j’ai vu.
   the-man that I-have saw
   The man that I saw.

Structurally, all these constructions are similar in that they project a CP. Moreover, the contexts in (36) and (37) do not have an antecedent in the sentence, as opposed to full relatives, which have to refer back to something specific (for a summary with examples of our three contexts and the full relatives, see King 1991 and Starets 2002). I will therefore not consider the full relatives, as in (38), to be equivalent contexts of variation (however, we will come back to the comparison between ‘full’ relatives (38) and free relatives (37) in Chapters 3 and 4). The free relatives and the indirect questions can be quite similar. We can distinguish the indirect questions by the verb of the matrix clause; it has to be a verb that permits an embedded question (or question-like statement) as its complement, such as dire “tell”, savoir “know”, demander “ask”, ignorer “not knowing”, etc.

The first results presented here (figure 6) are the dates of attestation with only the complementizer, for each wh- word. We can see that it takes less than 150 years for the construction to be used with all wh- words (vs. wh est-ce que that took over 400 years). Où ‘where’ appears first, then 50 years later combien ‘how much/many’, comment ‘how’, qui ‘who’ and pourquoi ‘why’, all attested at dates very close to each other. Finally quand ‘when’ gets attested with the complementizer at the beginning of the 20th century.
Figure 6. Order of attestation by wh- word

<table>
<thead>
<tr>
<th>Year</th>
<th>1784</th>
<th>1837</th>
<th>1842</th>
<th>1844</th>
<th>1850</th>
<th>1904</th>
</tr>
</thead>
<tbody>
<tr>
<td>où que</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>combien que</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comment que</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>qui que</td>
<td>----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pourquoi que</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quand que</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This order is slightly different from the order of attestation seen with the periphrastic *est-ce que* in Old and Middle French (Rouquier 2002: 99-101), see figure 7 below, although, once again, *quand* is the last one to be attested with the periphrastic form.

Figure 7. Order of attestation of wh *est-ce que* in Old and Middle French

<table>
<thead>
<tr>
<th>Century</th>
<th>12th c.</th>
<th>13th c.</th>
<th>Mid-14th c.</th>
<th>Early 15th c.</th>
<th>Late 15th c.</th>
<th>Late 16th c.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quel</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lequel</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pourquoi</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comment</td>
<td>Où</td>
<td>Quand</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The same was true in the study of the ALF maps in the previous section; wh+comp was more common with *quel*, *où* and *qui*, and was quite marginal with *quand*. This is an interesting observation, because different types of sources/data all converge on the same conclusion; wh+comp entered the language at different times for each wh word, but it took less time to spread to all of them than it did for the wh *est-ce que* construction.

Here is an example of one of the first attestations found for wh+comp, with *où* (Restif de La Bretonne, 1784: Frantext): 29

29 The occurrences extracted from Frantext have been transcribed as they appear in the database. No orthographical or punctuation changes have been made.
39. Mais si on a laissé sortir la fille le soir, alors le garson l’approche en-câlinant : où qu’vou allez donc, Jeanne? Donner de la pâille à nos vaches… j’vas donc vou ainder?

‘But if we let the girl go out at night, then the boy approaches her slowly: where are you going, Jeanne? To give some hay to the cows... – Can I help you?’

We find the occurrence in a dialogue, which is quite standard throughout the whole extracted corpus. The author, De la Bretonne (1734-1803), grew up as a peasant before starting a career as a printer, then a writer. His work often features lower class characters, which is probably why we find occurrences of the wh+comp construction in some of his novels. What is interesting is that in the same novel from which the quote in (39) was taken, we also find the following occurrences:

40. Et pourquoi se lier irrevocablement à une femme, par-exemple, avant l’âge qui qui nous rend habitudinaires? N’est-ce pas de gaîté-de-cœur, chercher un repentir?

‘And why get attached to a woman, for example, before the age that gets us “habitual”? Isn’t it to joyfully find a repentance?’

41. Je voudrais bien savoir quelle est ta politique avec tous les hommes? Je tiens la mienne de ma feue bellesoeur Manon, qui qui m’a très-bien endoctrinée pendant le peu de temps que j’ai vécu avec elle.

‘I would like to know what is your policy with all the men? I owe mine to my late sister-in-law Manon, who indoctrinated me during the short time that I lived with her.’

It seems that here, even though he used it in a naive and “incorrect” way, the author knew that the “double qui” was possible in lower class speech. Because I found two occurrences of it, I put aside the possibility that it could be a typographical error. I did not count these two occurrences in the total of tokens retrieved, but they represent a possibility that qui was first attested in 1784 as well, as early as où. Although these are qualitative remarks that can in no way be quantified due to their marginality, they may be a good indication that these forms existed in spoken language decades before they started to enter the written language, which would explain their wide distribution in the ALF at the turn of the 20th century.
The other results that will be presented here concern the context in which wh+comp and wh est-ce que are used in literary texts from mid-18th to mid-20th centuries.

Table 1. Context of usage for each wh- word, with comp only and with est-ce que

<table>
<thead>
<tr>
<th></th>
<th>DQ</th>
<th>IQ</th>
<th>FR</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>how much</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>how</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>where</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>why</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>who</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>when</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: DQ= Direct questions; IQ=Indirect questions; FR=Free relatives

Table 1 contains all of our findings to present a general view of the phenomenon. A closer look at this table reveals a few noteworthy trends. Ousque is the only ‘intermediate’ form presented in the data, and that is because it was the only one that could be found in the corpus under study.

\[30\] For all searches with est-ce que, I limited the search to the dates during which I could also find occurrences of the complementizer only constructions, to make sure that the occurrences of est-ce que retrieved were in competition with the construction at hand. The limited search only included the occurrences found between 1780 and 1950. Occurrences of wh est-ce que were retrieved only for comparison purposes, to verify if the context of usage of each construction was the same or different when wh+comp finally made its way into literary works.
With all the other wh- words, only the forms with the complementizer alone and with the full periphrastic est-ce que could be found. Ousque also has quite peculiar behaviour: out of 21 occurrences, a little over half of them (11) were found in embedded context. That is the greatest number of all the wh-forms except for quand, for which we only have 10 occurrences. We cannot say that, in general, où prefers the embedded context, since we do not find similar numbers with où que and où est-ce que (respectively 28 and 1 percent of occurrences in embedded context). It is quite interesting to find such an occurrence, and we might pose some hypotheses concerning its appearance in written texts. Où clearly seems to have a special status among the inventory of wh- elements. It is the one that was attested first in written texts with the complementizer only, in 1784. The next one, combien, is found more than half a century later, in 1837. Its evolution with the periphrastic form is also interesting: it was one of the last ones to be attested with the periphrastic form, only at the end of the fourteenth century (see Rouquier 2002), which we would not predict based on its appearance with the complementizer today. I do not have an explanation as to why its behaviour is different from the other wh elements, but these facts must be kept in mind for the analysis.

As was mentioned at the beginning of this section, the search in Frantext was conducted with two ideas in mind: to find the first attestation of each wh+comp in written texts (from our corpus), which was presented in Figure 6 above, and to see if the wh+comp behaved differently from the wh est-ce que when it came to the syntactic context (matrix vs. embedded). Figures 8 and 9 are the two graphs presenting the findings for the second question. They compare the two forms with each wh- element, the darkest bars representing the matrix context, and the lighter ones representing the embedded context (indirect question and free relative) (based on Table 1).

31 Note however, as mentioned before, that the variant wh + c’est is most likely represented, but it was not searched for. The ‘intermediate’ forms here that we tried to find are the orthographically fused ones.
The numbers in Table 1 presented a total usage of 85 percent in matrix context, with a 15 percent usage in embedded context. However, Figures 8 and 9 show that these numbers are not distributed equally; it is the wh+comp construction that is almost exclusively responsible for the occurrences found in embedded context. We notice right away the difference between the two graphs: the first shows almost none of the embedded context. The periphrastic form was used categorically in matrix context, i.e. in direct questions. The direct question context was also the
context in which the wh+comp appears most of the time, although this latter form seems to allow considerably more variation. Apart from quand, which only has 10 occurrences in total, none of the wh+comp prefers the embedded context over the matrix context, although they all allow it more than the wh est-ce que. Does this mean that the complementizer is a distinct lexical item that appears in broader contexts than the periphrastic form? It might very well have been the case historically. However, given the genre of the written texts included in the corpus, it is very possible that the wh est-ce que might be underrepresented compared with actual usage. One reason for this might be that since it seems quite clear that the wh+comp is associated with a lower (lowest?) register, perhaps the more nuanced wh est-ce que is less salient, and therefore less used in representations of lower registers of speech. So although the difference between the two forms according to the syntactic context is interesting and allows us to state a few hypotheses, it only applies to the literary genres that were included in the Frantext corpus. It would be interesting to see how the two constructions would be used in vernacular or colloquial usage, but this is quite difficult to verify with historical data. However, we can state with certainty that wh+comp appeared during the 18th century, maybe even during the 17th century, and that it is associated with vernacular speech, as the search in metalinguistic documents revealed.

The last piece of information that we get from this corpus search is presented in Table 2.

**Table 2. Raw numbers for each wh- word with comp only and with est-ce que**

<table>
<thead>
<tr>
<th></th>
<th>Wh+comp</th>
<th>Wh est-ce que</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combien</td>
<td>43</td>
<td>8</td>
</tr>
<tr>
<td>Comment</td>
<td>188</td>
<td>97</td>
</tr>
<tr>
<td>Où</td>
<td>178</td>
<td>82</td>
</tr>
<tr>
<td>Pourquoi</td>
<td>98</td>
<td>198</td>
</tr>
<tr>
<td>Qui</td>
<td>46</td>
<td>69</td>
</tr>
<tr>
<td>Quand</td>
<td>10</td>
<td>118</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>563</strong></td>
<td><strong>572</strong></td>
</tr>
</tbody>
</table>
Note that these are only raw numbers, not calculated in reference to anything, so statistically they are meaningless. However, they show that in the same texts for the same period, we found more occurrences of the wh est-ce que than the wh+comp for qui, quand and pourquoi. However, the opposite is true for comment, combien and où; these three appeared more often with only the complementizer than with the est-ce que form. It is also interesting to note that the total numbers are very similar for both constructions; when wh+comp started being used in texts, it was used with approximately the same frequency as the much older wh est-ce que construction. A breakdown of the numbers by decades would probably be necessary to get a clearer picture, although it is very possible that another variant (non-inversion variant or even in situ, for example) influences the usage. However, these numbers can predict what we saw earlier in the ALF: wh+comp was widespread enough in usage that it can even appear as often in fairly formal, published texts as the very old wh est-ce que variant.

2.2.4 Wh in situ variants

The last variant included in this work is the in situ one. The wh word overtly stays where it undergoes Merge. Note that for subject questions such as (45), the in situ version is homophonous to the fronted one, given the pre-verbal position of the subject.

42. Tu as vu qui?
   you have seen who
   ‘Who did you see?’

43. Elle vient quand?
   she comes when
   ‘When is she coming?’

44. Tu vas où?
   you go where
   ‘Where are you going?’

32 For où, this is also true when we add the ousque to the number of wh est-ce que. It still gives us 103 occurrences of wh est-ce que for 178 occurrences of wh+comp.
45. Qui vient?
     who comes
     ‘Who is coming?’

The in situ variant is the most recent one, according to Tuaillon (1975), who based his assumptions on map 25 from the Atlas Linguistique de la France (Gilliéron & Edmond 1902-1910), which showed that in situ is completely absent for the ALF (see previous section). This fact is corroborated by Elsig (2009), who compared two different spoken corpora, one with speakers that were born at the turn of the 20$^{th}$ century, and one with speakers who were born after the 1950s.$^{33}$ The oldest one had a percentage of 0.9 of usage of the wh-in situ variant (p. 149), whereas the corpus that included the speakers born 50 years later had an average of use of 7.6 percent (p. 147) of the in situ variant. In modern-day France, usage of in situ is slightly more spread than in LaF, representing around 17 percent of all wh interrogative occurrences (Coveney 2002: 189).

As far as my research could tell, no instance of non-echo wh in situ has been found in texts before the Modern French period (19$^{th}$ century). The studies done on the 17$^{th}$ century (Ayres-Bennett (2004) (who herself reviewed numerous studies that have been done on the same period), Elsig (2009)) have not found any instances either. Note however that there has not been any detailed study of texts from the 18$^{th}$ century, so there is a possibility that this variant appeared a little earlier (but not earlier than the 18$^{th}$ century, according to current evidence).

The non-echo meaning of French in situ is uncontroversial. It is not clear however if the in situ variant includes a special presupposition that the other forms would not require (see Boeckx 1999, Mathieu 2004, Chang 1997), because there seems to be much inter-dialectal variation. In this work, it will be assumed that there is no special presupposition associated to leaving the wh in situ, since we are talking about LaF; this view is also shared by Elsig (2009) and Coveney (2002), who suveryed the form using quantitative sociolinguistic methodology. The reader is

$^{33}$ Recall that Elsig used Shana Poplack’s corpora hosted at the University of Ottawa: the Ottawa-Hull French corpus (Poplack 1989), which features the speakers born after the 1950s, and the corpus Récits du français québécois d’autrefois (Poplack & St-Amand 2007), which features the speakers born at the turn of the 20$^{th}$ century.
referred to the works cited above for in-depths discussions of this matter. Since many proposals have been made regarding how to syntactically analyse this specific variant, in both French and other languages, we will come back to the issue in the proposal part of this thesis, since almost all analyses of wh in situ in French concern the optionality in movement, of which the in situ seems to be evidence. The next section reviews the possible intervention effects at play with this variant, although, much like for the interpretation of it, the exact restrictions are the object of much debate.

### 2.2.4.1 Intervention effects and in situ

Contrary to fronted wh variants, wh in situ in French is somewhat restricted. As the nature of the restrictions will tell, they may not affect the rate of usage of this variant (quantitatively, in the variationist sense), since they involve lower frequency constructions such as negated interrogatives and specific quantifiers. Moreover, those constructions might very well be associated with a register that is different from the one the in situ is usually used in. We will come back to this issue in Chapter 4. The data presented in this section was taken mainly from Butler and Mathieu (2004), but since the dialect described in that source might not correspond to the one described here, and since wh in situ famously involves dialectal variation (Chang 1997, Mathieu 2004), each example was ‘translated’ to fit the LaF particularities (e.g. absence of *ne* in negation), and tested on three LaF native speakers.

French wh in situ is not compatible with negation, but the fronted variant is.

46. *Il voit pas qui?*
   
   he sees not who

47. *Qui qu’il voit pas?*
   
   who that-he sees not
   ‘Who doesn’t he see?’

48. *Tu fais pas quoi ce soir?*
   
   you do not what tonight
49. Qu’est-ce que tu fais pas ce soir?
   What-is-it that you do no tonight
   ‘What are you not doing tonight?’ (Butler & Mathieu 2004: 35)

The licensing of wh in situ seems to be blocked by head elements; some XPs can also have a blocking effect, as in the following examples.

50. *Aucun étudiant a lu quoi? (negative quantificational subjects)
   no student has read what

51. Qu’est-ce qu’aucun étudiant a lu?
   what-is-it that no student has read
   ‘What did no student read?’

52. *Seulement Jean fait quoi? (focus markers, such as seulement, même)
   only Jean does what

53. Qu’est-ce que seulement Jean fait?
   what-is-it that only Jean does
   ‘What does only JEAN do?’

54. */?Il a beaucoup lu quoi? (iterative adverbs such as beaucoup, peu, trop)
   he has a-lot read what

55. Qu’est-ce qu’il a beaucoup lu?
   what-is-it that-he has a-lot read
   ‘What has he often read?’ (Butler & Mathieu 2004: 35-36)

Note however that for this last type of XP, the iterative adverbs, the judgements from my LaF speakers are not as straightforward. They still prefer the fronted version, but they do not categorically reject the in situ one (2 out of my 3 speakers said they thought it did not sound ‘that bad’). All these restrictions are important to note because ‘real’ in situ languages such as Chinese do not exhibit such restrictions on the licensing of in situ (see discussion in Butler & Mathieu 2004: 36-37).
Embedded context also seem to block in situ, although there might be dialectal variation in this particular case. The details of the distribution of in situ in embedded context are described in the next section.

Mathieu & Butler (2004) argue that these restrictions on wh in situ prove that the French system is not the same as the Chinese one. They argue that the restrictions are explained by “(i) a split configuration, and (ii) the semantic and discourse properties of the WH-phrase-in-situ.” (p. 37). Leaving aside the semantic and discourse properties of the in situ (see below, the summary of Adli’s (2006) work), we can explore the ‘split construction’ hypothesis. For Butler & Mathieu, (57) and (58) have similar properties and structures, whereas (56), the non-split variant, is different (examples have been ‘translated’ into LaF).

56. **Combien de livres** que t’as lu?
   how many of books that you-have read
   How many books did you read?

57. **Combien** que t’as lu **de livres**?
   how many that you-have read of books
   How many books did you read?

58. T’as lu quoi?
   you-have read what
   What have you read?

They define split constructions as “a construction with a bare operator [...], which is an operator which is structurally detached from its noun restrictor.” (p. 1). Without getting into the details of their system, which makes use of various tools taken from a formal logical framework, they assume that the operator takes scope over the noun, i.e. the wh word, it is attached to. We can imagine that such a relationship is translated into syntax via an operator located in SpecCP, c-commanding the rest of the clause. We will come back to this possibility in Chapter 4.

The completely opposite view is taken by Adli (2006); he found that in fact, there are no contextual or interpretive differences between in situ and its fronted counterparts, and there are no intervention effects on in situ either. He used a very specific methodology, asking his
participants for three types of judgements/tasks: a qualitative interview (grammaticality judgements within contexts, with reinforcement questions from the interviewer to verify answers, with 20 university students in Paris); graded grammaticality judgement tests (with 65 university students in Toulouse, France); and a reading time measurement experiment, to test the relative time of cognitive processing for each of the two variants, fronted and in situ (with 117 university students in Toulouse, France). 34

The three types of methodologies yielded consistent results: he did not find evidence for the restrictions on in situ mentioned in Cheng & Rooryck (2000), in Bošković (1998), and for some of the ones mentioned in Mathieu & Butler (2004) (Cheng & Rooryck’s and Bošković’s proposals are summarized in section 4.1). He tested for in situ in embedded contexts (see next section), for in situ with negation, quantifiers (such as *plusieurs* ‘many’) and modals, all of which were said to be incompatible with in situ according to the sources above. However, Adli found that his speakers accepted in situ in all those environments, and, importantly, that “we could not find the assumed differences of interpretation between wh-in-situ and wh-movement in French with respect to context presupposition” (p. 187). This weakens considerably the claims made by, among others, Bošković (1998) and Cheng & Rooryck (2000), in favour of LF movement for wh in situ variants. Adli favours an analysis that rules out LF movement, mentioning that: “an analysis of French wh-in-situ without LF-movement is at present backed by a broader ground of evidence that an analysis with LF-movement” (p. 190). He mentioned Reinhart’s (1998) analysis of wh in situ in terms of choice functions, but does not develop a further analysis.

Although speakers of LaF did seem to be somewhat sensitive to intervention effects with in situ, as demonstrated in examples (46) to (55) above, the judgements are not robust. To the best of my knowledge, Adli (2006) is the only contribution to the field of French in situ study that has systematically used three different types of data to get the most accurate judgements possible, as well as asked judgements from over a hundred participants for some of the experiments he conducted (compared to three speakers in the present study that were asked for only plain

34 I refer the reader to his paper (Adli 2006) for details on the methodology utilised, the exact contexts of his testing, and the reasons justifying such experiments.
grammaticality judgements). Moreover, he mentions problems with the current literature on in situ that were corroborated in the present thesis. He quotes one of his participants: “If there are sentences that one is not used to saying, one will rather say them in a better French.”\(^{35}\) (p. 173). This highlights the gap between the spoken LaF grammar described here and the grammar that speakers need to access in order to give grammaticality judgements on sentences such as (46)-(55): in other words, the judgements that were asked of LaF speakers are not part of the LaF grammar (we will come back to this issue in Chapter 4, and bring up Zribi-Hertz’s (2011) proposal for a diglossic situation within the French speaking world). Given those facts, and while no other studies involving LaF participants have been conducted, we will assume Adli’s conclusions when it comes to wh in situ in French, which are quoted below (emphasis my own):

59. “Firstly, the qualitative interviews uncovered considerably weaker evidence in favour of the assumption of LF-movement than has been claimed in some previous contributions.”;

60. “Secondly, a graded grammaticality judgement test revealed even in terms of fine nuances an identical level of grammaticality [between in situ and fronted]”;

61. “Thirdly, a reading-time experiment showed that both variants have the same cognitive complexity in processing”.

(Adli 2006: 199)

Note that additional asymmetries exist when we consider multiple wh in French, but we will leave these aside for now, since multiple wh interrogatives are not part of our variable context (as described in section 2.1.3) (see Bošković’s work (2007, among others) for more details on multiple wh in French and in other languages).

2.2.5 Embedded contexts

Most variants described in the previous section may also appear in embedded context in LaF, but not all of them. This section surveys the facts linked to wh elements appearing in embedded

\(^{35}\) Original quotation: “S’il y a des phrases qu’on n’a pas l’habitude de dire, on les dit plutôt dans un meilleur français.” » (Adli 2006: 173).
context. It is important to give a complete picture of the embedded context because many phenomena found in diverse languages (V2 in German for example) present an asymmetry when it comes to the clausal context, main or embedded. Very few studies have examined the usage of main wh interrogatives in day to day speech, but we have seen that even fewer studies have included wh embedded clauses. This section will therefore focus on what has been written on the theoretical and typological status of embedded wh, with a complete description of forms that are considered grammatical in LaF.

Some asymmetries are present between main and embedded contexts in French when it comes to wh clauses. First of all, inversion is impossible in embedded contexts.\(^\text{36}\)

62. Je sais pas comment il a fait.
   I know not how he has done
   ‘I don’t know how he did this.’

63. *Je ne sais pas comment a-t-il fait.

64. Ils lui ont dit quand il devait arriver.
    they him have told when he should arrive
    ‘They told him when he should arrive.’

65. *Ils lui ont dit quand devait-il arriver.

66. Je me demande qui vous voyez.
    I myself ask who you(pl.) see
    ‘I am wondering who you are seeing.’

\(^{36}\) If inversion happens it must reflect reported speech, and the ‘embedded’ wh becomes a direct question, as in:

   Je lui ai demandé: “quand arriveras-tu?”
   ‘I asked him: when will you arrive?’

Compared to:

   Je lui ai demandé quand il arriverait.
   ‘I asked him when he would arrive.’
67. *Je me demande qui voyez-vous.

(62) to (67) are examples of indirect questions, introduced by a verb that entails some kind of questioning, such as wondering, knowing, telling, etc., which select an embedded clause as their argument. Note that any form of interrogative reinforcers can appear in embedded context as well.

68. Tu regardes comment c’est qu’on en a, ...

   You look how it-is that we it have
   “You check how many (of these) we have, ...”

   [Les Bougon, épisode Citoyen du Monde, Radio-Can.]

69. Ils lui ont dit quand-ce qu’il devait arriver.

70. Je me demande qui que vous voyez.

The same facts are also true for what have been called ‘free relatives’, because the embedded clause is not selected as an argument by the verb in the main clause. This is distinct from questioning verbs, but free relatives seem to behave the same way when it comes to wh elements ((68) above is also an example of a free relative).

71. Tu vas où que tu veux.

   you go where that you want
   ‘You go wherever you want.’

Another difference between main and embedded contexts in French is that the form of the wh word what can vary in embedded context (72).

72. Jean a demandé ce que Marie a fait.37

   John has asked ‘it what’ Mary has done

But more importantly, the wh in situ variant seems impossible in embedded contexts (73).

37 Note that in spoken French, ce que is often replace by qu’est-ce que.
73. *Jean a demandé Marie a fait quoi/ce que.
    John has asked Mary has done what/‘it what’
    ‘John asked what Mary did.’

However, judgements are not as strong with all wh words in situ in embedded contexts. Although still mainly ungrammatical, some wh words seem to be a little more accepted in situ than others.

74. */? Je lui ai demandé il va là comment.
75. */? Tu m’as dit que t’allais où.
76. */? Je me demande il va arriver quand.

Shlonsky (2009) proposes that in situ is allowed in embedded contexts, whereas Rowlett (2007:196) states the opposite. I would not view things so categorically, but it is clear that judgements do not agree as to whether these examples are really instances of embedded clauses, or if they are instances of reported speech with direct questions. Most speakers – including myself – interpret these as direct questions, and would put a question mark at the end of examples (74)-(76).38 We would therefore lean toward’s Rowlett’s opinion. For the purposes of this work I will assume that wh in situ in embedded context is not accepted, but we will return to this in Chapter 4.39

38 Interestingly, we get similar judgements with (63), (65) and (67), which are examples of subject-verb inversion in subordinate contexts. (74) to (76) seem to be slightly better, but there is a tendency to interpret all of those as reported speech with two independent clauses.

39 In fact, I propose that my analysis presented in Chapter 4 can also be applied to embedded contexts, at least in the system of speakers who accept variants of wh est-ce que in embedded clauses. My claim does not extend to other variants such as (72) however, for which I do not have any additional analysis than what is usually proposed (see Rowlett 2007 for a good summary).
There are arguments in favour of proposing a different landing site for the wh element in matrix and embedded contexts (that is, if we assume wh movement). Rowlett (2007) shows the asymmetry using left dislocated elements:40

77. Toi, quand tu arrives?
   you, when you arrive?

78. *Quand toi, tu arrives?
   when you you arrive

79. *Dis-moi toi, quand tu arrives
   tell-me you, when you arrive

80. Dis-moi quand toi, tu arrvives
   Tell-me when you, you arrive

Rowlett’s claim is that if (77) is grammatical in matrix contexts, then (79) should be in embedded contexts as well, unless they have different landing sites for wh elements (one higher than TopicP in embedded context, and one lower than TopicP in matrix context, assuming that the position of TopicP stays constant). Two factors might prevent this analysis from applying in our case: first, in a dialect/register point of view, it is not clear if this difference really exists in LaF; second, theoretically, this analysis assumes wh movement, which might not be happening in wh est-ce que constructions in LaF.

---

40 It is not clear if these examples are part of LaF’s grammar. In LaF, there is (almost) obligatory clitic doubling (Courmane 2008), so LaF speakers would probably prefer wh in situ in such cases:

i. Toi t’arrives quand?
   You cl. arrive when
   ‘When are you arriving?’

De Cat (2002: 96, cited in Rowlett 2007: 194) suggests that right dislocation as in (ii) is preferred in interrogatives over left dislocation, which is slightly preferred in declaratives.

ii. Tu arrives quand, toi?
2.2.6 Summary: what does this all mean?

This section has taught us that a multitude of variants is not a recent phenomenon. As early as Old French we could find variation in the expression of interrogatives. We have seen that today, variants of wh *est-ce que*, including wh+comp, are the most commonly used variants in LaF, and that wh in situ seems to be spreading in usage. The wh fronted with inversion variant has now virtually disappeared from LaF, although it remains in written and highly formal registers (in which wh+comp and wh in situ are impossible, and wh *est-ce que* quite rare). What does this all mean for us?

First of all, I have answered the question ‘why so many forms?’: it comes from a historical evolution, and from the fact that certain variants are ‘specialised’ to certain registers, therefore taking a specific function that can be observed pragmatically. All of them have remained in competition because of this difference in usage context.

Second, we have seen that *est-ce que* was technically an innovation, since it appeared later than wh fronting with inversion, although not very much later. However, it spread somewhat quickly to all possible domains in which a wh element can be found, and even beyond (yes-no interrogation as well). It never became the majority variant in written registers, but it was never a highly stigmatised variant, even if it was at some point (and still is in some dialects) associated to lower registers (vernacular speech in general). It therefore seems that *est-ce que* played a major role in the evolution of the wh interrogative system (and, to a lesser extent, in the yes-no interrogative system as well, which we will briefly come back to at the end of Chapter 4).

These facts lead to an analysis that centers around the wh *est-ce que* form. It will be hypothesised, following the evolution of the system, that the wh *est-ce que* construction never fully lost its syntactic clefting properties. It appeared in the language as a way to emphasize the wh element by way of clefting, as we have seen was the case during the Old French period. The construction continued to be used and processes of grammaticalisation began structurally. The result is that two ‘systems’ are competing since the Old French period: the wh *est-ce que* ‘cleft’ system, and the wh movement ‘inversion’ system. It seems that in vernacular LaF, the former

---

41 A complete list of LaF subject and object interrogative variants can be found in Appendix I.
took over most (if not all) of the usage. Chapter 3 describes the exact processes of grammaticalisation that seem to have taken place with the est-ce que element, and also specifies what is meant by ‘clefting’ when we talk about French. The complete syntactic analysis of the wh est-ce que system will be presented in Chapter 4.
Chapter 3

3 What is ‘est-ce que’?

This chapter surveys the status of the copular element est-ce (que), ‘is-it (that)’. Since the wh est-ce que construction in LaF seems to have a lot in common with clefts, and since we know that it first appeared in the language as such, we need to know exactly what is meant by ‘cleft’. To do so, we first need to determine the exact status of the copular element, the est in est-ce que/c’est que, since this is what will determine the semantic and syntactic profiles of our constructions. This section summarizes the findings concerning the evolution of the copula within the wh est-ce que construction, and it also covers other functions that the copula be in French can fulfill, such as the cleft and pseudocleft constructions.\(^\text{42}\) I will propose a path of grammaticalisation for the wh est-ce que construction based on Roberts & Roussou’s (2003) (among others) view of grammaticalisation cycles.

Note that the question here is not if wh est-ce que is derived from the declarative cleft (as was the main question asked by Obenauer 1977, 1981), the question is rather if the copula, with or without a clitic element, fulfills the same function in interrogatives as it does in other copular constructions, such as declarative clefts. It will be shown that both constructions, declarative clefts and wh est-ce que, are equally devoid of special semantics, and that their uses are comparable, in that they are not marked in LaF. It is therefore argued that it makes sense to treat them in similar ways, and that just as much as the declarative clefting is no longer a tool to put major emphasis on a discourse element, the est-ce que in interrogatives is merely a discourse variant and a register marker.

3.1 The status of ‘est-ce/c’est’

Contemporary grammarians agree that the clitic ce ‘it’ no longer has any semantic content, and that it is completely dependent on the copula est ‘is’. Moreover, it has been shown that the que has to be treated separately from the copular element (Obenauer 1981), mainly because of its

\(^{42}\) Other types of copular construction that have been said to be copular even without the use of to be, are not included here. Such constructions include voici/voilà constructions (Morin 1985).
que/qui alternation, which happens in the complementizer position. Although it has been proposed (Rowlett 2007) that the full est-ce que is a complex complementizer, we have seen evidence that points towards considering the est-ce/c’est and que/qui as two different items. One of the main arguments would be that the variant wh+que/qui (complementizer only) is extremely frequent, and that the que/qui never lost their syntactic distinction (qui for animate subject and que elsewhere), while est-ce/c’est seem to have lost their internal meaning. The evidence presented in this section will make this claim more robust.

Brunot & Bruneau (1969) mention: “As early as the 15th century, [...] we can consider est-ce que as a simple ‘interrogative tool’, in which ce, like the other elements, has lost its [semantic] value”\(^\text{43}\) [my translation] (p.487). Many authors are vague when it comes to describing est-ce (que). Obenauer (1981: 104) talks about est-ce que causing a mise en relief ‘highlighting’ of the wh element, which is similar to the interpretation of a cleft (although still assuming that wh est-ce que questions are NOT clefts), and Brunot & Bruneau (1969: 487) talk about a forme d’insistance ‘emphasis form’ when the wh element is used with est-ce que. Although assuming the ‘special semantics’ associated with it, most people still propose that a ‘bare’ wh is freely interchangeable with a wh est-ce que: for instance, Elsig (2009) considers that the bare wh is only a lexical variant of wh est-ce que (the est-ce que is affixed to the wh – more details of his analysis are presented later in this chapter). This section aims at clarifying the issue, and determining once and for all the status, syntactic and semantic, of the est-ce que in LaF wh est-ce que constructions.

It has been shown by numerous authors (Druetta 2002, 2003; Rouquier 2002, 2003; Elsig 2009) that the cleft element, copula + clitic (est-ce/c’est) had grammaticalized into one single unit as early as the Middle French period (15th century), exactly as the quote above from Bruneau & Brunot (1969) states. Although, as noted above, most grammarians have come to that conclusion, very few actually provide enough data to support their claim. Rouquier (2003) did an extensive study on the evolution of the construction, and studied in detail each element of the wh-est-ce que interrogatives, the position and function of ce, est and of the wh element as well. Her

\(^{43}\) “dès le XVe siècle, [...] On peut considérer dès lors est-ce que comme un simple outil interrogatif, où ce, comme les autres éléments, a perdu sa valeur propre.”
conclusions are very insightful, and provide much needed evidence for what had previously been merely assumed.

First, she shows that ce had been the subject of the est, which was a full verb, not a copula, only until Middle French (14th c.). She cites Foulet (1921), who mentions that ce was likely the bearer of sentential stress in Old French:

81. qui est ce, diex, qui m’aparole?

who is it, god, who is me talking

“My lord, who is the person who is talking to me?”

(Renart IV 233, Foulet 1921 : 253, cited in Rouquier 2003 : 340)

As evidence for the grammaticalisation of the form, Rouquier mentions that the construction (est-ce que) went from being restricted to mainly ‘saying’ verbs (dire ‘tell’, demander ‘ask’, etc.) in Old French, to all kinds of verbs in Middle French (p. 351). Similarly, the ce went from being the bearer of sentential stress44 to clitic, and from being able to move around (before the wh word (82), before or after the copula (83 and 84)), to a fixed position (84), even sometimes fused orthographically with the copula (85).45

82. ce que est ore que vos dites?  (Renart 2080 (dated 1150-1200))

83. que ce est que vos dites?  (Queste 276 (dated 1200-1250))

84. qu’est ce que dit avez?  (Jugement Roy de Navarre 167 (dated 1350-1400))

85. Qu’esse que tu veulx?  (Rondeaux 457 (dated 1400-1450))

(Rouquier 2003 : 348, 351)

44 Indeed, the pronoun ce could appear before the 14th century in its strong form: iço.

i. Dex, fet il, qu’est iice que j’oi?

(Charrette 6551, cited in Rouquier 2003: 353)

45 Note that the ce before the copula as in (83) remained and became grammaticalised as well. It is still very productive to this day.
She proposes that the reanalysis, which is inherent to the grammaticalisation process, happened to the verb *to be* and the pronoun *ce* together. I will assume her view and use her data to demonstrate in the next section that the process of grammaticalisation that happened between Old, Middle and Modern French for *est-ce que* corresponds quite closely to the definition of grammaticalisation (Hopper & Traugott 1993) that is most generally assumed in the literature. I will also consider whether this specific case gives evidence for a theory of language change that is based on reanalysis between different generations of speakers.

### 3.1.1 Grammaticalisation of *est-ce/c’est*: from XP to $X^0$

This section\(^{46}\) reviews the evolution of the element *est-ce/c’est* in terms of grammatical language change, i.e., grammaticalisation.\(^{47}\) We will explore the theories put forward by Hopper & Traugott (1993) and by Roberts & Roussou (2003), to demonstrate that the changes that occurred within the interrogative CP from Old to Modern French are linked to the grammaticalisation of the *est-ce/c’est* element, which went from a lexical item to a functional affix and from an XP to an $X^0$. This change gives additional evidence to the claim that “grammaticalisation is always upward and leftward in the syntactic structure” (Roberts 2010).

I assume here the definition of grammaticalisation that allows a lexical element to become more functional “grammaticalisation is a process whereby lexical items lose phonological weight and semantic specificity and gain grammatical function” (van Gelderen 2008a: 245). Diachronically, we speak in terms of ‘cycle’ (as far back as Bopp (1868), and more recently by Hodge (1970), van Gelderen 2008a, b), and this can apply to a variety of linguistic elements, causing the language to become more synthetic. The typical linguistic cycle is presented in the example below:

\(^{46}\) I thank Ailis Cournane for her help and comments on this section, and for allowing me access to Cournane (2008).

\(^{47}\) I do not mention it further, but note that the changes described in this section could be explained through Kroch’s (1994) competing grammars. I chose to explain the facts in terms of grammaticalisation because it seemed to fit the data better, and that although the competition, as meant by Kroch, is definitely present in French, the preservation of the archaic grammar is unnatural under Kroch’s theory, which is one of the reasons why we can propose a diglossic situation in contemporary French (see Section 4.5.1).
According to this, some words have a universal tendency to go from lexical to functional, step by step as languages change. If we apply this cycle to the construction at hand, the element *ce* in *est-ce/c’est*, we hypothesize that the cycle would look something like what is presented in Table 3 below.  

<table>
<thead>
<tr>
<th>Status</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>N/A</td>
<td>Old French</td>
<td>Middle French</td>
<td>Modern French</td>
<td>N/A</td>
</tr>
<tr>
<td>Evidence</td>
<td>N/A</td>
<td>subject of <em>to be</em>; bearer of stress</td>
<td>fixed position; can no longer be stressed</td>
<td>cannot be separated from the then invariable <em>est</em></td>
<td>N/A</td>
</tr>
</tbody>
</table>

According to this view, the grammaticalisation process applies to the pronoun *ce* and the copula. The complementizer *que* and the wh element are not involved in this process, although they are the necessary context for it to happen. Rouquier (2003) proposes that the changes happened through lexical rebracketing, which would most likely affect the intonation of the construction (we will briefly come back to the intonational properties of this construction for contemporary data in Chapter 4).

In minimalist terms, van Gelderen (2008a) uses economy principles and grammaticalisation theory (such as Hopper & Traugott 1993) to justify the trend in (87) affecting the syntactic structure during grammaticalisation processes, which is part of her greater Feature Economy proposal (88).

\[
87. \ XP \rightarrow X^0 \rightarrow \text{Affix}
\]

---

48 The historical facts presented in Table 3 are taken from Rouquier (2003), based on what was described in the previous section, and simplified for the sake of clarity.
The assumption under this framework is that language change occurs through ‘micro-reanalyses’, and can be illustrated through feature changes (van Gelderen 2010). The claim made in this thesis is that there are in fact two interconnected processes of grammaticalisation in play in the evolution of the wh est-ce que construction, one involving the copula est, and another one involving the pronoun ce. Both types of grammaticalisation have been attested many times in the literature (Lohndal 2009, van Gelderen 2006, Roberts & Roussou 2003, etc.). Roberts (2010: 5) summarises in a very concise way the main ideas behind Roberts & Roussou’s (2003) approach to grammaticalisation, which, we will see, correspond exactly to how we can analyse the changes affecting both the copula and the pronoun in the wh est-ce que construction.

a. “the diachronic movement of a given morpheme, possibly tracked over many centuries through successive reanalyses, is always ‘upwards’ in the structural hierarchy of functional categories” (Roberts & Roussou 2003: 36);

b. “Successive upward reanalysis along the functional hierarchy is thus how we define grammaticalisation paths” (Roberts & Roussou 2003: 202);

c. “the path is traversed by the loss of steps of head movement, leading to changes from Move to Merge” (Roberts & Roussou 2003: 71).

The term ‘successive upward reanalyses’ illustrates the ‘micro’ changes that are in play during grammaticalisation; if we think in terms of language acquisition, each generation would make a very small reanalysis that would affect, say, one feature or the other, and it is these small

Note that all of these claims are of course empirically supported. Our data can be considered to provide further evidence for those claims, and we use Roberts & Roussou (2003) as a framework and an explanatory and predictive tool to understand the changes that took place within the interrogative CP since Old French. Note that Roberts & Roussou use exploded functional categories, within the cartographic approach (Rizzi 1997, Cinque 1999).
reanalyses that, over time, give rise to language change. This approach is a generalisation on what has been observed in different languages, and characterises what types of language change are possible, and predictively, in which direction they are most likely to occur.

The second grammaticalisation process, after the cliticisation of the pronoun *ce*, at play within the wh *est-ce que* construction concerns the copula. Lohndal (2009: 11) proposes, based on Van Gelderen’s Feature Economy Hypothesis (see (88) above) a pathway of grammaticalisation for the copula that seems to be universally observed.

89. Copula Cycle

<table>
<thead>
<tr>
<th>demonstrative/pronoun</th>
<th>copula</th>
<th>grammatical marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>specifier</td>
<td>head</td>
<td>affix</td>
</tr>
<tr>
<td>[iF]</td>
<td>[uF]</td>
<td>-</td>
</tr>
</tbody>
</table>

We are more concerned about steps 2 and 3, given the facts of French (in Old French, *est* was already a copula, i.e. a head). Lohndal (2009: 35) uses examples from languages in which the copula is homophonous with the general focus marker to demonstrate the passage from copula to grammatical marker.

90. Onişọwo *ni mi* [Yoruba]

merchant COP 1SG

‘I am a merchant’ (Ashiwaju 1968: 28)

91. Lále *ni wón dé*

in.the.evening FM they arrived

‘It was in the evening that they arrived’ (Bamgbọse 1966: 37)

Moreover, he shows (pp. 36-37) that with Turkish, there is diachronic evidence that the copular affix *y* was once a free morpheme *i*, to which tense and agreement markers could attach.

92. (ben) satci-ı-im [Turkish]

Affix

1SG seller-COP-1SG

‘I am a seller’

(Kornfilt 1997: 77)
These are examples showing how the Feature Economy Principle works, using real data that is connected to the construction under study; note however that I do not claim that est-ce went through the whole copula cycle as Lohndal proposes. Instead, we argue that we are looking at a diachronic window that illustrates part B of van Gelderen’s Feature Economy Principle (see 88 above). We wish to use this view to argue that est, in the case where it gets fused with the clitic c(e), is no longer a copula, but now an interrogative marker, most likely the lexical realization of the [Q] feature. It went from a head (I) to a higher head (C), and is now associated to A-bar features instead of A features. This change, happening between Old and Middle French, has certainly been triggered by the complete loss of V2, which was happening during this period as well. Since the loss of V2 involves a loss of movement from I to C, there is only one step towards a complete reanalysis, where the copula+clitic lexeme is seen as being Merged directly into the higher (interrogative) head C.\textsuperscript{50} Although both [Q] and [wh], associated with C, are [iF] (Roberts & Roussou 2003: 28), it has been proposed that [wh] does in fact behave more like case features than phi-features, exhibiting more [uF]-like properties than [iF] properties: “wh is in fact a feature which cross-linguistically requires checking on (all) wh-phrases, and so is actually similar in this to the –interpretable Case features assumed to be carried by argument DPs across (all) languages.” (Simpson 2000: 101-102). Together with the grammaticalisation of the pronoun ce, it explains the properties and distribution of the wh est-ce que construction from Old to Modern French, which are summarized in Table 4 below (some of the information in this table was already given in Table 3).

\textsuperscript{50} Note that this proposal is only for the wh est-ce que interrogative construction; I do not claim that it is a general process that happened during this period, i.e. that all verbs got reanalysed as being Merged under C. The hypothesis is that the clefting structure of this particular construction favoured the reanalysis, only for the copula in interrogatives.
Table 4. Grammaticalisation of *est* and *ce* in terms of Feature Economy

<table>
<thead>
<tr>
<th></th>
<th>Old French&lt;sup&gt;51&lt;/sup&gt;</th>
<th>Middle French</th>
<th>Modern French</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pronoun</td>
<td>Clitic</td>
<td>Grammatical marker (affix)</td>
</tr>
<tr>
<td><em>ce</em> Projection</td>
<td>NP (ou DP)</td>
<td>Parasite to copular I</td>
<td>Affix to C (see below)</td>
</tr>
<tr>
<td></td>
<td>specifier</td>
<td>head</td>
<td></td>
</tr>
<tr>
<td></td>
<td>iF – [phi]</td>
<td>iF - ?</td>
<td></td>
</tr>
<tr>
<td>Diagnostic</td>
<td>- subject of ‘to be’</td>
<td>- no longer stressed</td>
<td>- attached to <em>est</em> (see below)</td>
</tr>
<tr>
<td></td>
<td>- possible bearer of stress</td>
<td>- fixed position (before or after <em>est</em>)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Copular verb</th>
<th>Grammatical marker</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>est</em> Projection</td>
<td>I</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>head</td>
<td>higher head</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>iF/uF – [Q]</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>- Can vary in tense</td>
<td>- invariable (always <em>est</em>-ce or <em>c’est</em>)</td>
</tr>
<tr>
<td></td>
<td>- needs a subject</td>
<td>- fixed position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- only appears in wh interrogatives</td>
</tr>
</tbody>
</table>

Based on the facts highlighted in Rouquier (2003), the hypothesis is that by the end of the Middle French period, the Modern French system seemed to be in place. The claim is that the clitic fixed itself on the copula before this complex lexeme (copula+clitic) got reanalysed as a grammatical marker. Because both forms *c’est* and *est-ce* are widely attested since Old French, we hypothesise that the fixation of the clitic on the *est* was the first process to happen, but that it only took a few centuries for this to become an interrogative marker, bearer of the [Q] feature, i.e. the [Q] feature is expressed by the formally complex *est-ce* [es] or *c’est* [se], never *est* or

<sup>51</sup> Note that as in Table 3, the period during which each step occurred is approximated.
*ce, since the whole thing lexicalised before it became solely associated with the interrogative head C.  

3.1.2 Contemporary French: actual status of est-ce/c’est

For modern data, Druetta (2002, 2003) has studied the construction in detail using a modern written corpus from France, and he shows that the element est-ce ([es]) is not an emphasis tool, unless used with ‘special’ intonation. He uses a distributional approach, analysing each occurrence of wh est-ce found (he only includes occurrences without a complementizer, trying to determine the contexts that allowed or even favoured complementizer deletion). He makes several observations that are relevant here. First, he mentions that deletion of the complementizer is impossible in what he calls ‘averbal’ (94) and infinitival (95) sentences, i.e. in which there is no other tensed verb than the verb to be:

94. Qu’est-ce *(que) cela?
   what-is-it *(that) this
   ‘What is this?’

95. Qu’est-ce *(que) faire?  
   what-is-it *(that) to do
   ‘What is there to do?’

It has been pointed out by a member of my committee that there might be a possibility that the c’est, contrary to the est-ce, could have ‘escaped’ grammaticalisation, since grammaticalisation is said to be linked to the preservation of the SVO word order (c’est already being SV). However, in our case, it is not exactly true. The spread in usage of the wh est-ce que construction is indeed linked to the development of SVO since it avoided SV inversion in the main clause, but it was not linked to the word order within the construction itself. The dual est-ce/c’est merely exists due to the free(er) word order of Old French, and the grammaticalisation of the form happened via multiple reanalyses, as explained above.

Note that the grammaticality of these sentences with the complementizer, although Druetta assumes it for the dialect he studies, is not clear in LaF.
This is therefore a sign that in those cases, the *est-ce que* is not syntactically equivalent to the *est-ce que* found in wh questions. In the latter case, the complementizer can be deleted, since neither *est-ce* nor *que* have any semantic content or precise syntactic function.\(^{54}\)

The second part of his study included a questionnaire aimed mainly at checking the facts just described, since interrogative data are hard to find in a corpus. What came out of this specific methodology is that out of the hundred speakers surveyed, most corroborated the corpus study, and the morpheme [es] seemed to indeed appear where it was expected. Moreover, his conclusions are quite interesting when it comes to the grammaticalisation of the *est-ce que*: he demonstrates that there are, in fact, two *est-ce que* forms; one that has completely grammaticalised (fused, empty of content) and another that is still a cleft element (Druetta 2003: 25). His main argument concerns the different prosodic contours of the two forms; there is a pause between the *est-ce* and the *que* in the first construction, which is typical of the focus constructions of the type *c’est X que* (Rialland et al. 2002). It is also this form which would allow a lexical element in between *est-ce* and *que* (96), as well as variation in the tense of the verb *to be* (97).

96. Qui est-ce donc que vous avez invité?
who is-it disc.marker that you(pl.) have invited
‘Who is it again that you had invited?’

97. Où était-ce précisément qu’il aimait déjeuner? (Druetta 2003: 26)
where was-it precisely that-he liked to dine
‘Where was it exactly where he liked to dine?’

He admits however that variation like (96) and (97) is extremely rare (as mentioned in Blanche-Benveniste et al. 1990), and it is believed that LaF restricts these possibilities even more. The

---

\(^{54}\) He mentions that deletion of the complementizer is also not allowed in total interrogatives: Est-ce *(que) tu viens? ‘are you coming?’*. It might therefore very well be the case that the grammaticalisation of the *est-ce que* construction is different in partial interrogatives than it is in total interrogatives. The claims made in this section are consequently restricted to the context of partial interrogatives. The case of total interrogatives in LaF will be briefly addressed in the last section of this work.
most convincing argument is therefore the prosodic one. We will come back to the possibility of the two different est-ce que’s later in this thesis; they seem to be attested in LaF as well, although the grammaticalised/ungrammaticalised forms seem to have different distributions – Chapter 4, section 4.2.4, addresses this issue and present additional distributional evidence for LaF.

Elsig (2009) also proposes an analysis for the est-ce que construction, based on the data gathered through a LaF corpus (Ottawa-Hull). He proposes that all variants of est-ce que in wh questions are now affixed to the wh element, and are therefore no longer independent units (Elsig 2009: 192-194). His proposal assumes a step further in the grammaticalisation process: the complete affixation to the wh element. The fact that est-ce que can be phonologically reduced represents for him that: “the complementizer [esk] can no longer be analyzed as occupying an independent structural (head) position (such as C⁰). Instead, the particle seems to have built a lexical and structural unit with the wh-word through affixation.” (p. 192). He argues, contra Plunkett (2000), that the Doubly-filled-Comp-filter (DCFC)⁵⁵ is in fact active in LaF, and that sentences such as (98) are allowed because the C head is empty, the ce que being affixed to the wh word in Spec-CP. Plunkett (2000) proposes (99), assuming that there is no such thing as the DCFC in vernacular French.

98. c’est drôle [CP comment ce que [C-Ø[TP tout [T-tourne] dans la vie]]] (Elsig 2009: 193)
99. c’est drôle [CP comment [C- ce que[TP tout [T-tourne] dans la vie]]] (Plunkett 2000: 521)

It is funny how COMP everything goes in life

Elsig shows that although it was a possibility for the wh element to be separated from the complementizer by a lexical element in earlier stages of French (i.e. when the est-ce que complementizer was not yet fully grammaticalized), the data from his two corpora show that it is no longer an option, therefore corroborating the affixation proposal. The summary of his findings is illustrated in Table 5.

---

⁵⁵ The Doubly-filled-Comp-filter (DCFC) bans the realization of both the spec-CP and the C head when the CP is selected by a matrix verb (i.e. in embedded context CPs).
Table 5. Lexical element insertions between the wh word and the complementizer (by period)

<table>
<thead>
<tr>
<th>Corpus/Period</th>
<th>Percentage of Wh X Comp (out of all wh est-ce que occ.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15th-17th c. plays</td>
<td>17 % (6/36)</td>
</tr>
<tr>
<td>RFQ corpus (19th c. born speakers)</td>
<td>1 % (7/717)</td>
</tr>
<tr>
<td>OH corpus (20th c. born speakers)</td>
<td>0.25 % (2/779)</td>
</tr>
</tbody>
</table>

The findings presented by Elsig are interesting, since one of Druetta’s (2002) arguments for different est-ce que constructions is that the not completely grammaticalized est-ce que can be divided by one or more elements. It seems that, for LaF, there may be no such thing as a not completely grammaticalized est-ce que (Druetta was working with European French), at least not positioned after the wh word. As for Druetta’s other argument for a non-grammaticalized est-ce que, that is, that the copula can vary in tense and number, Elsig doesn’t mention any instances of past tense or plural agreement on est-ce que, although he does not specifically state that it is not present in the corpora either. However, we have to mention here that the verb to be with the pronoun ce, prescriptively, does not always agree in tense or number, even when used in declarative context. For example, if the verb to be is followed by a preposition, it stays invariable (100), or when used with the first and second person plural pronouns, it stays singular (101), etc.

100. C’est d’eux seuls qu’on reçoit la véritable gloire. (Corneille, *Horace*, vol.3)  
    It is from them only that we receive true glory  
    ‘We receive true glory from them only.’

101. C’était nous les voyous.  
    It was us the rascals  
    ‘We were the rascals.’  
    (Rey (dir.), *Le Grand Robert de la langue française*, electronic version)

It is therefore very probable that c’est is the default form used by speakers, since the tensed or pluralized form can only appear in such restricted contexts. There would be no reason for the interrogative context to trigger more agreement than the declarative context already does. However, note that the same distribution can be observed with pre-wh c’est, which seems to
behave exactly like the declarative one just described. It can sometimes vary in tense (example 102), usually not in number (as proposed in Rowlett 2007: 184), but most of the time it stays invariable. Unfortunately, we do not have any usage details in LaF on this particular variant, i.e. *c’est* wh, because it was not included in Elsig’s results.

102. C’était qui qui était là?
   It-was who that was there
   “Who was it that was there?”

103. C’est/c’était qui qui sont venus?
   It-is/it-was who that are come
   “Who is/was it that came?”

More details on the distributional differences between pre-wh *c’est* and post-wh *est-ce/c’est* are given in the analysis section, as well as in section 4.2.5. Note that the pre-wh *c’est* has been previously considered as a marker of clefts in wh questions (Shlonsky 2009, Boeckx 1999), which is also what will be assumed here.

To summarize, Elsig (2009) proposes an account that can explain quite well all pre-verbal wh questions in LaF. However, there is still the problem of the in situ variant. If *est-ce que* is an affix to the wh word, why is *est-ce que* incompatible with the in situ position?\(^\text{56}\) Moreover, it is not clear in Elsig’s proposal how the affix gets attached to the wh element; does he assume a

\(^{56}\) Lefebvre (1982) states that a subset of wh words can appear in situ with *est-ce que*:

i. Tu as fait ça quand-est-ce?
   You have done it when-is-it?

However, Elsig found no such occurrences in any of the corpora he studied (2009: 194). There is a possibility that some wh words have a grammaticalized counterpart, since they can be used in isolation, but those are quite rare and highly marked.

ii. Quand-est-ce? ([kätes]
   When?

iii. Que-c’est? ([kœse]
   What?
larger lexical inventory (the wh word inventory in LaF would include *quand*, *quand-est-ce*, *quand-ce*, etc.), or does he assume that the affix is linked to SpecCP and gets affixed when the wh word is merged? Of course, postulating a very large lexical inventory is not very desirable under an acquisition point of view, and if the *est-ce que* gets affixed to the wh word, what is the filter that allows it to appear only sometimes? I will put aside the lexicalist analysis for the moment, for the reasons aforementioned. As we have seen before, there are grounds to believe that *est-ce/c’est* is in fact a phonological realization of the [Q] feature, under C, following the grammaticalisation of the copula. This accounts quite well for the various possibilities available to a LaF speaker, assuming that pre-wh *c’est* is distinct from the post-wh *est-ce/c’est*, in line with Druetta (2002, 2003), who proposed that there were two different *est-ce que* in French. I will develop this idea further in the next Chapter.

To conclude, we have seen in this section that the wh *est-ce que* construction has undergone two interlinked processes of grammaticalisation since its appearance in wh questions in Old French. The copula *est* and the pronoun *ce* both had full lexical value when it was first being used, the copula as I and the pronoun as a full NP, checking and assigning the appropriate features in the structure. However, the evolution of the language led to an upward and leftward reanalysis, and the copula went from I to C, from assigning phi-features to being more similar to an uninterpretable element (such as Case). In parallel, the pronoun went from being an independent XP that could act as the subject of the clause, to a clitic, to an affix, dependent on the head *est*. This resulted in having two lexicalized forms under C, *est-ce* and *c’est*. The exact evolution of the derivation will be presented in link to our proposal in the first sections of Chapter 4. The next section reviews the current assumptions made about declarative clefting in French. This will allow us to draw parallels between declarative clefting and the clefting that seems to be (or was) going on with the wh *est-ce que* construction.

### 3.2 French declarative clefting

This section addresses the syntax and semantics of clefting constructions in French. Its goal is to give an overview of the syntactic structures that are associated with declarative clefting constructions and to survey their main interpretive particularities. Although the focus is on clefts and pseudoclefts, copular constructions in general will be briefly mentioned, mainly in terms of interpretation (we saw in the previous section that the copula has been grammaticalised over
time, yielding a wh *est-ce que* construction that no longer possesses the characteristics associated with copular constructions – we will come back to this issue throughout this section). The first subsection (3.2.1) addresses the definitions and evolutions of clefting in French. The second discusses the interpretation of such structures, and the last (3.2.3) presents the syntactic analysis that is adopted in this work.

My approach to this topic aims at summarizing the main characteristics of this type of construction, and to use the literature that is available on the subject for our investigative purposes. The structures presented here (Clech-Darbon et al. 1999, Den Dikken 2006, etc.) are, as far as I know, all widely accepted, but note that the exact structure corresponding to French clefts is of no consequence to my proposal. We will come back to this in more detail in subsection 3.2.3.

**3.2.1 Definition and evolution of clefting**

Clefting is defined in the following way:

“A cleft construction is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or a relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions.” (Lambrecht 2001: 467)

This is a very general definition that fits ‘real’ clefts (or it-clefts) as much as pseudoclefts (or wh-clefts). Rowlett (2007) gives the following examples for French (the canonical equivalent is presented in (106)):

104. C’est Luc qui veut te voir. Clefting (Rowlett 2007: 184)

    it-is Luc who wants you see

    ‘Luke wants to see you’ or ‘It is Luke who wants to see you’

105. Celui qui veut te voir c’est Luc. Pseudoclefting (Rowlett 2007: 173)

    the-one who wants you see it-is Luc

    ‘The person wanting to see you is Luke’.
106. Luc veut te voir.
    Luke wants you see
    ‘Luke wants to see you.’

Rowlett (p. 184) mentions that in Con[temporary]French (the name he gives to the French system corresponding more or less to vernacular French), the c’est can vary in tense (only simple tenses, so that present, imperfect and future are the most commonly found), but it does not vary in number, i.e. c’est is always singular, ce sont, the plural equivalent, is impossible (although permitted in more formal states of the language – what he calls Mod[ern]French).

Historically, Rouquier (2007) has found that cleft constructions using c’est X que (as in 104), were introduced during the mid-Old French period (based on the same corpus used in her 2002 study on interrogatives).

Mirroring the early behaviour of the interrogative wh est-ce que, declarative clefts (104) started being used only in restricted contexts, mostly only in direct discourse, and the clefted element (the X in C’est X que + verb) was more often a proper name, almost always [+human] (Rouquier 2007: 190). Clefts were also often used following ‘perception’ verbs, such as know, see, show (conoitre, veoir, moustrer in Old French) (Rouquier 2007: 191). The parallel appearance and evolution (wh est-ce que was also selected by a restricted class of verbs when it first appeared in the language) of both constructions is not surprising, and it gives one more argument to the theory that est-ce que in interrogatives appeared first to fill a clefting (focus, emphasis) function. The declarative clefting construction was used more freely during the Middle French period; contexts of usage started to broaden, as well as ways to express clefting (see footnote 50 on the placement of the preposition within the cleft construction later in this section) (Rouquier 2007). By the end of Middle French period, usage started to resemble the modern-day one.57

57 In the non-Minimalist literature on French clefting, authors distinguish cleft constructions from another similar c’est+que+verb construction, the relative clause (Muller 2003, Rouquier 2007). They refer to the ‘autonomy of the clefted element’, which means that the sequence complementizer+verb that follows the clefted element is not a modifier. This criterion is the main one used to distinguish clefts (i) from relative clauses (ii).

    i. C’est Jean qui est venu. (Rouquier 2007: 168)
Many authors (Belleti 2005, Hamlaoui 2007) have mentioned that declarative cleft sentences are the most natural way to answer a subject wh question today.

107. Qui est-ce qui a mangé un biscuit?
who is-it that has eaten a cookie
‘Who ate a cookie?’

108. C’est Ella (qui a mangé un biscuit).
it-is Ella that has eaten a cookie
‘Ella ate a cookie.’ (Hamlaoui 2007)

For this reason, we consider the ‘real’ cleft to be more closely related to interrogatives than the pseudocleft (even if in English pseudoclefts are sometimes called wh-cLEFTs, in French they bear no resemblance to (interrogative) wh words). More evidence for this assumption is given in the next section.

An embedded relative within a cleft is also a good way to determine that the ‘outer layer’ is a real cleft. Example (iii) is from Old French (the bolded part is the relative modifying ‘father’, and the italicized passage is the cleft).

iii. C’est vostre pere ki ci gist,
Que cist villarz a tort ocis.
‘It is your dad, who lies here, who was murdered by this old man’.
(Marie de France (XIIth c., Lais, Yonec verse 536) (Rouquier 2007: 193)

In generative terms, we talk about headed relatives and headless (free) relatives: the relatives from cleft constructions are not complements of an NP, whereas full (headed) relatives like (ii) have to be complements (some have said adjuncts) of an NP. The main approaches concerning clefting vs. headed relatives are presented in section 3.2.3.
3.2.2 Interpretation of clefts and relatives

Copular constructions in French can be divided into two types: the predicational sentence (109) and the identificational (or specificational (Higgins 1979, den Dikken 2008) or contrastive (den Dikken 2008)) sentence (110).

109. Il est linguiste.
    ‘He is a linguist.’

110. C’est un linguiste.
    ‘He is a linguist.’

(van Peteghem 1991: 30, cited in Demol & Tobback 2009: 2)

In the French declarative copular (attributive) construction, the two readings are easily discernable, since il ‘he’ is used in the former, and ce ‘it’ in the latter. The literature on copular constructions is extensive, and the semantics of them quite complex. For our purposes, only the identificational reading is relevant, since the predicational can only be converted into a yes/no question.

111. Il est linguiste. → Est-ce qu’il est linguiste? (i.e. is he a linguist?)
    Predicational

112. C’est un linguiste. → Qu’est-ce qu’il est? (i.e. what does he do?)
    Identification.

I do not wish here to put into question or to evaluate the various proposals that have been made about predicational and specificational cleft constructions; rather, my aim is to show that cleft-like properties can be associated with wh questions in LaF, and that although LaF wh questions have not kept all structural properties, their general interpretation and distribution match what we can find in declarative contexts. The main parallels will be established in the next section, 3.2.3, on the syntactic properties of clefting constructions.

3.2.2.1 Prepositions and interpretation

This section represents additional information, geared towards linking the interrogative and the declarative clefting constructions. The tests presented here, based on the placement of
prepositions, can of course only be applied to a small subset of examples, i.e. the ones which contain a prepositional phrase. However, I believe that despite its limitations, the test of the preposition shows quite clearly the parallel behaviours of wh est-ce que interrogatives and declarative clefts.

Throughout the evolution of French clefts, prepositions ‘moved around’ and could be found either with the fronted clefted element (113), or with the complementizer (114) (Muller 2003, Rouquier 2007). There also exists a redundant form, in which the preposition appears twice, as in (115).

113. C’est à ma mère que tu as parlé.
   It-is to my mom that you spoke

114. C’est ma mère à qui tu as parlé.
   It-is my mom to whom you spoke

115. C’est à ma mère à qui tu as parlé.
   It-is to my mom to who you spoke (Muller 2003: 2)

Today, the placement of the preposition is seen as a good way to distinguish two types of specificational clefts. Clech-Darbon et al. (1999: 87) make the distinction between presentational clefts (broad focus), in which the preposition appears higher than the complementizer but lower

58 Muller (2003) actually proposes an order of evolution for the cleft construction based on the position of the preposition: the first to appear placed the preposition in the embedded clause (114), the second one was the redundant one, in which the preposition appeared twice (115), and finally came the one that we know today, with the preposition inside the focussed part, as in (113). Rouquier (2007) modifies slightly this proposition, and dates each of these stages of evolution. She divides the different types of prepositions and her results are quite detailed, but in general she found that the construction with the preposition in the embedded clause (114) appeared in the first quarter of the 13th century, the preposition within the focussed element appeared mid-14th century (113), and that the redundant construction, with the preposition in both positions (115), was the last to appear, only in the mid-15th century (Rouquier 2007: 187).
than the focussed NP, and contrastive clefts (narrow focus), in which the preposition appears higher than the focussed NP.

116. C’est le garçon à qui j’ai parlé. [Presentational]
   ‘It’s the boy to whom I spoke.’

117. C’est au garçon que j’ai parlé. [Contrastive]
   ‘It’s to the boy that I spoke.’

*(Clech-Darbon et al. 1999: 86)*

Example (117) clearly distinguishes between two options: it’s to the BOY that I spoke, not to the girl. Example (116) however makes no such contrast; we could imagine that the person speaking is pointing to a particular boy: THIS is the boy to whom I spoke, hence the presentational meaning. Another argument is that (116) can be used with the presentational voici ‘here is’ instead of the clefted c’est ‘it is’, but (117) could not.

118. Voici le garçon à qui j’ai parlé.
   ‘Here is the boy to whom I spoke.’

119. *Voici au garçon que j’ai parlé.
   ‘Here is to the boy that I spoke.’ *(Clech-Darbon et al. 1999: 87)*

Although historically the placement of the preposition didn’t seem to bring any meaning contrast, it seems that today’s system is quite different. Foci can be differentiated by the syntactic placement of the preposition (in addition to the contextual contrasts).59

59 In fact, Clech-Darbon et al. distinguish between four different types of c’est ... que/qui foci. In addition to presentational and contrastive, they also mention even-related and exclamatory comment foci. The first one optionally allows the NP to follow the complementizer (iii), and has the meaning ‘Something happened; it is that...’, which is why it is called ‘even-related’.

i Tu sembles inquiète. Qu’est-ce qui se passe?
   You look worried. What happened?

ii C’est le petit qui est tombé dans l’escalier.
   It is the kid who fell down the stairs.
If we want to posit a parallel with the interrogative constructions, one test is to manipulate the position of the preposition and see if focus constructions containing a wh element have the same presentational/contrastive differentiation as the declarative specificational clefts.\(^{60}\)

120.  C’est avec qui que tu parlais?
      It-is with whom that you were talking

121.  ?C’est qui avec qui tu parlais?
      It-is who with whom you were talking
      ‘Who were you talking to? (or ‘To whom were you talking?’)

122.  C’est à quel endroit que tu dors?
      It-is at which place that you sleep

---

iii  C’est que le petit est tombé dans l’escalier.
      It-is that the kid fell down the stairs.  \(\text{Clech-Darbon et al. 1999: 84}\)

Note that (iii) would not be allowed with any other type of c’est ... que focus. The other kind is called exclamatory comment, and ‘it is heavily constrained lexically [...] and from a temporo-modal point of view (reference to the future is preferred).’ \(\text{p. 85}\). Moreover, it seems to be rarer in LaF than in France French (which is the dialect that Clech-Darbon et al. describe). This one can be distinguished in that the premise and the focus sentence can be inverted, which would not be allowed with other types of foci.

\(\text{iv} \quad \text{Papa a acheté trois gâteaux. C’est le petit qui va être content!}
      \text{Dad bought three cakes. It-is the kid who is going to be happy!}

\(\text{v} \quad \text{C’est le petit qui va être content: papa a acheté trois gâteaux.}
      \text{It-is the kid who is going to be happy: dad bought three cakes.}  \(\text{Clech-Darbon et al. 1999: 86}\)

Neither one of these two types of foci will be included in this study, since they are not as relevant for the LaF interrogative system, and they are not at all common in LaF. The discussion will therefore include presentational and contrastive foci exclusively, which are the two types that are more traditionally thought of as being cleft constructions, at least by speakers of LaF.

\(\text{60} \quad \text{This parallel is also evidence that wh questions ressemble (it-)clefts more than pseudoclefts.}\)
It is interesting to note that (120) and (122), which would be the equivalent of a contrastive cleft, are definitely grammatical in LaF. (121) and (123), which would be the presentational clefts, are not completely ungrammatical, but they are surely slightly more awkward. Even with interrogatives there is a difference in meaning between the two types, though it is more subtle. It seems that in ‘contrastive interrogatives’ (120 and 122), the focussed NP refers to a finite set of people (i.e. ‘it is with who that you were talking, your friend or your brother?’), whereas ‘presentational interrogatives’ are more open (i.e. ‘It is who that you were talking? I don’t know him/her...’). Similarly, it is possible to say that ‘personne’ nobody is not a felicitous answer to (121) and (123), but it is perfectly acceptable for (120) and (122). Given these facts, it seems that there would be no reason to treat wh interrogatives any differently than declarative cleft constructions. Moreover, Clech-Darbon et al. call contrastive clefts ‘real clefts’, whereas presentational clefts are ‘presentational focus sentences’ (1999: 88). If we follow their theory, wh interrogatives in LaF look and behave like contrastive clefts, or ‘real’ clefts – so we can narrow down the typology even more: we assume that wh questions in LaF are somewhat equivalent to specificational contrastive clefts, although their structural properties differ slightly due to the presence of interrogation.

3.2.3 Syntactic approaches to French clefts and relative clauses

What kind of syntactic structure do cleft constructions have in LaF? There have been a few different proposals, but the main point on which there is disagreement is if the focussed constituent is independent from the following clause (‘c-clause’, according to Clech-Darbon et al. (1999), for ‘cleft-clause’), or if the whole thing is a single constituent. Many authors have proposed the latter, sometimes as focus movement (Reeve 2011 – I will review his proposal in

---

61 These are judgements made by three LaF native speakers for my purposes, but Clech-Darbon et al. have specific terminology to refer to the semantic properties of clefts and other c’est...que/qui constructions: the contrastive cleft refers to a contextual set U that is closed/small, whereas presentational focus refers to a contextual set U that is open/large (p. 104).
section 4.2.1), whereas others have proposed a no-movement construction involving a small-clause. Recently, the most influential works have adopted the latter idea, and various proposals around it have been made (the most famous being works by den Dikken 2006, 2008, etc.). The next section summarises proposals involving some type of small clause (free (headless) relative), and I conclude by following Clech-Darbon et al. (1999), who propose that “the post-focal clause is base-generated as a relative clause right-adjointed to a copular or identificational IP specified by ce.” (pp. 83-84). Before presenting the main proposals, it is important to note that the exact structure that we pick is not crucial for our purposes; others may work just as well. The important point to make is that there are enough parallelisms between French ‘real’ contrastive clefts and LaF wh interrogatives that it makes sense to propose a unified analysis. We will see that the adjunction relationship between the clefted element and the free relative, as well as the gap and operator within the relative are relevant for an analysis of wh est-ce que.

The parallel that we outline between LaF wh est-ce que questions and declarative clefts is more structural and semantic than it is selectional. This means that although it is safe to consider declarative clefting as a type of copular or predicative construction, the grammaticalisation outlined in the previous chapters eliminates this possibility for interrogatives. The est-ce/c’est in est-ce que is no longer a copula, or a ‘relator’ (den Dikken 2006) or ‘predicator’ (Adger & Ramchand 2003), and it does not take a predicate as its argument. Under recent accounts of predication, the functional head is filled by the copula, which has the subject of the predication in its specifier, and the predicate as its complement (others have proposed an adjunction relationship instead; see De Vries 2002 for details).

124. Relator phrase

\[
\begin{array}{c}
\text{RP} \\
\text{YP} \quad \text{R'} \\
\text{R} \quad \text{XP}
\end{array}
\]

125. \[\text{[RP NP John} [\text{R: is [AP happy]]}],\text{ where YP=NP and XP=AP}^{63}\]

---

62 Adger & Ramchand’s (2003) structure is called PredP, which is a little more restricted in what can constitute a head, but in general the functional phrase has a very similar function as den Dikken’s RP.
It is hard to posit this kind of syntactic relationship for a wh est-ce que question, since what would be the copula is followed by a CP, and the predicative or equative meaning that is associated with a copular sentence such as (125) is absent – see section 3.1 for evidence of the grammaticalisation of the copula.

The reasons outlined above lead us to opt for an analysis of clefting that, although not on par with other copular constructions à la den Dikken, have been widely accepted since the proposal involving CP adjunction of Chomsky 1977 (see Chapter 4 for Reeve’s (2011) analysis, which is similar to Chomsky’s proposal).

The analysis outlined here is based on prosodic facts from declarative and interrogative clefts (interrogative clefts being exactly the same as declarative clefts, but with interrogative intonation: this is not related to wh interrogative clefts as proposed in the present work); Clech-Darbon et al. propose that contrastive clefts (declarative and interrogative) have the same intonation contour as right dislocated topics; there is a reduplication of the final tone between the focussed phrase and the c-clause (pp. 95-100). The same is not true for other c’est ... que/qui constructions such as presentational or exclamatory foci. They show that the prosodic and semantic facts (such as the closed and small contextual sets mentioned in the previous section) clearly show that contrastive clefts pattern exactly like other narrow focus constructions in French. Moreover, according to them, the best syntactic structure that mirrors these intonational and semantic facts is one that is similar to the commonly assumed one, which involves a free or headless relative as the subject of the main clause. Their proposal is in (126).

---

63 See French example (109) above ‘Il est linguiste.’
Here, the CP is base-generated in a right-adjoined position to the IP. The clefted NP with the cleft element c’est is a ‘copular or identificational IP’. The main point that Clech-Darbon et al. demonstrate here is that the CP has no argumental status, therefore the ‘agreement’ in intonation with the clefted NP is simply derived from normal application of a default rule (p. 105).

Their proposal is interesting, but the fact that they propose that the lower CP is adjoined to the IP has to be explored under a syntactic point of view. It has been independently shown (Mathieu 1999, Shlonsky 2009) that wh in situ patterns like an adjunct with respect to island restrictions, and so it would be “chain-linked to C and [...] this chain is non-argumental and hence subject to R[elativized]M[inimality]” (Shlonsky 2009: 2). Is there any way to bring these two proposals together via some kind of adjunction, perhaps of the interrogative CP? Wh movement is traditionally thought of as A-bar movement, so an analysis without movement could still propose

---

64 Mathieu (1999) noticed the similarities between the behaviour of moved adjuncts and wh in situ. Moved adjuncts cannot be separated from their variables by negation, and he argues that it is the same for both arguments and adjuncts wh in situ, which is why he proposes that wh in situ is chain-linked to C; this can explain the restrictions associated to wh in situ. We have seen in earlier chapters (section 2.2.4) that those restrictions are controversial; therefore, without assuming the complexities of Mathieu’s analysis, we keep in mind that parallels between adjuncts and wh elements have been noticed, and that it may represent evidence towards an analysis such as the one that Clech-Darbon et al. propose (example (126) above), which include an adjunction relationship between the IP for them, and the wh in our case, and the main verbal domain. I come back to this possibility in the next Chapter.
A-bar positions for wh elements, whether they ‘officially’ have argument status or not. My analysis pursues this idea, and proposes a structure for interrogatives that is very similar to the structure proposed by Clech-Darbon et al., thus keeping intact the unifying properties of interrogative wh clefts and declarative clefts. Moreover, we want to keep the restrictive relationship between the clefted element and the relative clause, as was first proposed by Partee (1975).

3.3 Summary

This chapter presented the main characteristics of the est-ce que construction, which was shown earlier to be playing a major role within the wh interrogative system (Chapter 2). First, I have proposed a grammaticalisation path for est-ce/c’est, based on recent minimalist theories of grammaticalisation (Roberts & Roussou 2003, van Gelderen 2008a, b, Lohndal 2009). We have seen that the copula went from an I element to a C element, moving ‘upward leftward’ in the structure, and that the pronoun ce has undergone a typical process of cliticisation (for a recent contribution on the notion of cliticisation and grammaticalisation, see Schiering 2010), attaching itself to the copula/grammatical marker. In the second part of the chapter, we explored the clefting construction in French, highlighting the interpretation similarities between the cleft and the wh est-ce que interrogative, as well as some structural particularities that can be transferred to the wh (est-ce que) interrogatives. I showed that although the copular element of est-ce que has completely grammaticalised, making it distinct from a real clefting construction, it is still possible to establish parallels between cleft constructions and wh interrogatives.

The next chapter comes back on the structural particularities of both declarative clefts and wh interrogatives, and presents my complete analysis. I take theoretical proposals from Clech-Darbon et al. to propose a unified structure for all LaF wh interrogative variants. I will show that although wh est-ce que variants are no longer full clefts due to the grammaticalisation of est-ce que, some characteristics of clefts are still present. The exact function of the grammaticalised est-ce que in today’s LaF wh interrogatives is also explained, within the context of our analysis, as well as in the context of universals in wh interrogative systems (section 4.3).
Chapter 4

4 Proposal – Accounting for Variation

Two main conclusions are to be remembered from previous chapters:

i. wh est-ce que and its variants dominate the wh interrogative system in LaF, and represent the main source of variation (Chapter 2);

ii. although no longer a full clefted element following the grammaticalisation of est-ce/c’est, the wh est-ce que still seems to possess some clefting properties (Chapter 3).

This chapter re-addresses these characteristics, and clarifies them by proposing a complete analysis of all variants.

Section 4.1 addresses the puzzle of the in situ within the French system: how do we account for it, given all the facts highlighted in the two previous chapters? It reviews previous accounts of this apparent optionality of wh movement in French, focussing on more recent Minimalist accounts, as many works have addressed this puzzle in the last decade.

Section 4.2 describes my proposal thoroughly, based on facts exposed in Chapters 2 and 3. Diachronic as well as synchronic evidence are given to justify a structure that lacks wh movement and that possesses some properties of a cleft construction, such as a null operator and an adjunction relation between the wh element and the rest of the clause.

Following sections describe various facts and consequences of the proposal, including variation and distributional facts, the issue of covert movement, and the possible interactions that such constructions might have with prosodic facts that are specific to French (Mathieu 2012, Hamlouhi 2007, Déprez to appear). The chapter is concluded with a discussion of the repercussions that the proposal has for theories of the French yes/no interrogative system (section 4.4), a discussion of the use of sociolinguistic data in syntactic proposals (section 4.5), and of the contributions that this work can have for the theories of oralité (section 4.6).
4.1 Previous accounts of optional wh movement

French as a wh-movement language is puzzling, given the very real and quite common option of leaving the wh in situ, but researchers are still reluctant to propose something dramatically different from traditional analyses of wh movement. Many proposals in the literature assume the binary typological classification of Cheng (1991): languages have wh movement, or they are wh in situ languages. Cheng’s (1991) widely adopted Clausal Typing Hypothesis\(^{65}\) states that languages need either a question particle or overt wh movement to type a clause as wh interrogative. Therefore, on a typological level, it follows that a language that does not have wh movement necessarily has a question particle (at least in yes/no questions). Because French poses a problem for such a binary hypothesis, since it allows both wh movement and in situ, in addition to the fact that LaF might have a question particle in yes/no questions,\(^{66}\) the French interrogative system has been the object of many studies since the publication of Cheng’s typological work (1991). This section summarizes a few analyses of the French wh system that have been proposed in recent years.

The traditional wh movement analysis for French is well exemplified in Rowlett (2007). According to him, French is a wh movement language, i.e., movement is not optional: the wh element needs to front in order to check the strong [Q] features that gets merged on one of the heads within the CP domain. Under this analysis, examples of wh in situ are in fact not equivalent to the examples of fronted wh: in situ questions are “strongly presuppositional in a way that [fronted wh are not]” (Rowlett 2007: 198). There is no [Q] that gets merged with in situ questions, so movement of the wh is impossible. There are in fact two types of proposals: the ones, like Rowlett, which assume that French has wh movement, and the others, such as Denham

\(^{65}\) Cheng’s Clausal Typing Hypothesis is widely accepted, but is not without opponents. Bruening (2007), for instance, shows that the presence of wh in situ in certain languages is actually not related to the presence of a question particle in the language. He claims that: “crosslinguistically, languages of all possible types are attested” (p. 139). I leave the reader to decide if the facts presented in this thesis support Cheng’s or Bruening’s proposal, but I present my arguments with respect to Cheng’s typology, since it is more widely known.

\(^{66}\) Details about the LaF yes/no interrogative system are given in section 4.4.
(1997), which explain movement of the wh in French by some other kind of mechanism, such as focus movement. The first group claims that movement is not optional \textit{per se}; the fronted and in situ constructions are somehow different in the numeration, by a feature, a head, or even a whole projection, which is not merged in certain semantic contexts. For the second group, movement of the wh element is not triggered by a strong interrogative feature, rather, it is more like focus movement. In those cases, fronted wh has a focus interpretation that in situ does not have.

Denham (1997), for example, combines these two ideas, and proposes that some languages have optionality because of different numerations, and some other languages have optionality because movement is triggered by something other than an interrogative ([Q] or [wh]) feature.\textsuperscript{67} Her grouping of ‘optional wh movement’ languages is as follows: some languages have \textit{apparent} optional wh movement, and others have \textit{real} optional wh movement. For her, French is part of the first group, alongside Egyptian Arabic, Akan and Igbo (Niger-Congo), and Kikuyu and Kiswahili (Bantu).\textsuperscript{68} Movement of the wh word in these languages is explained by a topicalization or a focus requirement, rather than a wh feature checking requirement. Fronted vs. in situ questions therefore usually have a slight difference in meaning (due to the interpretable topic or focus feature). Denham’s treatment of French is lacking in exhaustivity however; she only uses data with the wh word \textit{quoi/que} ‘what’, and therefore states that fronted wh questions in French take the \textit{est-ce que} element obligatorily, which we know is not necessary the case with other wh words, in arguments as well as adjuncts.

Her proposal is original in the sense that it considers French as a language more closely related to languages like Igbo or Arabic than English when it comes to wh interrogatives, which has not been a very popular idea (most researchers still consider French a wh movement language - Poletto & Pollock (2004); Cheng & Rooryck (2000); Rowlett (2007); Yeo (2010); Bošković (2000), presented below). However, we have seen that there is no meaning difference between wh \textit{est-ce que}, wh only and wh in situ (see sections 2.2.2 to 2.2.4). Her assumption that \textit{est-ce que}......

\textsuperscript{67} This results in different numerations as well, with focus features that are sometimes present, but not in the same way as the first approach, as described in the previous paragraph.

\textsuperscript{68} I refer to some of those languages, and the ones from the ‘real’ optional movement group described below, in section 4.3.
que is obligatory with ‘fronted’ wh words is only true with the wh word qu’/que ‘what’, quoi/que ‘what’, so it would be quite inconvenient to have to propose (at least) two types of systems to account for the behaviour of all wh words.

Lassadi (2003) follows the same idea. He conducts a comparative analysis of Egyptian Arabic and French wh movement, and he concludes that neither actually exhibit optional wh movement. Rather, fronting of the wh element is triggered by a focus feature. Again, it follows that like Denham, Lassadi does not attribute movement of the wh element to a strong wh feature, but to a focus feature instead. The attractive part is that the process of fronting of the wh gets ‘fused’ into another process which is quite common and uncontroversial: focus movement. However, it has been proposed that French does not, in fact, allow focus movement in the way that English does. Consider the examples below, taken from Rowlett (2007: 182), who based his claims on Zubizarreta (2001), among others.

127. My dad I called yesterday.


    my dad I-have called yesterday

Example (128) would need a resumptive pronoun to be grammatical, as in (129) below, which is in fact not a case of focus movement, but rather of left dislocation (Goosse 2000, cited in Rowlett 2007: 183).

129. Mon papa je l’ai appelé hier.

    my dad I him-called yesterday

Therefore, it seems that French uses other processes such as topicalisation and left dislocation, but not focus movement. There might be some dialectal variation when it comes to the grammaticality of sentences like (128) – Rowlett himself mentions that ModF (the higher, mainly non-spoken register) probably allows focus fronting, but ConF (the vernacular register) does not. LaF definitely does not allow it.

But again, regardless of the possibility of focus movement in French, for this proposal to work, one must show a semantic difference between fronted and in situ variants, and that has not been demonstrated satisfactorily (at least not for LaF), see section 2.4 (and 4.2.3 below).
The second group of languages that Dehnam (1997) talks about are the ones with ‘real’ optional wh movement. These languages are Babine-Witsuwit’en (Athabaskan), Ancash Quechua and Malay. For Denham, wh fronting in these languages exhibits “island violations characteristic of moved constituents” (p. 68). The optionality comes from the fact that movement happens only if C is selected from numeration. Therefore, languages exhibiting optionality have the choice to select C form the lexicon or not. She argues that scope assignment and wh movement are not necessarily related, and gives examples of languages for which they are not (German and Romani; Denham 2000: 213).

If we assume that French is not one of the languages that Denham calls ‘apparent’ optional wh movement languages, could it be part of the ‘real’ optional movement ones? The answer to this question is not straightforward, since, according to Denham, languages like Babine-Witsuwit’en and Malay exhibit island violations that would indicate that the wh element has been moved. However, it has been shown that these judgements in French are hard to get, and that the islandhood tests are not conclusive. For example, Shlonsky (2009) studied French wh interrogatives, and warns his readers at the beginning of his publication: “some of these weak island effects [...] are subtle and controversial, [...]” (Shlonsky 2009:1). For example, if we try to see if French is sensitive to subject islands, we get the following:

130. ?/*Que [Marie arrive quand] est pas certain? In situ
     that Marie arrives when is not sure

131. ?/*Quand (que) [Marie arrive qui] (qui) est pas certain? Fronted
     when (that) Marie arrives (that) is not sure 69

We can see that no conclusion can be drawn, since both the in situ and fronted versions are equally bad. One of the reasons behind the difficulty in getting reliable judgements is certainly the access to different grammars that a French speaker must have in order to get these judgements. We describe this issue, the idea of diglossia in French, in section 4.5.

69 The complementizers in parentheses have been added in an effort to the make the examples ressemble LaF, to facilitate grammaticality judgements.
Bošković’s (2000) proposal is also similar to what has just been discussed, although lexical insertion at LF is the basis of his proposal. He assumes, like Rowlett, that French has a strong wh feature, and that movement is thus obligatory for the wh element. He assumes that if a complementizer with a strong wh feature is not present in the syntax to make the wh word move, it is inserted at LF. Since it is phonologically null, it will not cause the derivation to crash, and it is faithful to Chomsky’s (1995) view of strong features (Bošković 2000: 57). Therefore, he analyses French wh in situ as covert wh movement, i.e. movement that happens at LF, after syntax and spell-out. Section 4.3 below will show that in fact, tests do not corroborate the covert movement hypothesis – akin to the islandhood tests – and that long-distance agreement à la Simpson (2000) might be a better fit for the data, at least for LaF.

4.2 Proposal

My proposal follows from the historical findings that were presented in the second chapter, as well as from the synchronic facts and observations that have been made throughout this work. Our proposal is that in the LaF’s wh est-ce que construction, the structure is a remnant of what was historically a full cleft, with a ‘cleft-like’ structure in the left periphery, and absence of wh movement. Within this construction are also included wh in situ and ‘bare’ wh without inversion, since it is predicted that speakers would reanalyse a wh structure without wh movement to one where the wh is merged directly in the argument position. Given the omnipresence of est-ce que variants in today’s LaF (over 80% of use without in situ – Elsig 2009, see Chapter 2), this construction seems to be the default one for contemporary LaF speakers, and possibly the only one used in speech (we will come back to this possibility in the conclusion – Chapter 5).

4.2.1 The structure of wh est-ce que interrogatives

My analysis relies on what has been proposed for clefts in French; the clefted element has a free relative as an adjunct. I presented in section 3.2 a possible structure for a declarative cleft, based on Clech-Darbon et al. (1999), who proposed exactly what was just described; the clefted element is in an adjunct relationship with the free relative that modifies it (example 126, reproduced below).
In this structure, we have the adjunction relationship between the main IP and the headless relative, which conforms to restrictive modification (Partee 1975). The operator in the relative undergoes A’-movement to SpecCP in order to locally agree with the preceding clause. This is in line with what has been generally proposed for clefts; the adjunction allows the restrictive meaning between the clefted element and the rest of the sentence.

Other proposals for clefts include ones like Reeve’s (2011), who proposes a fully biclausal cleft for English (the second clause of the construction is not a headless relative like above). His proposal is exemplified below.

---

70 Here I assume that the verb est head-moves from V to N, within the subject NP in SpecVP, after which the NP moves up to SpecIP for EPP requirements. It is not clear what Clech-Darbon et al. propose exactly with regards to verb movement, since they focus mainly on the mapping of form and interpretation. The exact details of this are not important however, and I mention this issue only to clarify that I do not assume a movement such as from V to SpecIP to be felicitous, since it would violate the Head Movement Constraint (Travis 1984). Moreover, note that Clech-Darbon et al. propose a movement of the operator, which, we will see, will not necessarily have to apply for interrogatives, as it has been debated for relatives (Bouchard 1982). Details are presented in the next section.
It was John that Mary saw.

(Reeve 2011: 144, following Hedberg 2000)

In this case, a full relative is adjoined to the DP John, making it a headed relative – we will come back to the issues behind the differences between headless and headed relatives below.

Hamlaoui (2007) follows Clech-Darbon et al., and she gives additional evidence in favour of a structure like (132). She states that a biclausal structure à la Reeve (133)$^{71}$ cannot account for the prosodic and intonational facts of French. Hamlaoui maps the structure in (132) with the French prosodic structure, yielding (134), and argues that the headless relative structure maps to stress patterns of French.

$^{71}$ Belletti’s (2005) and Lambrecht’s (2001) proposals for clefts are similar to Reeve’s (exemplified in 133), although in their accounts, clefting involves Focus movement to a Focus projection.
Another piece of evidence in favour of a structure in terms of small clauses or free relatives (contra a full biclausal like Reeve’s) is that the free relative that is adjoined to the cleft clause is a ‘regular’ free relative, meaning that it has the same properties as a regular nominal-headed relative, i.e. it contains a gap, and it is linked semantically and syntactically to the main clause (see example below, based on De Vries 2002). The structure presented below is an example of a nominal-headed relative, which is a modified (simplified) version of De Vries’s (2002) proposal.  

---

72 s = strong; w = weak; IntP = intonational phrase; PhonP = phonological phrase; PrWrd = prosodic word

73 Note that many different proposals have been put forward about the syntax of headed relatives since the 1970s. The main debates revolve around two important issues: is the relative an adjunct or a complement of the N (or D), and is there movement of the head noun from the inside of the relative to the NP (DP)? Nowadays, (restrictive) relatives are most often seen as complements of D, although there is no consensus. For a complete, clear and detailed summary of these issues, and a proposal in terms of complement of D, see De Vries 2002 (Chapter 3).
In this structure, the DP *le petit* moves up to SpecCP to check the wh feature present in the relative CP-head, and then only the determiner moves up to the position external to the relative CP. This is an alternative to proposing that the DP *le petit* gets merged directly externally to the CP, and that the wh feature is checked through an operator present in SpecCP (like tree (132) above), but it yields the same result; the relative CP contains some type of [+wh] feature, which is related, through movement or agreement with an operator, to a position external to the CP (DP in the case of headed relatives, IP in the case of clefts, and maybe another CP in the case of certain types of LaF interrogatives, see below). Note in addition that it has been proposed in fact that Quebec French lacks movement altogether within its relative clauses (Bouchard 1982, Vinet 1984, Roberge 1998), so the long distance checking between the operator and the wh element, as will be proposed later, seems to be closer to the reality, at least in LaF, than the movement version put forward by De Vries (2002).
Many proposals have been made for a parallel between relative constructions and wh-interrogatives: they both involve a wh-operator (even that-relatives), and leave a gap (see De Vries (2002: 116-117), although he assumes wh-movement, which need not happen, see Fabb (1990)). More precisely, headed relatives are defined the following way: “A headed relative is a syntactically complex modifier involving abstraction over an internal position of the clause (the relativization site) and connected to some constituent it modifies (the relative “head”). In the standard approach, abstraction is syntactically implemented by means of an unbounded dependency between the relativization site and a relative operator taking scope over the whole clause.” (Bianchi 2002: 197). The definition of a free relative is similar, except that the relationship with the “head”, i.e. the noun, is absent. We can see that the ‘relative’ part of both trees above is identical; both clefts and headed relatives involve a relative clause in which an operator has scope over a variable (gap). If we apply the syntax and semantics of these constructions in a purely mechanical way, as an exercise to compare the declarative and interrogative constructions, we get the structure below (note that we do not assume (136) to be right, for reasons that will become clear in the next section; this tree will be revised in (149)).

136. C’est qui qui est tombé?
   it-is who who has fallen
   ‘Who is it who fell?’

```
   IP
      /\   \\
     IP     CP
        /\       /\      \\
       C’estj   Opy     C'
         |       [ +wh ]  \\
        I         C       IP
          |         |       \\
         VP       qui     t_is tombé
            |       |         \\
           V     qui       t_j
             |     [ +wh ]  \\
            CP       [ +Q ]
```
We will see that this structure has many flaws, but it is a reasonable starting point, since it possesses the same properties as trees (132) and (135) presented above when it comes to the relative CP. The difference is within the ‘cleft’, where the argument of the copula is an interrogative CP instead of a DP. This is necessary since we need to have a place for the [Q] feature, and, more importantly, we need a place for the possibility of having est-ce or c’est after the wh word ((136) above could be pronounced ‘C’est qui est tombé?’, or ‘C’est qui c’est qui est tombé?’, for example). The interrogative CP is the complement of the copula, which might be a problem for our analysis, since we have seen in earlier chapters that the c’est that we most often find within LaF wh questions has lost its status as a copula (i.e., its selective properties). Additionally, we have also seen that the pre-wh c’est seems to have a different status from the post-wh one, which can appear as est-ce, c’est, or null. The ‘cleft’ part of the structure must therefore be re-worked to account for these facts. The next section will present the historical facts that have been exposed in the first part of Chapter 3, which argued for a ‘grammaticalisation’ of the clause. The facts related to the evolution of the construction lead us to propose a structure that includes a free relative, with an operator and a gap, possessing the same properties as described above, but which has been reanalysed as a matrix clause in direct wh interrogatives – more details to follow.

4.2.2 The structure of wh-est-ce que interrogatives: diachronic evolution

For Old French, it is uncontroversial to posit a clefting structure for all est-ce que questions. During this period, we have seen that there were two possible structures for wh questions, which were not freely interchangeable: one for wh est-ce que questions, and one for fronted wh with inversion. The first example below corresponds to the latter.
Por c’ai ocis tante bele jovente?
‘Why have I sacrificed such great youth?’
(Charroi de Nîmes (1000-1500), cited in Brunot & Bruneau 1969: 481)

The structure in (137) represents a typical example of fronted wh with inversion; the wh moves up from the lower, merged (probably adjunct) position within the VP domain to SpecCP in order to satisfy the strong wh requirement. Note that in the example given we have a null subject (Old French allowed pro-drop, although in a more restricted way than Modern Spanish, for instance; see Roberts (1993)), and the verb is under C (Old French was V2 in matrix clauses, and had V-to-I-to-C movement). I propose, not controversially, that this structure has not changed since; in Standard French, fronted wh with inversion is still formed with the same mechanisms, minus the default I-to-C movement\textsuperscript{74} and pro-drop. A modern wh fronted with inversion would have the same structure as the English one below.

\textsuperscript{74}I-to-C is still possible in subject-verb inversion contexts, see below.
Who did Medea poison?

In this case, wh movement triggers movement from I to C, of *do*-support in the case of English, and of the auxiliary or the main verb in the case of French (French has V to I movement, so the main verb can end up in C).\(^75\) Example (139) presents a Modern French example of inversion, to show that the structure is exactly the same as the Old French one in (137), as well as the English one in (138), with the auxiliary moving to I, and then to C.

---

\(^75\) The distribution of French inversion is more complicated than what has been described, i.e. complex inversion, etc. However, the exact details of it are not necessary for the present work, since I consider the inversion construction distinct from the wh *est-ce que* construction. For details on the analysis of French wh inversion, see Rizzi & Roberts 1989, Kayne 1994, and Kayne & Pollock 2001 (on stylistic inversion).
In addition to the inversion structure available in Old French, there was also the clefted construction, characterised by the use of the clefted elements *est-ce que* or *c’est que* (with or without the pronoun *ce/c’*). As was shown in Chapter 3 of the present work, linguists (Rouquier 2002, Marchello-Nizia 1995, etc.) agree that this construction was one of emphasis, and that it could not be used interchangeably with inversion. It is also agreed that this emphatic meaning disappeared some time during the Middle French period. The clefted wh construction could be represented in the following way, assuming the clefted structure proposed by Clech-Darbon et al. (1999), as presented above (tree (132)).
Qui est chou qui tient sa court...
who is it who holds his/her court...
‘Who is the person who holds his/her court where I should have held mine?’
(example adapted from Rouquier 2003)

Both structures, the inversion one and the clefted one, involved traditional wh-movement to SpecCP, except that in the latter it is restricted to the main clefted clause. The embedded clause is a regular free relative, with an operator in SpecCP and a gap at the place of the argument. Old French did not have I-to-C movement in embedded clauses, which is why the complementizer can appear under C2.

Notice that in the clefting construction, the verb to be is a full verb, able to assign case to its subject, in this case the NP\textsuperscript{76} chou (strong form of today’s clitic ce). The chou is a full NP that can satisfy the EPP feature of I, and can receive phi-features from the verb.

As mentioned above, the structures in (137) and (139) are not exactly in competition; the cleft construction is marked pragmatically, and is still not compatible with adjunct wh words during this period (pourquoi ‘why’ will be the first one to be allowed, during the 14\textsuperscript{th} century; see

\textsuperscript{76} It is not clear if the pronoun at this stage was a full DP, or an NP. It is not crucial to the analysis, the important thing being that it is still a full XP, not just an X\textsuperscript{0}.
Chapter 2 for details). The main only difference between the cleft structure presented in the previous section and the one in (140) is that in the latter, the relative CP is adjoined to an interrogative CP instead of an IP. I keep the exact same structure because I assume the same interpretation, and most probably the same type of prosodic contour assumed by Clech-Darbon et al. (see section 3.2).

It was during the Middle French period that the element est-ce que extended its usage to all wh words, and that we find the first examples of merged orthography (esse instead of est-ce; see Chapter 2). Moreover, we have seen in Chapter 3 that the element est-ce/c’est has undergone complete grammaticalisation since Old French; the consensus among historical linguists is that the marked emphatic interpretation was disappearing during the Middle French period. The exact dates of the changes, semantic and structural, are not known, but since grammaticalisation is assumed to be very gradual (van Gelderen 2008a, b), it most likely happened over many centuries. We can consider the Middle French period to be the ‘intermediate’ phase, from one state of the language, where the structure involved a full cleft structure, to the other, in which the structure most likely looked like some type of ‘impoverished’ cleft structure. Other than the grammaticalisation of lexical elements, which would have consequences in the syntax, I assume that the structure presented in (140) did not change much during the Middle French period, except that I-to-C movement disappeared in declarative main clauses with the loss of V2.

The grammaticalisation processes happening during that period concerned the pronominal element ce (in est-ce and c’est), and the copula est ‘to be’ (in est-ce and c’est). The change affecting ce can be described as a typical cliticisation process; we can see the intermediate stage as being the one in which the element being reanalyzed goes from XP, in Old French, to X^0. Recall the grammaticalisation cycles discussed in Chapter 3:

141. \textbf{word} \rightarrow \textbf{clitic} \rightarrow \textbf{affix} \rightarrow \emptyset \quad \text{(from Hopper & Traugott 1993: 7)}

142. \textbf{Feature Economy} \quad \text{(van Gelderen 2008b, 2010: 145)}

\begin{align*}
\text{a. Adjunct/Argument} & \quad \textbf{Specifier} & \quad \textbf{Head} & \quad \textbf{Affix} \\
\text{semantic} & > & [iF] & > & [uF] & > & - \\
\text{b. Head} & > & \text{Higher Head} & > & 0 \\
[iF]/[uF] & & [uF]
\end{align*}
With the pronoun *ce*, this period can be seen as a window allowing us to witness steps 1 to 3 of the Hopper & Traugott’s cycle (bolded), and steps 2 to 4 of the cycle presented in the Feature Economy Hypothesis (bolded) (which represent in reality the same process, see details in Chapter 3). In Old French, as exemplified in the structure (140) above, the pronoun is a full XP, generated in the specifier of IP as the subject of the copula. Note that *ce* in this example has the strong form *chou*, which has now disappeared; this is another piece of evidence in favour of the grammaticalisation of the pronoun: strong forms of this particular pronoun, *cui*, *chou*, etc., no longer exist in Modern French, which might suggest that the lexical item is no longer a full XP, since it can no longer bear stress (see section 3.1). According to Rouquier (2003: 351), the form loses the option to move above the wh word; word orders such as the example below, from the Old French period, stopped being attested mid-fourteenth century (Rouquier 2003: 351).

143. Ice que est que je di ore?
   It (*ce*) what is that I said ADV.
   ‘What is it that I said then?’
   (Narcisse (1150-1200), cited in Rouquier 2003: 348)

The only possible place of realisation of *ce* is after *est*, or sometimes just before, it gets closer and closer to the copula, and by Modern French it could no longer be separated from it. This is evidence in favour of the pronoun going from lexical word to clitic, and maybe even to affix (see details in Chapter 3) – which is a change that has been described numerous times in the literature, as cliticisation. The intermediate stage, where the pronoun is no longer a full XP but now a head that gets attached to the head I, with the copula, is represented below, and I hypothesize that this ‘intermediate’ stage was most likely active during the Middle French period.

Note that *est-ce que* was never found in embedded contexts, whereas *c’est que* was found, both in Old and Middle French, both in matrix and embedded contexts (Rouquier 2003: 355-356, based on an observation made by Jensen 1990).
Qu’est ce que je voy?
what-is it that I see
What am I saying?

In this structure, the pronoun was reanalysed as a head, which gets attached to the head I by a process of cliticization.\(^{78}\) The head I is most likely defective, since it is not clear if it still possesses all phi-features associated with the IP domain: since this is the intermediate stage, it is very likely that the copula would show variation when it comes to tense agreement, see explanation below. The head C is empty, but it still bears the interrogative features necessary to type the clause interrogative and to have the wh element front.

As for the copula itself, it is more difficult to trace its exact evolution. We know however that in Modern French, *est* in *est-ce* and *c’est* no longer has the possibility of varying in tense and number. This loss of phi-features is also most likely linked to the fact that since the pronoun could no longer satisfy the EPP requirements of the subject position of the copula, the copula itself lost its EPP. We have seen that evidence for each step of the copula cycle, from a copula to

\(^{78}\) Whether the clitic undergoes lowering to become attached to the copula, or it is merged directly under I is of no consequence for our purposes. Any theory of cliticization could apply (see Auger 1994 for details on cliticization and grammaticalisation in Quebec French, or Rizzi 1986).
a grammatical marker (Lohndal 2009), are scarce, both in French and crosslinguistically. What we can state here is that *est* in wh *est-ce que* questions is now completely devoid of phi-features, whereas it seemed to possess them in Old French (Elsig 2009; see Chapter 3). The grammaticalization of the copula seems to be linked to the loss of V2 in Old French, although, as just mentioned, the evidence for the exact date of the completion of the grammaticalisation is lacking. Since the loss of V2 means the loss of I to C movement, it is very likely that the reanalysis happened exactly during this period, and that the copula got reanalysed as C; however, since V2 was never the only word order option to Old French speakers, the date of this change remains an hypothesis.

More evidence in favour of the hypothesis that the copula loses its verbal properties (phi-features, EPP) is that *c’est* (no longer a bare copula, but now a complex lexical item) dissociates itself from the verbal domain. During the Middle French period, the type of verb that could select *est-ce que* constructions was no longer restricted. Whereas in Old French it was almost exclusively *tell*-type verbs, in Middle French other verbs such as perception verbs (see *voir* ‘see’ below), can be used with the *est-ce que* construction. We can explain this fact by proposing that the functional head I, associated with copulas, was reanalyzed as a higher functional head, in this case C, so it no longer had any selectional effect. *C’est* went from being merged in I to being merged directly into C. This change might have also happened earlier and have been a reanalysis of I-to-C movement as Late Merge (see Chapter 3); we do not have enough evidence here to know without a doubt if the process has been only an upward reanalysis or a loss of movement (Late Merge). I hypothesize that both forces contributed to this change, which is an expected one according to the Feature Economy Hypothesis; if a head is going to be reanalysed, it will go upward in the structure.

145. Feature Economy Hypothesis (van Gelderen 2008b, 2010: 145)

b. \[
\text{Head} \rightarrow \text{Higher Head} \rightarrow 0
\]

\[
[iF]/[uF] \rightarrow [uF]
\]

Note that this change from I to C for the copula is also quite plausible in terms of context: the grammaticalisation process seemed to affect only *est-ce/c’est* when following a wh word, and only in interrogative contexts. It is therefore conceivable that learners would reanalyse this element as more closely connected to the higher interrogative CP domain than to the lower IP
tense domain. Note that this whole process only happened within the construction treated here, i.e. not all ‘to be’ copulas in French are now under C. In fact, as was discussed in Chapter 3, no claim is made at all for any other French copular construction.

A consequence of the copula no longer behaving like a verb is that it could no longer select the wh element; in traditional wh movement systems, the wh is merged within the VP domain to satisfy case requirements, and then moves up to SpecCP to check the strong wh features. However, if the IP within the cleft does not have the ability to assign case anymore, then merging the wh element low is no longer needed. In such a structure, the case features on the verb in the free relative (CP₂) are the only ones in need of being checked (valued), which is why the operator is there. Moreover, the operator agrees with the wh DP, without a need for the DP to move at all (we will come back to the mechanics of this long distance agreement in section 4.2.3). It is therefore hypothesised that, along with the reanalysis of the copula under I, which is most likely linked to the loss of V2 (which is a loss of V to C movement), wh movement also disappeared within the clefted clause, all approximately simultaneously (some time between the end of Old French and the end of Middle French), each phenomenon being the cause or consequence of the other.

Finally, if est is no longer a copula and if it is reanalysed as being under C instead of I, and if the wh word no longer undergoes movement from VP to CP, the whole IP projection within the clefted clause does not fulfill any syntactic function. The interrogative clefted CP no longer needs to select an IP, yielding a much reduced, verbless cleft construction. Such a structure, although not syntactically sound, would look like the one presented below (which will be slightly revised in (147)).
Qu’est-ce que tu vois? (or ‘Que c’est que tu vois?’)

‘What do you see?’

This is the structure that yields the one that we have today. Such a structure must lead to reanalysis, since it is not felicitous; a ‘normal’ CP would need to select its IP – how can a main clause have no verb? The proposed reanalysis made by speakers is that the headless relative is now considered the main clause, and the CP₁ is now an ‘atrophied’ clefting device, which has become an interrogative device. I do not propose that the structure in (146) was ever active for LaF speakers; rather, I believe that reanalysis happened to ‘fix’ this structure, and that speakers assume (147) instead, in which the wh element is adjoined to the main clause instead of the wh element that contains the main clause.⁷⁹

⁷⁹ Note that the lack of operator movement within CP₂ is a good consequence of our proposal, which lacks wh-movement entirely. An argument that makes us believe that this is in fact the right analysis is that ever since the beginning of the 1980s, linguists have said that relative clauses in Quebec French (LaF) lack movement altogether, and that the relative pronoun agrees with its antecedent in a resumptive-like way (see Bouchard 1982, Vinet 1984 and Roberge 1998). I do not claim here that the process in relative clauses is exactly the same as the one proposed here, but since the lack of movement is already attested in relatives in LaF and that the clefting structure adopted here is quite similar to the one for full relatives, these similarities are highly desirable.
A structure like this one can account for all variation in LaF, and explains in a straightforward way the distributional facts of all elements (see section 4.2.5 for details). However, this is not a structure that has been associated to wh interrogative constructions in French, so one might be tempted to analyse the data using a more familiar structure, such as Rizzi’s (1997) exploded CP. I present in (148) a structure, adapted from Tailleur (2009), which shows that Rizzi’s theory could be detailed enough to provide all the projections needed for the distributional realities of LaF interrogatives. However, for the specific distributional facts that have been described so far in this thesis, we will see that such a structure would not be ideal, since it contains flaws that can be avoided if we assume (147).
C’est qu’est-ce que tu vois?
It is what is-it that you see
‘What is it, that you see?’

Two key elements are missing from the structure in (148) to account for all the LaF facts. First, the semantic relationship between the wh word and the main verb is lost; in the structure in (147), the wh word is adjoined to the main clause, which conforms to restrictive modification (Partee 1975), as mentioned earlier in this chapter – this restrictive meaning is in fact a remnant of the full cleft, since there is a restrictive modification relation between clefted elements and the free relative that is adjoined to it (see Clech-Darbon et al. 1999). This restrictive meaning is also the main reason why a single CP with multiple specifiers is not proposed; the Spec-Head, just like the exploded CP, would not allow the desired semantic relationship. Moreover, we have seen that the wh est-ce que construction has an intonational pattern that can be directly mapped to a structure like (147) (Hamlaoui 2007) – see example (134). It would be difficult to try to explain why, in example (148) above, the complementizer is not part of the same intonational phrase as the wh word, whereas when they are located in different CPs, the mapping is straightforward. Finally, in a structure like Rizzi’s, one could hardly propose that the wh element is merged directly in the left periphery, since agreement created by the presence of the operator is lacking. For all of these reasons, I propose a structure that resembles more a clefting construction, i.e. (147), than a traditional wh movement structure. We will see in section 4.3 that
such a proposal is more in line with analyses of languages that have wh interrogative systems that mirror LaF’s, such as Malay and Niuean.

The structure in (147) above accounts for all variants with a wh word followed by any version of est-ce que. It is proposed the est-ce/c’est, in variation since Old French, became in fact phonetic realizations of the [+]Q feature, merged directly under C⁰ – which has been proposed before, see Rizzi (1997) and Rowlett (2007). The possible realizations of [+]Q in LaF are therefore est-ce [œs], c’est [se] or null. When speakers want to utter a ‘real’ clefted question, which contains a special emphasis on the wh word, they can use the clefted element before the wh word (example (149) below). I follow in this sense Druetta (2003) who proposed that there are, in fact, two different est-ce que in French; one completely grammaticalized, and another one that is not (see Chapter 3 for details). In fact, as was discussed earlier in this thesis, the c’est that is not completely grammaticalized is the equivalent of the clefting element that we find in declarative clefts: the copula est following the inanimate pronoun c’. The copula corresponds to a defective I, as mentioned by Reeve (2011), and can vary in tense but not in number. The ‘real’ clefted questions in LaF would therefore be exactly like the structure we ended with in the previous section, reproduced in more detail below. Note that the pronoun c’ has a clitic status; although its exact status does not change our proposal, I consider it to be merged directly under I, as a ‘parasite’ to the copula (it could also get merged in SpecIP – Rowlett 2007: 184).
C’est que c’est que tu vois? (or ‘C’est qu’est-ce que tu vois?’)

‘What is it, that you see?’

In such a structure, the copula in the higher projection can vary in tense (and sometimes in number, but we have seen (p. 66) that this is rarer in French, even in declarative clefts (Rey: Le Grand Robert)).

80 We are not excluding examples such as the following:

i. Qui c’était qui était jaloux?

who it-was that was jealous

‘Who was it who was jealous?’

In this case, we would posit a Focus or Topic projection above the high IP, to allow for the wh element to precede the copula. Note however that examples such as (i) seem to be rarer and more restricted (mainly in embedded clauses, not necessarily possible with all wh elements) than examples such as (149), in which the copula precedes the wh element. I refer the reader to Druetta’s (2003) work, which explains in detail the two different types of est-ce/c’est, and he mentions the possibility – although not very common
To summarize, we have seen that real clefted questions in LaF are now generated as in (149) above, whereas regular wh questions, any variant of the wh *est-ce que*, are generated as (147), reproduced below.

150. Qu’est-ce que tu vois? (or ‘Que c’est que tu vois?)
what is it that you see
‘What do you see?’

All variants of wh *est-ce que* have the above structure; however, we will see that the in situ one, where the wh word is spelled out in the argumental position, lower in the structure, cannot be accounted for in the same way. The next section addresses this issue.

---

– of (i), and highlights the fact that (i) is explained through a structure similar to (149), not (147), since in (147) the copula could never vary in tense, and would not carry an emphasis meaning.

Let us be precise with respect to the use of the term *in situ*. Structure (150) above does not involve wh movement, and so the wh element is in fact in situ. However, we use the term ‘wh in situ’ as it has been traditionally used in the French linguistics literature, to refer to the construction that spells out the wh element in the argumental position, low in the structure (within the main VP domain). This is what is meant by ‘real’ wh in situ.
4.2.3 Accounting for in situ

Finally, the last variant that we need to account for is the in situ. We saw in Chapter 2 that the in situ variant represents slightly under 10 percent of all wh interrogatives in LaF (Elsig 2009), compared to almost 20 percent for France French (Coveney 2002). We also determined that it seems to be in the same system as the wh est-ce que variants, since it is not associated with a more formal or higher register (in fact, the social value associated with this particular variant seems to be fairly neutral (Elsig 2009)). Given the analysis that I have just proposed for the wh est-ce que construction, the presence of a ‘real’ in situ variant in such a system is to be expected; if there is no wh movement at all within this cluster of variants, then it is only one step further to a reanalysis from speakers merging the wh word lower. The mechanics of such a construction are also already available to speakers; there is long distance agreement within a wh est-ce que construction, since the operator in SpecCP₂ agrees with the wh word in CP₁ through m-command – I assume Simpson’s (2000) definition of long distance agreement, which is described in section 4.2.3.2.2 below.

For in situ, we cannot propose the same structure as the one proposed above, since the ‘clefted’ part would be completely null, which would make the more complex structure redundant and useless. Instead, what is proposed here is, in fact, the minimal structure that is required to account for the in situ.
Tu vois quoi?

you see what

‘What do you see?’

The wh word is merged as an argument (or adjunct, depending on the wh word) to the verb, and the interrogative CP merged above the IP domain (the exact details of the IP have been omitted for clarity’s sake) contains an operator that agrees with the lower wh element, therefore satisfying the wh requirements of the clause. This is the exact CP2 that was proposed earlier (150), without the second adjunct CP being merged. The CP in (151) also contains a [Q] feature to type the clause interrogative, which is under the head C, and is also satisfied via a spec-head configuration by the wh operator. In this case, the wh operator is obligatorily merged in SpecCP; the wh word itself could not appear there, since it would cause the derivation to crash – the wh DP needs to be merged in the argument position to preserve the argument structure of the verb. This is similar to what Butler & Mathieu (2004) (B&M) have proposed, that the wh in situ is like a split construction, involving a null operator high in the structure, and the wh word low. They do assume wh movement in related constructions, so their proposal differs from mine, but the concept of having long distance agreement (via either long distance agreement or covert movement) is not new. The following section argues that in situ in French does not involve any
type of movement, overt or covert, and that Simpson’s (2000) proposal of long-distance agreement for wh in situ can account for the facts observed.

4.2.3.1 The in situ variant and negation

The first part of this chapter was dedicated to summarizing the numerous proposals that have been made on the status of the in situ variant in French, since the fact that we have both fronted and in situ wh in the same language is puzzling. As we have seen in section 4.1, since Cheng’s (1991) typological work, it is generally assumed that a language checks the wh requirements of interrogatives either by overt movement to the left periphery (like English), or via a (null) wh particle (like Chinese). This section provides more evidence in favour of the proposal that wh in situ in French does not involve movement at all, overt or covert. I adopt Simpson’s (2000) proposal of long distance agreement.

We have seen in Chapter 2 that most proposals that aim at accounting for the presence of wh in situ in French assume that wh in situ is not compatible with negation.

152. *Il (n’})a pas rencontré qui?  
He (cl.) has not met who  
(Adli 2006: 17)

Even if Bošković, Cheng and Rooryck, Shlonsky, etc. all agree that (152) is ungrammatical, Adli (2006) demonstrates that, based on grammaticality judgements, (Parisian) French speakers accept (152). He mentions that it could not be the case that NEG would block LF movement, since most of his speakers, given the appropriate context, accepted wh in situ in negated sentences (without an echo interpretation). It is hard to explain such contradictory statements; are negation and wh in situ compatible in French, or not? The explanation for such an unclear status probably lies within the fact that negation and wh questions are not exactly compatible to begin with. As Coveney (2002: 165) points out: “Negative questions are a particularly complex area, especially pragmatically, and their history in French has been the subject of considerable controversy”. In fact, Coveney demonstrates, following Givón (1978), that even fronted wh negatives are more or less unacceptable as a request for information (and of course, negative wh are virtually absent from corpora – only one example in Coveney’s whole corpus). Most negative wh questions are in fact ‘positively-biased negative interrogatives’, and are often rhetorical or
echo (Coveney 2002: 174). Below is the only example found in Coveney’s corpus, which is clearly rhetorical.

153. bon – euh – qui ne fait pas d’erreurs?
   Ok – hum- who not make no mistakes?
   ‘Who doesn’t make mistakes?’ (Coveney 2002: 172)

These facts lead to positing that negation effects on movement, although real theoretically, are not to be taken literally, since speakers most likely have trouble accessing such judgements, because negative interrogatives are not part of their ‘accessible’ grammar (i.e. they don’t use such forms every day). Of course, judgements can be accessed given the appropriate context, but in the case of negative interrogatives, it is not clear if it is compatible with the LaF grammar discussed in this work. Adli (2006) convincingly shows, through different experimental methods, that French in situ does not seem to be sensitive to different interveners, as described in section 2.2.4.1. The next section shows that in fact, evidence for covert movement in wh in situ constructions is quite scarce.

4.2.3.2 Covert movement?

It has been proposed that wh in situ are in fact examples of covert movement, and that although the wh word is spelled out at the place of Merge, it still moves up to the left periphery to satisfy the [wh] requirements (Bošković 2000, see section 4.1 for details of his proposal). Given that I propose that overt wh movement does not occur in LaF, even when the wh is spelled out high, positing covert movement for in situ variants would be problematic for a unified system. This section argues that French in situ does not involve LF movement, and that the wh word is interpreted at the place of Merge, via long distance wh agreement with a wh operator (Simpson 2000) (section 4.2.3.2.2).

4.2.3.2.1 Island restrictions and adjuncts

The restrictions of wh in situ in French were presented in section 2.2.4.1. Different claims have been made about the behaviour of this construction, in particular when it comes to island restrictions, because, as Shlonsky (2009: 1) puts it: “some of these weak island effects – in the sense of Cinque (1990) and Rizzi (1990) – are subtle and controversial”. Mathieu (1999) and
Shlonsky (2009) show that wh in situ patterns like adjuncts with respect to island restrictions, and therefore it would be “chain-linked to C and [...] this chain is non-argumental and hence subject to R[elativized]M[inimality]”\(^{82}\) (Shlonsky 2009: 2). Wh movement is a type of A-bar movement, so an analysis without movement could still propose A-bar positions for wh elements, whether they ‘officially’ have argument status or not.

In this case, it explains why in situ is ungrammatical in the scope of negation and within wh islands (these two constraints, according to Shlonsky, are the only ones that are not controversial in the literature on French in situ (see quote above), but we saw that there still exists disagreement, which we come back to below).

154. */?Elle a pas fait quoi?
She has not done what

155. Qu’est-ce qu’elle a pas fait?
What-is-it that-she has not done
‘What hasn’t she done?’

156. */? Tu te demandes comment aider qui?
You cl. Wonder how to help who

---

\(^{82}\) **Relativized Minimality Condition** (Rizzi 1990):

X antecedent-governs Y only if there is no Z such that

i. Z is a typical potential antecedent governor for Y

ii. Z ccommands Y but does not c-command X.

**Typical Potential Antecedent Governor** (Rizzi 1990):

Z is a typical potential antecedent governor for Y iff

i. Y is in an A-chain, and Z is an A-specifier c-commanding Y, or

ii. Y is in an A’-chain, and Z is an A’-specifier c-commanding Y, or

iii. Y is in a head-chain, and Z is a heading c-commanding Y.
157. (?)Qui(-est-ce que) tu te demandes comment aider?
Who(is-it that) you cl. Wonder how to help
‘Who do you wonder how to help’?83

Again, these judgements are not the most robust, but they seem to hold when compared to their
fronted counterpart ((155) sounds better than (154), and (157) sounds better than (156)). In fact,
Adli (2006) demonstrates that in situ is acceptable in all contexts that were previously thought to
block it; in embedded contexts, with negation, with quantifiers, and with modals. The speakers
that he interviewed (twenty in total) were students from a Parisian university, and so not speakers
of LaF. However, this, in my view, has more weight when it comes to theory, since most
theoretical proposals on in situ (Mathieu & Butler 2004, Shlonsky 2009, Cheng & Rooryck
2000, etc.) had been based on a French (France) dialect, most likely Parisian (see Chapter 2
section 2.2.4.1 for details of Adli’s proposal). I assume here that evidence in favour of LF-
movement, as proposed by Cheng & Rooryck (2000) and Bošković (1998), is not strong enough
and that LF-movement cannot be assumed in the case of French wh in situ. This is in agreement
with the analysis proposed here, since there would be no logical purpose to propose LF-
movement when overt wh movement does not seem to occur either.

Boeckx (1999) had, in fact, previously proposed an analysis of wh in situ in terms of clefting.
The assumptions behind his proposal are contrary to what Adli has proposed and what has been
argued for in this thesis. However, I presume here that his analysis can work if the variants are
grouped differently.

Boeckx mentions that both (158) and (159) have the same interpretation, and that rien ‘nothing’
is not a felicitous answer for neither of them; however, answering nothing to (160) and (161), the
fronted and reinforced fronted variants, is completely felicitous.

83 These four examples are freely taken from Shlonsky (2009: 2), and adapted for LaF; note however that
Shlonsky considers (154) and (156) to be completely ungrammatical (there is no question mark beside the
star), though he opens the door in a footnote about the possibility to accept (154) given the right
intonation. He also does not consider (157) to be completely grammatical, even if better than (156).
These divisions do not correspond to the data presented in this thesis so far (see Chapter 2, and the distributional facts on variants in the present chapter, section 4.2.5). First of all, we have seen that the ‘reinforced fronting’ variants pattern more like in situ than with what he calls ‘fronting’ (which is in fact fronting with inversion, which has completely disappeared in LaF and almost completely disappeared in France French, and therefore is not part of the same grammar as the other variants).

Nonetheless, the arguments and the analysis in terms of clefting can explain some characteristics of the system, and seem in line with what is proposed here. He proposes that in situ variants are “focused, ‘covert’ cleft structures” (p. 71). Interestingly, he brings forward prosodic evidence – as well as syntactic and semantic – to support his claim, which is important, given the prominence of stress patterns within the syntactic theory of French clefting, and the focus/stress parallel that we can observe in focus-type constructions such as clefts. Boeckx assumes however the intervention effects mentioned above, and proposes that “wh-in situ is licensed by the closest CP head” (p. 78), and that “CP makes ‘movement’ of the wh in situ upper-bound ( [...] some movement is involved to license the wh in situ)” (p. 77). Note that Shlonsky (2009) explicitly demonstrates that Boeckx’s assumptions that (158) and (159) have the same interpretation are
misleading; in fact, in situ and fronted variants are the same, whereas what Boeckx calls ‘cleft’ (but with the c’*est* obligatory, see (159) above) are slightly different, since they involve a presupposition that is not present with either of the other variants (Shlonsky 2009: 8). This proposal is independently motivated by the clefting element itself: c’est, when it appears before the wh element, has the exact same value and distribution as the c’est in declarative clefts. It can vary in tense (usually either past (*c’était*) or present (*c’est*)), and it cannot be inverted (*est-ce*). The proposal here is that this particular variant has an extra topic projection, which gets merged above the wh projection (tree (149), reproduced below).

162. C’est que c’est que tu vois? (or ‘C’est qu’est que tu vois?’)

it-is what it-is that you see

‘What is it, that you see?’

To conclude, we have seen that although intervention effects have been proposed in the literature, no proposal assumes the same facts. To decide on a proper analysis, we relied on a study that has demonstrated extensive testing with a large number of participants (Adli 2006), and I also relied on usage facts as described in the first chapters of this thesis. My conclusions are the following:
i. There is not enough evidence for LF-movement of the in situ, so I posit that it is interpreted at the place of Merge;

ii. The only variant of the LaF wh system that seems to involve some slight difference in interpretation is the one involving a clefting element BEFORE the wh word.

This led me to posit a structure that lacks wh movement. But for such a structure to be felicitous, we need some mechanism that will allow the wh word and the wh operator ‘to see each other’, in order for the appropriate features to be valued/checked. The next section summarizes Simpson’s (2000) proposal that wh in situ does not involve LF movement, but rather the features on the wh element are checked via long distance agreement, which would be the only possibility for a construction without movement, overt or covert, as the one I proposed above (151).

4.2.3.2.2 Simpson (2000): long distance licensing of wh in situ

This section presents Simpson’s (2000) proposal that wh in situ does not involve covert (LF) movement. I will show how his analysis can be applied to the system developed here.

Simpson’s (2000) system relies on notions of parametrical locality; languages do not vary in terms of strong/weak features, they instead vary in terms of the locality constraints that apply to licensing (checking) of (wh) features. He argues that all island-type restrictions observed in Chinese and other languages can in fact be explained in alternate manners, which would render LF (covert) movement “theoretically unnecessary” (p. 66).84 His views are extreme in the sense that he proposes that within the Minimalist system, “the syntactic derivation is actually taken to terminate at the point of Spell-Out” (p. 71). This work will make no such assumptions. It will instead demonstrate that the ‘checking domain’, as proposed by Simpson, can explain straightforwardly how wh in situ can be licensed in the LaF system. I propose here, following Simpson (2000), that wh in situ can be licensed, i.e. ‘checked’, within a domain that is larger than the usually assumed Spec-Head configuration (Chomksy 1995); similar to multiple wh in

84 One of his main arguments concerns the asymmetries found in various languages between wh fronted and in situ; if there was wh movement at LF, we would not expect any asymmetries at all, since both variants would be identical, assuming that overt and covert movements are governed by the same restrictions, which is highly desirable within the Minimalist framework.
English, LaF can check the wh features on wh phrases anywhere within the sentence containing the interrogative CP.

163. The *wh*-checking domain in English [and now, in LaF] includes all sentence-internal positions m-commanded by the +Q Comp.

(Simpson 2000: 103)

Simpson explains the movement of single wh phrases (and the movement of the first wh in multiple wh) in terms of ‘ambiguous’ C-heads:

164. **The Triggering Hypothesis**: Wh-movement to an ambiguous C⁰-head is necessary to trigger C⁰ as a licensor for specifically wh-elements.

(Simpson 2000: 104)

The C-head is said to be ambiguous, that is, either focus, +wh+Q, or +yes/no+Q, prior to the movement of the first wh, as shown in (165) and (166).

165. *Did John give what to who?*

166. What did John give to who?

The raising of what in (166) allows the question to be interpreted as a wh question, not just a focus construction or, in the case of (165), a yes/no question (the example (165) is ungrammatical as a wh question).

For LaF, I do not assume movement at all, so there is no need for the Triggering Hypothesis. However, for the checking requirement to work, we do need to assume that wh in situ constructions are monoclausal, or at least that the wh element is not embedded.⁸⁵ This is exactly what was proposed in (151), reproduced below.

---

⁸⁵ If it was embedded (like the structure proposed in (147), repeated in (150)), it could not m-command or be m-commanded by the wh operator.
We can see that the wh element is m-commanded by the C head containing [Q], and, additionally, the wh operator c-commands the wh, checking the required wh feature.\textsuperscript{86}

This system, when applied to LaF, has many desirable consequences:

(i) all wh phrases have to be checked prior to Spell-Out, which makes them, according to Simpson, more like Case, therefore more like bearers of uninterpretable features rather than interpretable, which makes sense in terms of grammaticalisation (see Chapter 3);

\textsuperscript{86} As for the treatment of subject wh, whether we should treat them as in situ or not, see discussion in section 4.2.5.
We saw in Chapter 3 that the direction of grammaticalisation, in addition to being upward and leftward (Roberts & Roussou 2003), usually goes from lexical to functional, and from interpretable features to uninterpretable features (van Gelderen 2008b) – see section 3.1.1 of the present work. Since wh in situ is the most recent variant in LaF, it seems that the direction of change would be expected according to theories of cyclic grammaticalisation, as proposed by van Gelderen. I do not propose that wh elements in LaF now have only uninterpretable features, but since we saw in section 3.1.1 that there is some evidence to treat the [Q] feature spelled-out as est-ce or c’est as an uninterpretable feature more than as an interpretable feature, the idea that wh elements also carry uninterpretable features could be worth pursuing.87

(ii) long distance checking is analogous to the agreement between the relative clause and the clefted clause when the wh appears fronted.

We therefore have a unified checking system for wh interrogatives in LaF – agreement between the wh operator and the overt wh element happens minimally through m-command, since no wh elements move at all.

4.2.3.3 Licensing wh in situ – interfacing syntax and prosody?

Although wh in situ is part of the same system as the wh est-ce que variants, its structure is different, since it does not involve any type of clefting projection. We therefore need to have two interrogative CPs (two different CPs where [Q] can be merged): one which is selected as an adjunct by a main CP; and one which selects a regular IP as its complement. It is not clear why a speaker would pick one structure over the other. Mathieu (2012), although he does not assume the same analysis as the one proposed here, puts forward the hypothesis that licensing of the second kind of CP, the one for wh in situ, is triggered by prosody (suprasegmental properties). This section briefly describes the motivations behind linking prosody and wh interrogatives, by summarizing a few recent contributions to the field. I explore how the prosodic facts could fit into my proposal, by posing the hypothesis that licensing of the in situ option is conditioned by

87 Of course, this is an isolated observation, and more tests would need to be done in order to corroborate this hypothesis. I leave developing this idea, which could have unique theoretical implications, to future research.
prosody. The exact details of this hypothesis will remain to be developed in future research, but this section is important in a work on wh interrogatives, since the link between focus, stress (prosody) and wh, independently of clefting, have been exposed multiple times in recent literature (see Kahnemuyipour 2006, 2009, for instance).

In addition, it was mentioned above that the clefting structure proposed for French (by Clech-Darbon et al. 1999) was based on prosodic facts about the clefting/focus construction in French, and others have described the link between focused elements and prosody (Riallant et al. 2002). More recently, prosody has also been linked to variation within the wh system; see Mathieu (2012); Cheng & Rooryck (2000); Déprez et al. (to appear); Hamlouai (2007), (2010). Below is a summary of the main ideas behind the proposals that have been made for French.

An innovative proposal came from Cheng & Rooryck (2000), who suggest that an intonation morpheme is responsible for the possibility of wh in situ in French; they claim that the intonational contour of wh in situ questions exactly mirrors the intonation of yes/no questions, and that this intonation morpheme was therefore present in yes/no and wh in situ questions, but was absent from fronted wh questions. Their proposal has been challenged by several authors (Mathieu 2004, Hamlouai 2009), mainly based on native speakers intuitions, but Déprez et al. (to appear) are the first ones to have experimentally tested those claims. To determine the prosody of wh in situ questions, they tested 12 native speakers of French (from France) with sentences preceded by a detailed context (given the possible contextual influence on the licensing of wh in situ). They found that although a majority of examples exhibited a final rising contour, similar to yes/no questions, it was not categorical. For wh in situ content questions, there was inter-speaker variation, as well as a percentage of rising intonation lower than what was found for both intonation (i.e. declarative word order with an interrogative rising intonation) and est-ce que yes/no questions (Déprez et al. to appear: 7-8). They concluded that the rising contour was in fact not so much optional, but rather variable; it highly depends on the focus put on the wh word (p. 9). They therefore do not negate Cheng & Rooryck’s analysis; they nuance it: “we find their [Cheng & Rooryck’s] claim to not only be tenable, but also compatible with our experimental results, provided the target morpheme is allowed a variable phonetic realization, influenced by other factors that could manipulate the shape and height of the intonation contour.” (p. 9). They propose that those ‘other factors’ are related to the information structure, and they agree with Hamlouai (2010), who claims that the wh element is highly prominent in a wh in situ question.
Indeed, Hamlaoui claims that the wh word is the only new information in the utterance, the rest of the sentence having already been activated either overtly or through contextual cues. In the same vein, Déprez et al. bring up Kučerová’s proposal (2007, 2008, based on Czech data), who formalises this new vs. old information through a G-operator – a givenness marker that makes everything merged before it (lower) new information, and everything merged after it (higher) old/given information. Déprez et al. therefore claim that the licensing of French wh in situ happens through the influence of both syntax and prosody, as well as information structure. In their view, there is a G-operator that gets inserted right before the wh word, which makes all projections merged above the wh word given information.

All of these proposals could be linked to what Richards has recently proposed (2010) regarding wh interrogative domains – both Mathieu (2012) and Déprez et al. (to appear) heavily rely on his claims. Based mainly on Japanese data, he links the licensing of wh with boundaries of prosodic domains – a *wh-domain*, in which the wh word and the head C form the boundaries. This results in a pitch compression between the wh word and the C (in Japanese, where the C in clause-final). Richards’s proposal is interesting in that it formalises the links that have been said to exist between the syntax of wh interrogatives and prosodic domains/boundaries.88

Although the exact mechanism of licensing varies between authors – Mathieu, Richards, Hamlaoui, Cheng & Rooryck, Déprez et al. – there seems to be a consensus that prosody and information structure have to interact with syntax in order for French wh in situ to be felicitous. Whether it is through an intonation morpheme or a givenness operator, it is agreed that there needs to be some extra syntactic process going on, and that these seem to be more robust than the presuppositions or the semantic particularities that had been proposed previously. More work needs to be done to know which of the mechanisms proposed, if any, could account for the LaF facts, but regardless of how they are described, those intonational and informational cues may

[88] The exact details of his proposal can be found in Richards (2010), which I will not review here, but an analysis such as his merits consideration, since interfacing prosody and syntax, as a field of study, still has much to offer, being only in its first stages. In fact, French (and LaF) represents a great deal of data from which to develop such a theory, and much work must still be done to understand the ramifications of this interface.
share responsibility for licensing the wh in situ structure over the wh *est-ce que*, thus minimizing true optionality (the issue of ‘optionality’ was addressed in section 4.1).

### 4.2.3.4 Summary

This section has shown that the wh *est-ce que* variants in LaF are compatible with an analysis without any wh movement. The syntactic structure is one of an ‘atrophied’ clefted element adjoined to a main CP, due to historical evolution and synchronic distributional properties (see section 4.2.5 below), following many reanalyses by speakers. The in situ variant, which is part of the wh *est-ce que* variants in the sense that it is also found in spoken, fairly neutral registers (speakers would not use in situ in a written register, for example), is to be expected in a system without wh movement. The structure proposed for this particular variant is similar to the one described above since it also involves a wh operator in SpecCP and long distance agreement, and it lacks wh movement (both overt and covert); the difference lies in the fact that it only projects one CP layer. The next two sections bring more specifications to this proposal: section 4.2.4 provides details on the specific distributional properties of wh *est-ce que* variants in LaF, showing that although the variation looks extensive, ordering the elements and placing them within our structure renders the actual variation to be quite minimal; and, finally, section 4.2.5 discusses the asymmetries found in the French wh interrogative system caused by the presence of two wh elements for ‘what’: *que* and *quoi*.

### 4.2.4 Distributional facts about *est-ce que* questions

This section summarises some of the particularities that are associated with the set of variants of wh *est-ce que* questions in LaF. We have seen the different variants, and we have seen the structure projected when they are uttered. However, there are distributional details that are yet to be described. The structure proposed above gives some indication of word order and possible variations, but we must still explain examples such as the one below, taken from Elsig (2009: 244).
168. Il dit *qui c’est que c’est qui* frappe sur la maison comme ça, […]?  
‘He says: “who [lit. who it-is that it-is that] knocks on the house like that, ?”’

(OH.096.1089)

First, it is important to mention that, except for the wh- word, all the elements appearing in wh questions – the clefting element, wh- word, interrogative marker (*est-ce/c’est*), extra *que c’est* and complementizer – are optional. That the wh- word is the only obligatory element is easy to account for; it is the only element bearing the [wh] feature, so it has to appear in order to make the proposition a wh interrogative. Moreover, it has an argument status.

The overt complementizer is optional like the other elements, but the matrix C head is always generated. The complementizer can however be phonologically null; this is a similar rule to the complementizer deletion in relatives in English.

169. The guy (that) I talked to you about just arrived.

Under this analysis, the examples in (170) and (171) below are exactly equivalent, semantically and structurally, it is only that the complementizer is phonologically overt in the first one and null in the latter, which is a phenomenon that has been attested many times in LaF (see Martineau (1988), Dion (2003)):

170. Qui que tu vois?  
who that you see  
“Who do you see?”

171. Qui ∅ tu vois?  
who you see?  
“Who do you see?”

---

89 As mentioned in Chapter 2, OH refers to the Ottawa-Hull corpus (cf. Poplack 1989), and RFQ refers to the *Récits du français québécois* corpus (cf. Poplack & St-Amand 2007), both hosted by the Sociolinguistics Laboratory of the University of Ottawa.
The interrogative marker, which can be realised as either \textit{est-ce} or \textit{c’est}, is also optional in exactly the same way as the complementizer – it can be null, but the interrogative C head is always projected.

172. \begin{align*}
\text{C’} & \text{est qui } \varnothing \text{ qui est venu?} \\
& \text{it-is who that came} \\
& \text{“Who is it that came?”}
\end{align*}

173. \begin{align*}
\text{C’} & \text{est qui } \textit{est-ce} \text{ qui est venu ?} \\
& \text{it-is who is-it that came ?} \\
& \text{“Who is it that came?”}
\end{align*}

Note also that the first \textit{c’est} in (172) and (173), the one appearing before the wh element, can only have that form; it cannot be an inverted \textit{est-ce}. We have seen in Chapter 3 that this is a ‘real’ clefting element, with exactly the same properties as the one appearing in declarative clefts – it can vary in tense, but not in number (at least in vernacular French and LaF, see Rowlett 2007), and it projects its own IP.

Finally, the extra \textit{que c’est} is completely optional: it can appear or not, but always adjacent to the interrogative marker. Despite its complete optionality, the extra \textit{que c’est} appears only under certain conditions:

- It can only appear as an atomic element: the \textit{que} cannot appear without the \textit{c’est}, and vice-versa (examples 174 and 175), the \textit{que} cannot be stressed (examples 176), and it cannot appear as the subject variant of the complementizer (\textit{qui}) (example 177);

174. \begin{align*}
*\text{Qui est-ce } & \textbf{que} \text{ que tu vois ?} \\
& \text{who is-it that comp. you see}
\end{align*}

175. \begin{align*}
*\text{Qui est-ce } & \textit{c’est} \text{ que tu vois ?} \\
& \text{who is-it it-is comp. you see}
\end{align*}

176. \begin{align*}
*\text{Qui est-ce } & \textbf{QUE} \text{ c’est que tu vois ?} \\
& \text{who is-it \textbf{THAT}-it-is comp. you see}
\end{align*}
177. *Qui est-ce qui c’est que tu vois ?
   who is-it who-it-is comp. you see

   • It cannot be inverted; the est-ce cannot appear in this slot;

178. *Qu’est-ce que tu veux ?
   what-is-it is-it comp. you want

179. *Que c’est que tu veux?
   what it-is that-is-it comp. you want

   • It cannot be separated from the interrogative marker, it needs to be directly adjacent (to the right) to it;

180. *Que c’est tabarnak que c’est que tu veux?
   what it-is interj. that-it-is that you want

181. Que c’est que c’est tabarnak que tu veux?
   what it-is that-it-is interj. that you want

Repeated below is example (168), taken from Elsig (2009). He describes (p. 152) this extra morpheme as a marginal variant, mainly associated with an older demographic. This extra morpheme is quite rare nowadays, especially among younger people.

182. Il dit qui c’est que c’est qui frappe sur la maison comme ça, […]?
   ‘He says who [lit. who it-is that it-is that] knocks on the house like that, …?’
   (OH.096.1089 – Elsig 2009: 244)

If we consider all these different elements in terms of ‘slots’, we can visualise the possible variants as in Figure 10:

90 Tabarnak in LaF is a swear word, or a discourse marker, associated to a very popular (angry) register.
Figure 10. Possible elements appearing in fronted wh-interrogatives in LaF

<table>
<thead>
<tr>
<th>(0) Clefting element</th>
<th>1 wh-word</th>
<th>2 Interrogation reinforcer</th>
<th>3 Extra morpheme</th>
<th>4 Complementizer</th>
<th>rest of the clause (IP...)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c’est</td>
<td>qu’, que, où, qui, quand, etc.</td>
<td>est-ce, c’est</td>
<td>k(ə)se</td>
<td>que, qui</td>
<td>rest of the clause (IP...)</td>
</tr>
</tbody>
</table>

Although this way of presenting the data is quite abstract, we can transpose these slots to separate syntactic positions, XP or X₀, where each element is Merged. The tree below illustrates this, based on the tree proposed in (147) above.

183. Où est-ce que (que c’est) que tu vas?

When Slot 0 is filled, as in C’est où est-ce que tu vas?, then the tree is like what was proposed in (149), repeated in (162). The structure is the same as above, except that there is an extra IP merged above CP₂ to account for the topicalisation of the clefted element. Since the extra morpheme in slot 3 is associated with an older demographic, I believe that it is not compatible with a null interrogative marker. In Table 6 below are the possible lexical items for slots 2 and 3 in older LaF speakers’ grammar (grammar 1). In contrast, in a more recent grammar (grammar 2), the extra morpheme is no longer available, and the interrogative marker can be null.

---

91 See Appendix 1 for a table exemplifying all possible variants.
Table 6. Vocabulary items available for insertion in two different stages of LaF grammar

<table>
<thead>
<tr>
<th>Slot 2 (interrogative marker)</th>
<th>Grammar 1 (earlier)</th>
<th>Grammar 2 (later)</th>
</tr>
</thead>
<tbody>
<tr>
<td>est-ce; c’est</td>
<td>est-ce; c’est</td>
<td>est-ce; c’est; Ø</td>
</tr>
</tbody>
</table>

This is why I do not posit an extra projection in the syntactic structure, since the extra morpheme is no longer productive in LaF. Note that this is in accordance with the grammaticalisation process which affected est-ce/c’est: the structure got simpler as the interrogative reinforcer became more static and invariable, dropping the possibility of the extra morpheme. In addition, the more an element gets grammaticalised, the more likely it is to drop, which is why we only find the null option in the later grammar. For the period during which Grammar 1 was active, I hypothesize that the extra morpheme affixed itself to the interrogative marker (given the facts described above), possibly in a post-syntactic operation, but I leave the exact details aside, since it is not part of the state of language described in this thesis.

4.2.5 Que or quoi?

This section describes a reality of the French wh interrogative system which does not really have any influence on the proposal put forward. However, it is very likely that the asymmetry in the French wh inventory described here is one of the causes behind the preservation of the cleft structure, which is why I include a description of it here.

French has strong and weak pronouns, such as moi vs. me, toi vs. te, etc., and so do wh words. There are two versions of the inanimate what pronoun; the strong (tonic) quoi and the weak

---

92 In an earlier work (Tailleur 2009), I proposed that the extra morpheme behaved like a dissociated morpheme, inserted post-syntactically (Embick & Noyer 2006).
(unstressed) _que_ (which can undergo contraction in front of a vowel, such as in _qu’est-ce_, for example).\(^93\) The two inanimate _what_ forms are in complementary distribution:\(^94\)

184. a. Que cherchez-vous?\(^95\)
    
    what seek you
    “What are you looking for?”

    b. Vous cherchez quoi?
    c. *Vous cherchez que?
    d. *Quoi cherchez-vous? \(\text{(Plunkett 2000: 512)}\)

However, this is only true with the inanimate direct object pronoun, since the animate counterpart of _que/quoi_ only has one form, _qui_.

Subject questions are special in that they appear not only to disallow subject extraction, but do not allow in situ either:

185. a. *Que flotte dans l’eau?
    
    what floats in the-water
    “What floats/is floating in (the) water?”

    b. *Quoi flotte dans l’eau? \(\text{(Plunkett 2000: 513-514)}\)

However, this restriction disappears when the subject is animate (186), or when the _wh_- element is more complex (187):

\(^93\) It is not clear if _quoi_ can actually contract when followed by an interrogative reinforcer, such as “_quoi c’est que_” (/kwasek/ or /kɔseq/?). For a more detailed discussion, see Lefebvre (1982).

\(^94\) For a discussion about the properties of weak and strong _wh_- pronouns in French questions and relatives, see Plunkett (2000), Hirschbühler (1980) and Goldsmith (1978), among others.

\(^95\) Standard French only. In LaF, the _que_ would be replaced by _qu’est-ce que_. See the discussion about subject questions below.
186. Qui peut flotter?
who can float
“Who is able to float?”

187. Qu’est-ce qui flotte dans l’eau?
What-is-it that floats in the-water
“What floats/is floating in (the) water?” (Plunkett 2000: 514)

The exact reason and motivations for this distribution is beyond the scope of this thesis (it appears to be an ‘evolutionary accident’), but it is important to keep in mind that while there are only three morphological argumental whpronouns in French, there are technically four different forms; animate weak and strong (qui) and inanimate weak and strong (que, quoi). The choice of one or the other is never arbitrary, it entirely depends on the context.

The behaviour and position of qui also makes it impossible to tell, when it appears bare (without any est-ce que), if we should favour an in situ structure as in (167), or an atrophied cleft like other wh est-ce que questions (as in (150)). The two examples below show that both structures are possible, given that they yield the same word order. Example (188) shows the in situ possibility, and (189) shows the atrophied cleft possibility.
188. Qui vient?

who comes

‘Who is coming?’

189. Qui vient?

who comes

‘Who is coming?’
In other words, we hypothesize that both structures exist. We would most likely be able to differentiate the two structures via prosody; we have seen earlier that structures such as (189) would have CP$_1$ stressed, but since C is null, the stress on the wh word, although present, would be less obvious to hear (see Hamlaoui’s prosodic mapping in (134)). The structure in (188) would have no such stressing pattern. I leave this issue as an hypothesis, to be verified in future research.

Finally, to conclude this section, I can mention that the fact that the est-ce que is obligatory, in all dialects and registers of French, to ask an inanimate subject question is most likely one of the factors that contributed to the spread of the wh est-ce que construction. Indeed, we have seen that until the 18th century, est-ce que was predominantly used with qu’, the other wh words all possible, but used to a lesser extent (although this might have to do with pure lexical frequency). Since it is quite hard to measure or to determine the exact cause of language change, this piece of information represents at least a step towards finding the cause behind the overwhelming spread of the wh est-ce que construction.

4.3 Typological considerations: French as an in situ language

Many languages, related or not to French and with or without any equivalence to the wh est-ce que construction, express their wh-questions with a complementizer, or complementizer-like, element. Ambar (2003) noticed the correlation between having the possibility of wh in situ and the presence of a complementizer with fronted wh:

> If a language prefers non-inversion and wh-in-situ, then it tends to use complementizer elements in root wh-questions. Conversely, if a language prefers inversion and fronted wh-questions, then it tends to not have complementizer elements in root wh-questions.

(Ambar 2003: 41, emphasis my own)

Brazilian Portuguese, a Romance language like French, presents such a correlation, as do Malay, an Austronesian language, and Lamnso’, an African language from the Niger-Congo family.

190. Ali membeli apa?
    Ali buy what
191. **Apa Ali beli?**

what Ali buy

192. **Apa yang Ali beli?**

what **that** Ali buy

‘What did Ali buy?’

(Malay: Cole, Hermon & Aman 1999: 2-3)

193. **[dzə ká ] wo a ki-ji**

FM\(^{96}\) what **that** 2sg **PAST-eat**

‘What did you eat?’

(Lamnso’: Moghaddam 2010: 29)

Many other languages behave in a similar way when it comes to the complementizer appearing alongside the wh- word; Yoruba (West-Africa), Krio (Sierra Leone), most Northern Italian dialects, Cypriot Greek, some dialects of Norwegian, some dialects of Dutch, etc. These kinds of data have often been analysed as being instances of clefted questions (see, among others, Cattaneo 2006; Cole et al. 1999).

Although most languages that have been analysed as allowing clefted questions exhibit ‘reinforced wh’ when the wh element is fronted, it is sometimes hard to determine what type of lexical element actually appears with the wh. It does not seem to always be a complementizer, like Ambar suggests. In some languages it is a focus particle, in some a copula, in others a complementizer-like element, etc. In reality, this element is hard to identify or to translate into English. For example, in languages that do not have copulas (such as most Austronesian languages), it is hard to determine without a doubt the exact nature of the element that appears alongside the fronted wh. In Lamnso’, what has been glossed as a focus marker in the example (191) also appears in declarative clefts, so it might be equivalent to a copula, although it has been analysed as a focus marker by Moghaddam (2010). In contrast, Massam (2003) analyzes the marker that comes with wh elements in Niuean (an Austronesian language) as a predicate particle (**ko** in the example (192) below), which makes the interrogative construction closer to a

\(^{96}\) FM= Focus Marker.
cleft than to a focus construction, since she argues that the Niuean left periphery does not allow for D elements.

194. Ko hai ne figita e Moka
    pred who Nfut kiss Abs Moka
    ‘Who did Moka kiss?’ (Massam 2003: 97)

Given these facts, LaF seems to fit nicely within this group of languages, for many reasons: i. the fronted wh often appears with a ‘complementizer’; ii. the ‘complementizer’ is in fact an interrogative marker and a complementizer, although there is some debate (Elsig 2009, Rowlett 2007) – which is what we find in other languages as well, a reinforcer is sometimes called a copula, sometimes a focus marker; iii. the wh in situ variant is possible, and it appears bare. It thus follows that, according to the traditional dichotomy put forward by Cheng (1991), LaF possesses more properties of an in situ language than of a wh movement language. More work needs to be done to determine if all languages exhibiting an ‘apparent’ optionality systematically lack wh movement. It seems however that Cheng’s typology is still quite robust, but might need some type of modification if all languages that have apparent optionality have very similar properties; a third class of languages might be necessary to account for these systematic differences.

Finally, I include a word on the confusion between copulas, focus markers and complementizers, to link this issue to theories of grammaticalisation. We have seen in Chapter 3 that the grammaticalisation cycle involving a copula has a certain directionality; it goes from copula to grammatical marker, usually from IP to CP (Lohndal 2009). It is therefore not surprising that confusion might occur in some of the languages named above; there is a possibility that the grammaticalisation process is not yet complete. There is also a reason behind the fact that a copula goes from IP to CP: there are functional similarities between a copula and a focus marker (emphatic, presentational meaning for instance), and it is possible that the two get interchanged, even without the influence of a grammaticalisation process. Those relations between two functional elements are important when it comes to theory; either from a grammaticalisation or a predication point of view, they have to be taken into consideration, and further work might allow us to determine what role wh interrogatives actually play within this clefting/focus domain.
4.4 Interrogation in LaF: unified system?

Could my proposal fit into a unified analysis of interrogatives in LaF, i.e., could it be extended to include total (yes/no) interrogatives? I have briefly mentioned that the system of yes/no interrogatives in LaF is distinct from the one found in France French; it allows one more variant, namely the -tu variant, as in:

195. Elle aime-tu le beurre puis toi tu aimes-tu le beurre?  
   she likes-INT the butter and you you like-INT the butter  
   ‘Does she like butter, then do you like butter?’

(OH.099.1037 – Elsig 2009: 87)

The -tu in LaF has been analysed as an interrogative marker/particle (Morin 2009, in progress; Vinet 2001). Given Cheng’s typology (1991), it is not surprising that LaF, as an in situ language, would have a yes/no particle to mark interrogation. This provides one more piece of evidence towards an analysis in terms of long distance agreement instead of the traditional wh movement, since it is expected in an in situ language to have an overt interrogative particle in their yes/no interrogation system.

However, it is important to keep in mind that the two systems – total and partial interrogatives – in LaF are very different. The distribution of variants, as well as the factors favouring one over the other, differs largely from the partial interrogative system (Elsig 2009). For instance, in total interrogatives, est-ce que is only marginally used, in under 8 percent of all occurrences (Elsig 2009: 104). Inversion is still alive, and it shares the system with the intonation and -tu variants.

Table 7. Variants of yes/no (total) interrogatives (based on Elsig 2009: 104, OH corpus (20th century) only)

<table>
<thead>
<tr>
<th>Variant</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inversion</td>
<td>32.2</td>
<td>205</td>
</tr>
<tr>
<td>Intonation</td>
<td>29.2</td>
<td>186</td>
</tr>
<tr>
<td>-tu</td>
<td>30.8</td>
<td>196</td>
</tr>
<tr>
<td>Est-ce que</td>
<td>7.8</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>637</td>
</tr>
</tbody>
</table>
Usage is therefore divided fairly equally between the three main variants, inversion, intonation and -tu. Note however that inversion (pronominal inversion only, not complex or stylistic inversion) is favoured in only certain specific contexts: for instance, it is exclusively used with second person subjects tu and vous (Elsig 2009).

It seems that the est-ce que present in yes/no interrogatives is not the same as the one present after wh words, since it is the marked variant in this system, associated to a higher register; Elsig (p. 170) even mentions that the est-ce que variant in yes/no interrogatives is not part of the vernacular system at all. Therefore, we can hypothesize that est-ce que in yes/no questions has not grammaticalised in the same way as it has in wh questions. Recall that its appearance in the language is much more recent than with wh words, its first attestation dating back to only the 16th century, four to five centuries later than the wh est-ce que. The est-ce que variant never took over usage in yes/no interrogatives as it did with wh interrogatives.

Additional evidence towards proposing a different stage of grammaticalisation for the two est-ce que is the fact that in yes/no questions, est-ce que always appears in its full form [esk], whereas, as we have seen, it can be reduced in many ways when used as a grammatical marker in wh questions. Both Elsig (2009) and Druetta (2002, 2003) corroborate this fact in the corpora they studied, which leaves no doubt concerning their differences: est-ce que in yes/no questions and est-ce que in wh questions may be homophonous and originate from the same lexeme, but they are now two distinct lexical elements, both in LaF (Elsig) and in France French (Druetta).

More work would need to be done to determine if one of the two interrogative CPs postulated in this thesis for the wh system could be compatible with the yes/no system. The default variant of vernacular LaF is the -tu variant, which is also a grammatical marker (interrogative particle), the same grammatical function proposed for the est-ce/c’est. However, as exemplified in (195) above, -tu is a post-verbal particle, whereas est-ce/c’est appears adjacent to the wh word. To posit that -tu is under C, one would need to assume I-to-C verbal movement (which is roughly what Morin (2009) proposes), or some type of lowering process to get the right word order at spell-out. Moreover, the distribution of -tu has restrictions (the definiteness of the subject, presence of negation, etc.) that est-ce/c’est do not – see Morin (2009, in progress) for details. More work needs to be done to determine if the interrogative CP that is projected in yes/no questions is the same as one of the two posited in sections 4.2.2. and 4.2.3; do yes/no questions
have an operator in their SpecCp as well? If so, does it agree with the interrogative particle? But regardless of the analysis that one can attribute to the -tu particle, the fact that LaF wh questions can be explained without the use of wh movement is quite desirable, given Cheng’s (1991) generalisation that a language with wh in situ usually has an interrogative particle in its yes/no system. Future research that would include the whole paradigm of LaF interrogatives could certainly shed more light on the parallels (and differences, if applicable) of the two systems.

4.5 Sociolinguistic data and linguistic theory

There is great difficulty in working with usage data, that is, taken from a corpus gathered with a variationist approach in mind, to propose a syntactic analysis. In syntactic literature, there are often distinct analyses proposed for specific constructions that are unique to certain dialects. However, in order to understand the structural particularities of such constructions, the speakers can be asked to give grammaticality judgements that would force them to access a different grammar than the one that is being tested. For example, when proposing a wh est-ce que construction without wh movement, one needs to use diagnostics to prove if a certain lexical element shows the particularities of movement or not, overt or covert. However, LaF speakers must turn to a more standard French grammar in order to answer, since some quantifiers, negators, etc. included in those tests are used only in a grammar (register) that is still characterised by wh movement. How can we be sure that our speakers’ judgements come from the grammar (register) under study? I hypothesise that this discrepancy between judgements asked of speakers and the ‘working’ grammar of these same speakers is one of the reasons why the proposals to explain the occurrence of in situ, in particular, are so polarised in the literature; since in situ is part of the spoken, vernacular grammar, those judgements do not match. Moreover, the data presented in Chapter 2 point towards a system that is not quite stable; the wh interrogative system, in LaF but most likely in France French as well, is changing, and multiple grammars are competing. According to what has been proposed in the first part of this Chapter, LaF’s multiple grammars for wh interrogatives can be divided roughly into two types: one grammar includes all variants of wh est-ce que, including in situ, and it does not involve wh movement, and a second grammar includes variants that have subject-verb inversion, in which wh movement is mandatory. The first one is oral, vernacular speech and is acquired, whereas the second one is formal, found in written registers, and is learned. This corresponds to Zribi-Hertz’s (2011) definition of diglossia, which is described in the next section.
4.5.1 The diglossic approach

Zribi-Hertz (2011) proposes a diglossic approach to the study of French; since all speakers have access to registers that are, arguably, so different from each others that they might correspond to different grammars, it is necessary to approach the study of the language accordingly.

There are two definitions of diglossia that are used in Zribi-Hertz’s approach, which are reproduced here.

“Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language […] there is a very divergent, highly codified [...] superposed variety, the vehicle of a large and respected body of written literature [...], which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.”

(Ferguson 1959: 336, cited in Zribi-Hertz 2011: 3)

“Diglossia (diglossic). A term used in sociolinguistics to refer to a situation where two very different varieties of a language co-occur throughout a speech community, each with a distinct range of social function. Both varieties are standardised to some degree, are felt to be alternatives by native speakers and usually have special names. Sociolinguists usually talk in terms of a high (H) variety and a low (L) variety, corresponding broadly to a difference in formality: the high variety is learned in school, tends to be used in church, on radio programmes, in serious literature, etc., and as a consequence it has greater social prestige; the low variety in family conversations, and other relatively informal settings. Diglossic situations may be found in Greek (H: Katharevousa; L: Dhimotiki), Arabic (H: classical; L: colloquial), and some varieties of German (H: Hochdeutsch; L: Schweizerdeutsch, in Switzerland).”

(Crystal 1991: 104, cited in Zribi-Hertz 2011: 3-4)

Both definitions are very similar. For the sake of uniformity, I will use here Crystal’s terminology of H and L varieties, despite the obvious lack of social objectivity associated to these terms. The diglossic approach was first put forward in reaction to the traditional way of

97 Admittedly, there are very few terms that can be considered objective; ‘standard’ has been associated with specific dialects spoken by certain speech communities (Parisian French, for example), when what linguists usually mean by standard is an abstract variety that is mainly used in written registers, and that is
treating variation in French as a distanciation, more or less important, from the ‘central’ French (i.e. Paris – we have seen this reality referred to as central vs. peripheral French) (Massot 2010: 87).

Note that we are discussing here one H variety and many L varieties, but this might not be exactly true; since H refers to a normative state of language, which is learned and used in formal situations, it has been argued that the norm coming out of France is not the only one that speakers of L varieties have a knowledge of: Valdman (2000b) refers to a ‘français standard québécois’, which might vary slightly from the ‘français standard’ coming out of France, especially where language teaching is concerned. The exact definition and which language particularities it includes are not crucial for my purposes; I will come back to what this means for the LaF wh interrogative system below.

The main characteristics of each grammar available to a French speaker, are summarized in Table 8 below, based on Zribi-Hertz (2011) and Massot (2010). Note that I use Crystal’s terminology, and that issues that can be applied to the LaF wh interrogative system only are included.

learned in school (more details below); no speech community actually speaks the standard variety as a default (acquired) variety. Similarly, the term ‘informal’ has been rejected by Zribi-Hertz, mainly on the basis that in French the meaning of informel is not the same as the English term informal. Zribi-Hertz uses the term dialectal, which is not very popular in North American literature, because it is often used in contrast with Parisian French. I consider L varieties to include all spoken varieties, Parisian French as well as Louisiana French, for example. The distinction will become clearer following the argumentation presented in this section. Massot (2010) uses the terms français démotique for L varieties and français classique tardif for the H variety, and, as mentioned in previous chapters, Rowlett (2007) used the terms ‘Contemporary French’ and ‘Modern French’.
Table 8. Distinctions between French High and Low varieties (H vs. L)

<table>
<thead>
<tr>
<th><strong>High Variety (H)</strong></th>
<th><strong>Low Varieties (L)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired passively before school (mainly through media)</td>
<td>Acquired actively (according to language acquisition principles)</td>
</tr>
<tr>
<td>Learned (actively) in school</td>
<td>Usage often discouraged in school</td>
</tr>
<tr>
<td>Used in written language (used orally in highly formal contexts only)</td>
<td>Used orally</td>
</tr>
<tr>
<td>Grammar used for the most part in media (news, newspapers, radio, etc.)</td>
<td>Might be used in media (in sitcoms, for example)</td>
</tr>
</tbody>
</table>

Zribi-Hertz (2011: 11-12) makes a distinction between ‘passive’ and ‘active’ acquisition/learning. For her, children have access to the H variety before entering school only in a very limited way, usually through media. It is the L variety that is acquired through normal language acquisition (which means, according to the link between acquisition and language change described in Chapter 3, that the L varieties are the systems going through language change, not the H variety). Education is linked to the use of the H variety, since less educated adults may never acquire a fully active proficiency of the H variety (and not mastering the H variety is often socially stigmatized). The implication of these definitions of the H variety is that one cannot acquire full proficiency without formal education.

Note that, as Massot (2010: 168) mentions, these two types of grammars are not étanches ‘closed’, rather, they are merely concurrentes ‘in competition’. He illustrates the reality of this ‘communication’ between grammars using Venn diagrams, as in Figure 11.
Figure 11. Representing a diglossic situation

![Diglossic Situation Diagram](image)

(based on Massot 2010: 90)

The reality of diglossia is therefore continuous; it is not the case that speakers have access to two distinct grammars without any interactions between the two: the two grammars are constantly intersecting. Why then propose a discontinuous diglossic situation and not a variation continuum? My arguments are solely based on the construction described in this thesis. I have proposed an analysis for the different variants of the wh *est-ce que* construction. I have largely ignored the oldest wh variant, namely the variant that involves wh movement and that triggers inversion of the subject and the verb, as in English. My claim is that the variants of the wh *est-ce que* are part of LaF speakers’ L variety, and that the wh fronting with inversion is part of the H variety of French. I propose a diglossic situation in LaF mainly because the two grammars are quite distinct in this particular dialect. Massot (2010: 89) (based on Zribi-Hertz 2011) provides three conditions for speakers to be in a diglossic situation [my translation].

i. they consider that they speak only one language (to be distinguished from bilingualism);

ii. they acquire two linguistic competences that are identifiable and competing, each one independent but not closed (to be distinguished from simple variation);

iii. the two grammatical competences are used for two distinct language functions: one for Low and one for High (i.e., informal and formal situations).

I believe that these three conditions are present for LaF speakers, especially when it comes to the wh interrogative system. The first criterion is not debatable, since all speakers of LaF consider themselves speakers of French (some might say ‘bad French’ or ‘broken French’, but French nonetheless). The second criterion, when applied to the wh interrogative system, is also clearly met; I have shown that the wh *est-ce que* variants are very different from the wh fronting with
inversion variants. Finally, the third criterion is also filled; Elsig (2009) has shown that in the sociolinguistic corpus he used (in which the data is gathered so that the register used in the interviews mirrors as much as possible the register of an informal conversation), wh fronting with inversion represents barely 2 percent of all occurrences. It therefore seems to be the case that the discourse situations in which a speaker would use the H variety over the L variety are very specific.

For all those reasons, we consider the LaF wh interrogative system to be characterised by a diglossic situation, vernacular LaF being their L variety, and ‘standard’ French being their H variety. In terms of wh interrogation, Table 9 below shows the distribution of all variants according to L or H varieties. For the sake of comparison, I also included variants of the yes/no interrogative system, based on Elsig’s (2009) observations.

Table 9. Distribution of interrogative variants within the LaF diglossic situation

<table>
<thead>
<tr>
<th></th>
<th>L variety</th>
<th>H variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh system</td>
<td>- wh est-ce que and variants;</td>
<td>- wh est-ce que (only with an inanimate subject wh);</td>
</tr>
<tr>
<td></td>
<td>- wh fronting;</td>
<td>- wh fronting with inversion.</td>
</tr>
<tr>
<td></td>
<td>- wh in situ.</td>
<td></td>
</tr>
<tr>
<td>Yes/no system</td>
<td>- inversion (only with tu and vous);</td>
<td>- inversion;</td>
</tr>
<tr>
<td></td>
<td>- intonation;</td>
<td>- est-ce que.</td>
</tr>
<tr>
<td></td>
<td>- tu.</td>
<td></td>
</tr>
</tbody>
</table>

This thesis included an analysis for all variants of the L variety, since they are all part of the same grammar. Wh fronting’s relative formality is hard to determine, as we have seen in Chapter 2 (section 2.2.2, in which it was said that wh fronting is the less well known of all variants), but since it is quite common in LaF (17 percent of all wh interrogatives (Elsig 2009: 147)), I can assume that it is only a variant of the wh est-ce que, with the grammatical marker est-ce/c’est as well as the complementizer spelled-out as null (see details earlier in this chapter, section 4.2.4).
Notice also that in yes/no interrogatives, I hypothesized, following Elsig (2009) and Druetta (2002, 2003), that the est-ce que variant is not part of the same grammar as the wh est-ce que. Finally, we know that in a diglossic situation, all speakers have a complete mastery of their L system, but variable knowledge of the H variety, which explains the relative distribution of each variant in usage.

4.6 The syntax of speech: l’oralité

The data and analyses presented in this thesis have concerned mainly a system used by speakers in vernacular, conversational contexts. The wh interrogative system data fit very well within the approaches that have recently been put forward to describe what has been called oralité ‘orality’. Françoise Gadet has been a long time advocate for an approach to dialectal differences in terms of orality; the approach assumes that most dialectal characteristics are to be attributed to the oral (spoken) status of most of the dialects. This approach puts forward the idea that French dialects all share a lot more characteristics than previously thought, since they all have in common a few traits that are to be attributed to their status as spoken language. Such traits include, but are not restricted to (based on Neumann-Holzschuh 2012: 3):

- expressive linearization (more topicalization, dislocation, etc.);
- structural ambiguities;
- lack of overt morphosyntactic markers;
- usage of multi-purposed elements.98

This list of characteristics is not exhaustive, nor does it all apply in the case of the LaF wh interrogative system. However, it has been shown (Gadet 2007, 2011; Neumann-Holzschuh

98 Most of these characteristics are enclosed within the definition of the term parataxe, which is widely used in the literature on regional varieties of French. Parataxe describes the characteristic of oral speech that makes the flow spasmodic, because of the missing linkers, conjunctions, etc. that create grammatical links between each proposition (grammatical links are often replaced by pauses and context in speech). Parataxe is usually used in contrast to hypotaxe, which characterizes a profusion of conjunctions used in very long and highly embedded (most often written) sentences.
that many elements are systematically found in dialects of French, which might share a common source, but are no longer in constant contact. For example, we find in most dialects of French the omnipresence of the conjunction *que*. Where Standard French can introduce adjunct subordinate clauses with conjunctors such as *dont, où, etc.*, most dialects of French will use *que* in all instances (Gadet 2011, Neumann-Holzschuh 2012, Valdman 2000a, Bauche 1946).

196. *c’est les deux endroits *que* j’ai travaillé* (cf. *où*) *New Brunswick French*

it-is the two places that I-have worked

‘These are the two places where I worked.’

(Wiesmath 2002: 400, cited in Neumann-Holzschuh 2012: 5)


my parents worked really hard that-we go to the-school

‘My parents worked really hard so that we could go to school.’

Similarly, some characteristics of the LaF wh interrogative system are not unique to LaF, but can also be found in other vernacular varieties. For example, we know that we find various forms of *wh est-ce que* in Prince-Edward Island (PEI) French (King 1991) as well as in Ontario French (Starets 2002). Note however that I do not believe that variation with respect to Standard French can be reduced to registers (oral vs. written). Even if PEI French can make use of similar variants to those found in LaF, some of them are crucially absent, like the wh+complementizer only. This may be evidence that the wh *est-ce que* bundle of variants in that variety might not be explained through a reduced cleft construction, but via some other syntactic mechanisms. For example, if the complementizer only variant is impossible, it might be a sign of a monoclausal structure. Of course, more work needs to be done to understand each variety, but although they all share characteristics of spoken registers, I do not claim that their underlying syntactic system is the same. Although most dialects of North American as well as ‘popular’ French from France (*français populaire*, Gadet 1992) all share common characteristics due to their vernacularity, their syntactic system can still vary, and it is only with studies such as the present one, describing the underlying structures of variation, that we can know for sure the limits of this variation.
“... mais on peut suggérer que la variabilité est une intrication complexe entre effets de facteurs historiques, écologiques et sociolinguistiques d’une part, et facteurs systémiques ou cognitifs, deux ordres de modalité qui vont interagir, par faisceau ou divergence selon les phénomènes”

“we can suggest that variation is a complex mix between historical, ecological and sociolinguistic effects, and factors from the internal system (cognition), which are two modalities that will interact in different ways according to each phenomenon” [my translation]


This thesis includes these two types of modalities: the external factors (historical, ecological and sociolinguistic above) described mainly in Chapter 2, as well as the internal factors (internal system (cognition) above), described mainly in Chapters 3 and 4. It was shown that the variation that characterises oral speech can be explained in a fairly straightforward way when considering the different factors at play.
Chapter 5

5 Conclusion

In this thesis, I have provided an overview of the LaF wh interrogative system. The main questions that were answered throughout the chapters of this thesis concerned the origins of the extensive variation found in today’s system, and how we can account for it synchronically. This type of work was much needed in French linguistics; addressing the theoretical problem caused by such important variation after surveying systematically many types of empirical evidence – synchronic and diachronic. To my knowledge, it had not been done, and it was necessary given all the debates surrounding not only the apparent optionality of wh movement in French, but also the status of LaF itself, as a system that is different from the more standard variety of French.

The contributions to the field were divided into three domains; I provided new details in historical linguistics, in grammaticalisation theories, and in theoretical syntax, in addition to providing reflexions about the linguistic study of the French language, such as diglossic situations or oralité. The main original findings/proposals are threefold:

i. I showed that the extensive variation is mainly due the evolution of the wh est-ce que construction, present in the language since Old French;

ii. I provided more data to the study of the linguistic cycle; the est-ce que element went through two different grammaticalisation processes between Old and Modern French. The copula est exhibited all characteristics of the last two steps of the Copula Cycle (Lohndal 2009), which is important since data to support this particular cycle are sparse;

iii. I showed that a unified analysis in terms of absence of wh movement – an avenue that had never been explored in detail before – can account for the facts in LaF. I provided evidence, both historical and synchronic, to show that there are two interrogative CPs in LaF, one which is selected as an adjunct by a main CP, and one which selects a regular IP as its complement. The latter is more recent than the former; it was first attested in texts during the nineteenth century, whereas the former is a product of the evolution of the Old French wh est-ce que interrogative cleft. Moreover, I argued that most – if not all – variants of LaF are in fact evolutions of the Old French interrogative clefted construction,
not of the wh fronted with inversion variant. I showed that comparing the wh fronting with inversion to the wh *est-ce que* variants may not only yield blurry results, but it also does not make sense historically.

This contribution has assembled different types of data, and attempted to answer theoretical questions while relying on corpus or experimental data. We have seen that the type of data considered when proposing an analysis often has a great influence on that analysis; although grammatical judgements are necessary to overcome the realities of negative evidence, they are not very reliable in a diglossic situation, and can create confusion. I showed that, concerning the restrictions on wh in situ, what has been attributed to dialectal differences was in fact for the most part only due to speakers confusing their different grammars when asked to access a low frequency construction/structure.

The consequences of proposing a wh system without any wh movement for LaF are numerous, and such a proposal is quite innovative. More work needs to be done on the prosodic facts and their specific effect on the licensing of each CP (the wh *est-ce que* one and the in situ one), but the lack of movement unifies the ‘fronted’ and in situ variants in a way that is more economical than what has been proposed in terms of covert movement (Bošković 1998) or remnant movement (Poletto & Pollock 2004). The overwhelming similarities between LaF and Austronesian languages provide even more evidence towards considering French as an in situ language, rather than a wh movement language like English, as it has been traditionally described. In such a system, the in situ variant becomes more of a normal consequence than an anomaly in need of an explanation.

The French, and specifically the LaF, wh interrogative system provides linguists of all subfields a great deal of interesting data. Of course, many more things need to be said on the system. For instance, a detailed quantitative historical study of embedded interrogatives, as well as of matrix and embedded interrogatives before the 16th century, would be necessary for us to better understand the conditions, internal and external, that were necessary for the reanalyses proposed in this thesis to take place.

Theoretically, recent proposals regarding the interface of syntax and prosody are only at their first stage, since many things still need to be tested and understood.
Finally, it is important to note that I consider the wh interrogative system in LaF to be a changing system. The interaction between focus, topicalization and interrogation makes the wh interrogative system especially susceptible to change, since many variants are possible. If we follow what is predicted from grammaticalisation cycles, we can expect the structure to get more fixed, and the variation within the different wh *est-ce que* variants to be more and more reduced. Given the fact that economy principles are at the center of the linguistic cycles (van Gelderen 2006), it will be interesting to see if the wh in situ will take over the whole system, since its structure is simpler and requires fewer projections to be merged. In terms of grammaticalisation theory, we will have to continue to watch both the wh and the yes/no interrogative systems in LaF, since both seem to be constantly evolving, due to their susceptibility to variation – they represent a pool of data that could provide insightful information for the development of new theories of language variation and change.
References


Demol, Annemie & Els Tobback. 2009. The distribution of *IL* and *CE* in French copular sentences: Semantic and informational aspects. In *Studies van de BKL/Travaux du BCL/Papers of the LSB* 4, ed. Svetlana Vogeleeer, Frank Brisard, Philippe De Brabanter, Patrick Dendale, Bert Le Bruyn. Published online.


Massot, Benjamin. 2010. Le patron diglossique de variation grammaticale en français. Langue française 168.4, pp. 87-106.


Moghaddam, Safieh. 2010. Ex-situ Focus Versus In-Situ Focus, manuscript, University of Toronto.


Neumann-Holzschuh, Ingrid. 2012. Le français louisianais et la syntaxe de l’oral. Presentation given at the Summer Institute : Quand la Louisiane parle français. May 28-June 1, Tulane University, New Orleans, LA.


Rey, Alain (dir.). *Le Grand Robert de la langue française*. [electronic version], 2nd ed.


Tailleur, Sandrine. 2009. *What is que c’est que c’est? The Paradigm of Wh-Questions in Quebec French*. Manuscript, University of Toronto.


### Appendix

Complete list of contemporary LaF subject and object interrogative variants

<table>
<thead>
<tr>
<th></th>
<th>Subject animate</th>
<th>Subject inanimate</th>
<th>Object animate</th>
<th>Object inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Who came?</em></td>
<td>Il est venu qui?</td>
<td>Il brille quoi?</td>
<td>Tu vois qui?</td>
<td>Tu veux quoi?</td>
</tr>
<tr>
<td><em>Who do you see?</em></td>
<td>Qui qui est venu?</td>
<td>? Quoi qui brille?</td>
<td>Qui que tu vois?</td>
<td>? Quoi que tu veux?</td>
</tr>
<tr>
<td><em>What do you want?</em></td>
<td>Qui c’est qui est venu?</td>
<td>C’est quoi qui brille?</td>
<td>C’est qui que tu vois?</td>
<td>C’est quoi que tu veux?</td>
</tr>
<tr>
<td><em>Wh in situ</em></td>
<td>Qui c’est qui est venu?</td>
<td>Qui c’est qui brille?</td>
<td>Qui c’est que tu vois?</td>
<td>Qui c’est que tu veux?</td>
</tr>
<tr>
<td><em>Fronted Wh + inversion</em></td>
<td>Qui c’est qui est venu?</td>
<td>Qui c’est qui brille?</td>
<td>Qui c’est que tu vois?</td>
<td>Qui c’est que tu veux?</td>
</tr>
<tr>
<td>Wh + <em>comp</em></td>
<td>Qui qui est venu?</td>
<td>? Quoi qui brille?</td>
<td>Qui que tu vois?</td>
<td>? Quoi que tu veux?</td>
</tr>
<tr>
<td>c’est + Wh + <em>comp</em></td>
<td>C’est qui qui est venu?</td>
<td>C’est quoi qui brille?</td>
<td>C’est qui que tu vois?</td>
<td>C’est quoi que tu veux?</td>
</tr>
<tr>
<td>Wh + <em>est-ce</em> + comp</td>
<td>Qui est-ce qui est venu?</td>
<td>Qui est-ce qui brille?</td>
<td>Qui est-ce que tu vois?</td>
<td>Qui est-ce que tu veux?</td>
</tr>
<tr>
<td>Wh + <em>est-ce</em> + que c’est + comp</td>
<td>Qui est-ce que c’est qui est venu?</td>
<td>Qui est-ce que c’est qui brille?</td>
<td>Qui est-ce que c’est que tu vois?</td>
<td>Qui est-ce que c’est que tu veux?</td>
</tr>
<tr>
<td>Wh + c’est + comp</td>
<td>Qui c’est qui est venu?</td>
<td>Qui c’est qui brille?</td>
<td>Qui c’est que tu vois?</td>
<td>Qui c’est que tu veux?</td>
</tr>
<tr>
<td>Wh + c’est + que c’est + comp</td>
<td>Qui c’est qui est venu?</td>
<td>Qui c’est qui brille?</td>
<td>Qui c’est que tu vois?</td>
<td>Qui c’est que tu veux?</td>
</tr>
<tr>
<td>c’est + Wh + c’est + comp</td>
<td>C’est qui c’est qui est venu?</td>
<td>C’est quoi c’est qui brille?</td>
<td>C’est qui c’est que tu vois?</td>
<td>C’est quoi c’est que tu veux?</td>
</tr>
<tr>
<td>c’est + Wh + c’est + que c’est + comp</td>
<td>C’est qui c’est qui est venu?</td>
<td>C’est quoi que c’est qui brille?</td>
<td>C’est qui c’est que tu vois?</td>
<td>C’est quoi que c’est que tu veux?</td>
</tr>
<tr>
<td>c’est + Wh + <em>est-ce</em> + comp</td>
<td>C’est qui est-ce qui est venu?</td>
<td>C’est qu’est-ce qui est venu?</td>
<td>C’est qui est-ce que tu vois?</td>
<td>C’est qu’est-ce que tu veux?</td>
</tr>
<tr>
<td>c’est + Wh + <em>est-ce</em> + que c’est + comp</td>
<td>C’est qui est-ce que c’est qui est venu?</td>
<td>C’est qu’est-ce que c’est qui brille?</td>
<td>C’est qui est-ce que tu vois?</td>
<td>C’est qu’est-ce que c’est que tu veux?</td>
</tr>
</tbody>
</table>

| 99 | When the subject is questioned, an expletive *il* is necessary when using the in situ variant, and the fronted with inversion variant is impossible (the standard form being the wh *est-ce que*).

| 100 | The *est-ce* or *c’est* immediately following the wh element can be phonologically reduced to [s], which is why it is often represented orthographically by ‘s’. However, the ‘que c’est’ that is sometime s present right before the complementizer cannot be reduced in the same way. The ‘que’ part can be reduced to [k], but the ‘c’est’ part always stays [se], never just [s]. The comp also always has the possibility to be phonologically deleted.
<table>
<thead>
<tr>
<th>Wh + que c’est + comp</th>
<th>Qui que c’est qui est venu?</th>
<th>? Quoi que c’est qui brille?</th>
<th>Qui que c’est que tu vois?</th>
<th>? Quoi que c’est que tu veux?</th>
</tr>
</thead>
<tbody>
<tr>
<td>c’est + Wh + que c’est + comp</td>
<td>C’est qui que c’est qui est venu?</td>
<td>C’est quoi que c’est qui brille?</td>
<td>C’est qui que c’est que tu vois?</td>
<td>C’est quoi que c’est que tu veux?</td>
</tr>
</tbody>
</table>